Technical Performance and Patient Outcomes in Surgery: Another Piece of the Puzzle

by

Andras Botond Fecso

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Institute of Medical Science
University of Toronto

© Copyright by Andras Botond Fecso 2018
Technical Performance and Patient Outcomes in Surgery: Another Piece of the Puzzle

Andras Botond Fecso
Doctor of Philosophy
Institute of Medical Science
University of Toronto
2018

Abstract

**Introduction:** Patient outcomes depend on a variety of variables related to patients themselves, to disease or treatments. In surgery, operations are the mainstay of management and the operating room is a high-stakes, high-risk environment where the surgeon plays an invaluable role. Extensive literature demonstrates how the aforementioned variables affect patient outcomes, however when it comes to surgeon technical skills, the existing evidence is limited. Therefore the main goal of this thesis was to provide evidence and to emphasize the importance of technical skills and to strengthen the current literature.

**Material and Methods:** Research designs contained within this dissertation include: (1) a systematic review to summarize the existing evidence, (2) a retrospective technical skills analysis, (3) a survey study to define the participants in the operating room and a (4) prospective observational study to determine the relationship between technical and non-technical intraoperative performances.

**Results:** Four research studies were completed. The systematic review demonstrated that technical performance affects patient outcomes, however the quality of evidence is low. The retrospective skill analysis addressed the main methodology deficiencies identified in the review.
and confirmed that technical performance does affect patient outcomes in laparoscopic gastric cancer operations. The survey study identified that although technical skills have been shown to be important, trainees do not routinely perform the main steps of these operations. Finally, the prospective observational study demonstrated that technical and non-technical performances are related and both are important complements to each other.

**Conclusion:** Technical performance affects patient outcomes. Technical and non-technical performances are related. Ongoing assessment and enhancement of these skills, using modern strategies such as structured feedback, deliberate practice, coaching etc., might improve patient outcomes. Future studies should perform a comprehensive assessment of the operating room and design educational interventions for all participants, regardless of their level of training or experience.
Acknowledgments

This thesis would not have been possible without the help and guidance of all the fantastic people I was fortunate enough to work with during my journey towards this PhD degree.

First and foremost, I would like to thank my supervisor, Dr. Teodor Grantcharov for giving me the opportunity to be part of his research team, for introducing me to the world of surgical education and patient safety as well as for offering me his continued support, encouragement, guidance and mentorship.

I would also like to thank my Program Advisory Committee (PAC), Drs. Ori Rotstein, Guylaine Lefebvre and Allan Okrainec for their invaluable advice and guidance, constructive criticism feedback and support throughout my studies.

I further thank my fellow researchers Drs. Esther Bonrath, Nicolas Dedy, Marisa Louridas, Peter Szasz, Alaina Garbens, Mitchell Goldenberg, James Jung as well as the rest of the research team, Drs. Sara Elkabany, Bijan Dastgheib, Bojan Macanovic, Mr. Karthik Raj and Mr. Georgi Kerezov, for their continuous support, feedback, contribution, teamwork and friendship.

Thank you, Dr. James Rutka in the Department of Surgery, Dr. Norman Rosenblum in the Clinician Investigator Program, Drs. Michael Fehlings and Andras Kapus in the Surgeon Scientist Training Program, Dr. Carol Swallow in the Division of General Surgery and my Program Director, Dr. Najma Ahmed, for giving me the opportunity to undertake this research opportunity as part of my training.

Furthermore, I wish to acknowledge all the surgeons and residents at the University of Toronto who participated and provided data for my studies. Without their contribution these research projects would not have been possible.

Finally, without the ongoing love, help and support of my wife, Krisztina, our two little daughters, Gabriella and Viktoria, and my entire family, I would not be where I am today. Thank you for your unconditional love and support, I am forever indebted to you all.
Contributions

I, Andras B. Fecso, solely prepared this thesis and am the first author of all manuscripts associated with this PhD degree and presented in this dissertation. My contributions include the conception, planning, design, data acquisition and interpretation as well as the preparation of the manuscripts.

Dr. Teodor Grantcharov (supervisor) guided me throughout all phases of this degree and was involved in the conception, planning, design, data interpretation and critical review of all manuscripts in Chapter 1, 3, 4, and 5 as well as this thesis.

My Program Advisory Committee, Drs. Ori Rotstein, Guylaine Lefebvre and Allan Okrainec guided me throughout the studies, offering exceptional advise regarding planning, execution and interpretation of the results, as well as critically reviewed this thesis.

Dr. Peter Szasz co-authored Study No. 1 and was involved in the data interpretation and critical review of the manuscript in Chapter 1.

Medical student, Mr. Georgi Kerezov, co-authored Study No. 1 and was involved in the data acquisition and critical review of the manuscript in Chapter 1 as well as contributed to the data acquisition for the manuscript in Chapter 3.

Christine Nielson, information specialist at St. Michael’s Hospital, Health Science Library, was involved in assisting with the literature search for the data acquisition in Chapter 1.

Dr. Junaid Bhatti co-authored Study No. 2 and was involved in the design of the methodology, the data interpretation and critical review of the manuscript in Chapter 3.

Drs. Peter Stotland and Fayez Quereshy co-authored Study No. 2 and were involved in the study design, data acquisition and critical review of the manuscript in Chapter 3.

Dr. Esther Bonrath co-authored Study No. 3 and was involved in the conception, planning, study design, data acquisition and interpretation as well as the critical review of the manuscript in Chapter 4. Dr. Bonrath also contributed to data acquisition for Study No. 4 in Chapter 5.
Drs. Sebnem Kuzulugil, Ceni Babaoglu and Ayse Bener co-authored Study No. 4 and were involved in the data interpretation and critical review of the manuscript in Chapter 5.

Drs. Nicolas Dedy, Sara Elkabany and Bijan Dastgheib contributed to data acquisition for Study No. 4 in Chapter 5.
# Table of Contents

Abstract ......................................................................................................................... iii

Acknowledgments ........................................................................................................ iv

Contributions ............................................................................................................... v

Table of Contents ........................................................................................................ vii

Funding ........................................................................................................................ xiv

List of Abbreviations ........................................................................................................ xv

List of Figures ................................................................................................................ xviii

List of Tables ................................................................................................................... xix

Thesis Overview ........................................................................................................... xx

CHAPTER 1 ...................................................................................................................... 1

1 Introduction ................................................................................................................... 2

1.1 Technical performance ......................................................................................... 4

1.1.1 Definitions ........................................................................................................ 4

1.1.2 Assessment of technical performance ........................................................... 4

1.1.2.1 Global Assessment Scales ........................................................................ 5

1.1.2.1.1 Objective Structured Assessment of Technical Skills ............................... 5

1.1.2.1.2 Global Operative Assessment of Laparoscopic Skills ............................... 8

1.1.2.1.3 Direct Observation of Procedural Skills ..................................................... 8

1.1.2.1.4 The Ottawa Surgical Competency Operating Room Evaluation .............................. 8

1.1.2.2 Task – Specific Assessment Scales ............................................................... 9

1.1.2.2.1 Generic Error Rating Tool ....................................................................... 10

1.1.2.3 Procedure - Specific Global Assessment Scales .......................................... 12

1.2 Non-Technical Performance ............................................................................... 13

1.2.1 Definitions ....................................................................................................... 13

1.2.2 Assessment of non-technical performance ..................................................... 14
1.2.2.1 Individual assessment scales

1.2.2.1.1 Non-Technical Skills for Surgeons

1.2.2.1.2 Objective Structured Assessment of Non-Technical Skills

1.2.2.1.3 Scrub Practitioners’ List of Intraoperative Non-Technical Skills

1.2.2.1.4 Anaesthetists’ Non-Technical Skills

1.2.2.2 Team assessment scales

1.2.2.2.1 Non-Technical Skills

1.2.2.2.2 Observational Teamwork Assessment for Surgery

1.3 Validity of Performance Assessments

1.4 Patient Outcomes

1.4.1 Definitions

1.4.2 Adverse outcomes in surgery

1.4.3 Classification of adverse outcomes

1.4.4 Contributing factors to adverse outcomes

1.4.4.1 Patient factors

1.4.4.1.1 American Society of Anesthesiologists classification

1.4.4.1.2 Charlson Comorbidity Index

1.4.4.2 Disease factors

1.4.4.3 Hospital factors

1.4.4.4 Surgical factors

1.5 The effect of technical performance on patient outcomes in surgery: a systematic review

1.5.1 Abstract

1.5.1.1 Objective

1.5.1.2 Summary Background Data

1.5.1.3 Methods

1.5.1.4 Results
3.1.1 Introduction........................................................................................................70
3.1.2 Methods ..........................................................................................................70
3.1.3 Results ............................................................................................................70
3.1.4 Conclusion .......................................................................................................71
3.2 Introduction..........................................................................................................72
3.3 Methods ...............................................................................................................73
  3.3.1 Patients ..........................................................................................................73
  3.3.2 Measures ........................................................................................................73
  3.3.3 Surgical Technique .......................................................................................74
  3.3.4 Data extraction .............................................................................................74
  3.3.5 Statistical analysis .........................................................................................75
3.4 Results ..................................................................................................................76
  3.4.1 Patient factors ...............................................................................................76
  3.4.2 Surgical factors .............................................................................................78
  3.4.3 Postoperative complications .......................................................................80
  3.4.4 Predictors for postoperative complications ...............................................82
3.5 Discussion ............................................................................................................84
3.6 Conclusion ..........................................................................................................87

CHAPTER 4 ..................................................................................................................88
4 Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges and Potential Opportunities at a Large University Centre........................................89
4.1 Abstract ..............................................................................................................89
  4.1.1 Objective .......................................................................................................89
  4.1.2 Design ..........................................................................................................89
  4.1.3 Setting ..........................................................................................................89
  4.1.4 Participants ....................................................................................................89
  4.1.5 Results ..........................................................................................................89
5.3.5 Statistical Analysis .......................................................................................... 110
5.4 Results .................................................................................................................. 110
  5.4.1 Patient characteristics ...................................................................................... 110
  5.4.2 Technical Performance .................................................................................... 111
  5.4.3 Non-Technical Performance .......................................................................... 112
  5.4.4 Interaction Between Technical and Non-Technical Performances .............. 112
5.5 Discussion ............................................................................................................ 118
5.6 Conclusion ........................................................................................................... 120

CHAPTER 6 .............................................................................................................. 121
6 General Discussion ................................................................................................. 122
  6.1 Summary of Study Findings ............................................................................... 122
    6.1.1 Study No. 1 (Chapter 1) ............................................................................. 122
    6.1.2 Study No. 2 (Chapter 3) ............................................................................. 123
    6.1.3 Study No. 3 (Chapter 4) ............................................................................. 124
    6.1.4 Study No. 4 (Chapter 5) ............................................................................. 125
  6.2 Thesis Summary .................................................................................................. 126
    6.2.1 Strategies to enhance technical skills ............................................................. 126
    6.2.2 Technical and Non-Technical Skills .............................................................. 134
    6.2.3 Added Validity ............................................................................................. 135
    6.2.4 Outcome measures ...................................................................................... 137
  6.3 Thesis Conclusion ............................................................................................... 139

CHAPTER 7 .............................................................................................................. 140
7 Limitations ............................................................................................................. 141
  7.1 Generalizability of the results ............................................................................. 141
  7.2 Comprehensive assessment of the intraoperative environment ....................... 142

CHAPTER 8 .............................................................................................................. 143
8 Future directions .............................................................................................................. 144

8.1 Comprehensive assessment of the intraoperative environment.............................. 144

8.2 Increase generalizability .......................................................................................... 146

8.3 Design of educational interventions ......................................................................... 146

References ....................................................................................................................... 148

Copyright Acknowledgements ....................................................................................... 177
The research work in this thesis was funded by:

I. Physicians’ Services Incorporated (PSI) Foundation Resident Research Grant

II. John L. Provan Fellowship in Surgical Education, University of Toronto

III. Lawrence, Ila and William Gifford Scholarship in Radiation Oncology and Surgery, University of Toronto

IV. Edward Christie Stevens Fellowships in Medicine Postgraduate Research Award from the University of Toronto

V. The Department of Surgery and the Division of General Surgery with tuition and salary support
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGME</td>
<td>Accreditation Council for Graduate Medical Education</td>
</tr>
<tr>
<td>ANTS</td>
<td>Anaesthetists’ Non-Technical Skills</td>
</tr>
<tr>
<td>APACHE</td>
<td>Acute Physiology and Chronic Health Evaluation</td>
</tr>
<tr>
<td>ARDS</td>
<td>Acute Respiratory Distress Syndrome</td>
</tr>
<tr>
<td>ASA</td>
<td>American Society of Anesthesiologists</td>
</tr>
<tr>
<td>B2</td>
<td>Billroth II</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>CCI</td>
<td>Charlson Comorbidity Index</td>
</tr>
<tr>
<td>CDC</td>
<td>Clavien-Dindo Classification</td>
</tr>
<tr>
<td>CIHI</td>
<td>Canadian Institute for Health Information</td>
</tr>
<tr>
<td>CNS</td>
<td>Central Nervous System</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic Obstructive Pulmonary Disease</td>
</tr>
<tr>
<td>CVA</td>
<td>Cerebrovascular Accident</td>
</tr>
<tr>
<td>DIC</td>
<td>Disseminated Intravascular Coagulation</td>
</tr>
<tr>
<td>DISI</td>
<td>Disruption in Surgery Index</td>
</tr>
<tr>
<td>DOPS</td>
<td>Direct Observation of Procedural Skills</td>
</tr>
<tr>
<td>EBM</td>
<td>Evidence Based Medicine</td>
</tr>
<tr>
<td>EG</td>
<td>Esophago-Gastrostomy</td>
</tr>
<tr>
<td>EJ</td>
<td>Esophago-Jejunostomy</td>
</tr>
<tr>
<td>ERAS</td>
<td>Enhanced Recovery After Surgery</td>
</tr>
<tr>
<td>ESRD</td>
<td>End-Stage renal Disease</td>
</tr>
<tr>
<td>GERT</td>
<td>Generic Error Rating Tool</td>
</tr>
<tr>
<td>GJ</td>
<td>Gastro-Jejunostomy</td>
</tr>
<tr>
<td>GOALS</td>
<td>Global Operative Assessment of Laparoscopic Skills</td>
</tr>
<tr>
<td>IBM</td>
<td>International Business Machines</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>IHI</td>
<td>Institute for Healthcare Improvement</td>
</tr>
<tr>
<td>IO</td>
<td>Intraoperative</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile Range</td>
</tr>
<tr>
<td>JJ</td>
<td>Jejun-Jejunostomy</td>
</tr>
</tbody>
</table>
KLASS QC  Korean Laparoscopic Gastrointestinal Surgery Study Quality Control
LRYGB   Laparoscopic Roux-en-Y Gastric Bypass
LTMB    Long-Term Morbidity
LTMT    Long-Term Mortality
MERSQI  Medical Education Research Study Quality Instrument
MESH    Medical Subject Headings
MIS     Minimally Invasive Surgery
M&M     Morbidity and Mortality
MSKCC   Memorial Sloan Kettering Cancer Center
NOTECHS Non-Technical Skills
NOTSS   Non-Technical Skills for Surgeons
NO      Not Observed
N/R     Not Required
OR      Operating Room
OR      Odds Ratio
ORBB    Operating Room Black Box
OSATS   Objective Structured Assessment of Technical Skills
OSANTS  Objective Structured Assessment of Non-Technical Skills
OTAS    Observational Teamwork Assessment for Surgery
PCA     Post Conceptual Age
PCO     Patient Centered Outcomes
PGY     Post Graduate Year
PHD     Doctor of Philosophy
POD     Postoperative Day
POSSUM  Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity
PRISMA  Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM    Patient Reported Outcome Measures
R0      Negative resection margins
R1      Microscopically positive margins
RnYGJ   Roux-en-Y Gastro-Jejunostomy
RnYEJ   Roux-en-Y Esophago-Jejunostomy
SD      Standard Deviation
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPLINTS</td>
<td>Scrub Practitioners’ List of Intraoperative Non-Technical Skills</td>
</tr>
<tr>
<td>SST</td>
<td>Surgical Safety Technologies</td>
</tr>
<tr>
<td>STMB</td>
<td>Short-Term Morbidity</td>
</tr>
<tr>
<td>STMT</td>
<td>Short-Term Mortality</td>
</tr>
<tr>
<td>VFA</td>
<td>Visceral Fat Areas</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1 – PRISMA algorithm (pg. 45)
Figure 2 – Non-Technical Performances for Surgeon Group (pg. 114)
Figure 3 – Non-Technical Performance for Fellow Group (pg. 115)
Figure 4 – Bigrams for Surgeon Group (pg. 116)
Figure 5 – Bigrams for Fellow Group (pg. 117)
Figure 6 – Synchronized surgical timeline produced by Operating Room Black Box (pg. 133)
Figure 7 – Opportunities with the ORBB (pg. 145)
List of Tables

Table 1 – Objective Structured Assessment of Technical Skills (pg. 7)
Table 2 – Generic Error Rating Tool (pg. 11)
Table 3 – Non-Technical Skills for Surgeons (pg. 16)
Table 4 – Scrub Practitioners’ List of Intraoperative Non-Technical Skills (pg. 19)
Table 5 – Validity Evidence based on Messick’s Framework (pg. 25)
Table 6 – Classification of Surgical Complications (pg. 30)
Table 7 – ASA-Physical Status Classification System (pg. 33)
Table 8 – Weighted index of comorbidity (pg. 35)
Table 9 – Overview of the studies (pg. 46)
Table 10 – Technical performance and patient outcomes assessed (pg. 49)
Table 11 – MERSQI overall and per category score for each study (pg. 58)
Table 12 – Sources of evidence contained within the studies, based on the Messick’s framework of validity (pg. 59)
Table 13 – Patient Factors (pg. 77)
Table 14 – Surgical Factors (pg. 79)
Table 15 – Postoperative Complications (pg. 81)
Table 16 – Predictors of Postoperative Complications (pg. 83)
Table 17 – Junior Residents and Staff responses on case sharing (pg. 96)
Table 18 – Senior Residents and Staff responses on case sharing (pg. 97)
Table 19 – Fellows and Staff responses on case sharing (pg. 98)
Table 20 – Spearman's Correlation Matrix for Surgeon and Fellow Groups (pg. 113)
Table 21 – Added Validity Evidence by this Thesis (pg. 136)
This thesis follows a paper format.

Chapter 1: Introduction

This chapter provides a general, broader introduction on surgical performance and patient outcomes, and then gives a thorough literature review specifically on how technical performance affects patient outcomes. By summarizing this evidence, it also identifies gaps in knowledge and highlights potential future research opportunities, some of them being addressed by the projects included in this PhD thesis. Throughout the thesis, the word “performance” is used interchangeably with “skills”.

The introduction has three main sections. The first section focuses on surgical performance, with the emphasis on both technical and non-technical performances, their definitions, importance and methods of assessment. The second part describes patient outcomes, definitions, and classifications. Also, just as important is to describe what are the main factors and variables shown to be associated with adverse patient outcomes. The final section of the introduction summarizes the current evidence in the literature on how technical performance affects patient outcomes. This is a systematic review and was published as an article, entitled “The Effect of Technical Performance on Patient Outcomes in Surgery: A Systematic Review” by Fecso AB, Szasz P, Kerezov G and Grantcharov TP, Annals of Surgery 2017 Mar; 265(3): 492-501. Reprinted with kind permission from Wolters Kluwer Health Inc. Formatting has been adapted for the purpose of this thesis.

Chapter 2: Aims, Objectives and Hypotheses

The purpose of this chapter is to describe the rationale for the overall thesis in the form of purpose statement, to provide the main and secondary research hypotheses, and to list the specific aims and objectives for each of the studies included.
Chapter 3: Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery

The study in this chapter provides further evidence on the effects of technical performance on patient outcomes. The methodology used in this study addresses the main limitations that have been identified in the systematic review, discussed in Chapter 1, namely using an objective, trained rater to review full length, unedited video files for performance assessments. This study has been submitted for publication with the title “Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery” by Fecso AB, Bhatti JA, Stotland PK, Quereshy FA and Grantcharov TP.

Chapter 4: Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges, and Potential Opportunities at a Large University Centre

Although the focus is mainly on the performance itself and not the level of the operating physician, and since the previous study assessed de-identified videos, the purpose of the study in this chapter was to explore the current educational environment in the operating rooms at a large university center, to determine how is an operation, in this case a laparoscopic gastrectomy for cancer distributed between operators (Staff vs. Fellows vs. Residents). This study was published as an article, entitled “Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges, and Potential Opportunities at a Large University Centre” by Fecso AB, Bonrath EM and Grantcharov TP, Journal of Surgical Education 2016 Jul-Aug; 73(4): 749-55. Reprinted with kind permission from Elsevier. Formatting has been adapted for the purpose of this thesis.

Chapter 5: The Relationship Between Intraoperative Non-Technical Performances and Technical Events in Bariatric Surgery

Chapter 1 introduced the discussion regarding how surgeons work closely in the operating room with a number of other professionals, to achieve the common goal, which is the success of the operation. Also initiated, that non-technical performances, such as teamwork, communication, etc. affects patient outcomes as well, however, the interaction between technical and non-technical performances remains under-investigated. This chapter aimed to investigate these interactions and demonstrate the relationship between the two. The study in this chapter has been
accepted for publication in British Journal of Surgery (BJS-1409-Oct-17) with the title “The Relationship Between Intraoperative Non-Technical Performances and Technical Events in Bariatric Surgery” by Fecso AB, Kuzulugil SS, Babaoglu C, Bener AB and Grantcharov TP. Formatting has been adapted for the purpose of this thesis.

Chapter 6 – General Discussion

This chapter summarizes the main findings of the studies presented in this PhD thesis as well as gives an overall summary of the thesis. This chapter ties the findings together, and with the added new evidence narrows the knowledge gaps identified in Chapter 1. This general discussion is in addition to the study specific discussions detailed in the previous chapters.

Chapter 7 – Limitations

Although the studies in this thesis added significantly to the current literature and existing knowledge base, the presented results need to be interpreted while keeping their limitations in mind. Specific limitations of each individual study have been discussed in the relevant chapters; therefore in this chapter the general limitations of the thesis as a whole are discussed.

Chapter 8 – Future Directions

The studies in this thesis not only added to the current evidence, but also provided opportunities to generate future research hypotheses. This chapter highlights the possibilities for future projects targeting those opportunities, including the limitations presented in the previous chapter.
CHAPTER 1

The purpose of this chapter is to:

I. Describe surgical (technical and non-technical) performance and their assessment methods and instruments

II. Describe patient outcomes and classify their contributing factors

III. Review the current knowledge in respect to the effect of technical performance or skills on patient outcomes

IV. Summarize the gaps in knowledge and provide future directions
1 Introduction

Worldwide, the estimated number of major surgical procedures performed in 2004 was over 234 million, which translates into approximately one operation for every 25 human beings (Weiser et al., 2008). Looking at the trend, these numbers are to increase even more, due to gains in life expectancy and the epidemiological transition that has accompanied industrialization (Omran, 2005; Weiser et al., 2008). Surgical interventions are not without risks and the reported rates of major perioperative complications are between 3 and 16% in developed countries with permanent disability or death of about 0.4 – 0.8 % (A. A. Gawande, Thomas, Zinner, & Brennan, 1999; Kable, Gibberd, & Spigelman, 2002; Weiser et al., 2008). Derived from these impressive numbers, potentially 7 to 37 million major complications or adverse events occur annually worldwide.

When caring for a surgical patient, it is important to acknowledge that the surgical care and management doesn’t start or end with the operation itself. It starts well in advance, especially in the case of operations. Even for emergency operations, the management and surgical care starts in the emergency department or hospital wards. Complications can occur at anytime and at any stage of the patient’s journey through the healthcare system, from the very beginning of the initial admission until the end of the final discharge or even post discharge. However, approximately 39 to 54 % of these complications occur in the operating room and nearly one third of them are potentially avoidable (de Vries, Ramrattan, Smorenburgh, Gouma, & Boermeester, 2008; A. A. Gawande et al., 1999; Kable et al., 2002; Zegers et al., 2009).

The operating room is a high-stake, high-risk environment (Kohn, Corrigan, & Donaldson, 2000). In this unique environment, professionals work together in a closely coordinated fashion to achieve a common goal, which is the successful surgery for the patient (D’Addessi, Bongiovanni, Volpe, Pinto, & Bassi, 2009). In this environment the surgeon is one of the participants and without a doubt their competence and performance is imperative in ensuring a successful outcome. According to the Accreditation Council for Graduate Medical Education (ACGME), in general surgery, surgical performance is one of the components of competency along with patient care, medical knowledge, practice-based learning and improvement,
interpersonal skills, professionalism and finally system-based practice ("ACGME Program Requirements for Graduate Medical Education in General Surgery," 2016). Similarly, in the CanMEDS 2015 Physician Competency Framework as Medical Experts, physicians integrate all of the six intrinsic Roles (e.g. Communicator, Collaborator, Leader, Health Advocate, Scholar and Professional), applying knowledge, skills and professional values in their provision of high-quality and safe patient-centered care (Frank JR, Snell L, Sherbino J, & editors., 2015).

Therefore, for a surgeon to be competent has to be proficient in all of the aforementioned components (Satava, Gallagher, & Pellegrini, 2003).

Although a mix of the above-mentioned competencies and skills of the surgeon are required from the very beginning of the care, in the non-operative settings as well, the operating room (OR) represents the main area of activity for surgeons. It has been estimated that cognitive skills outweigh the technical ones during the performance of a surgical procedure (Satava et al., 2003), however technical skills remain an important component since, most of the errors made in the operating room are technical in nature (A. A. Gawande et al., 1999; A. A. Gawande, Zinner, Studdert, & Brennan, 2003; Rogers et al., 2006) and technical performance might contribute to patient outcomes more than the perioperative management the patient receives (Birkmeyer et al., 2013).

For the purpose of this thesis, out of the above-mentioned competencies, technical and non-technical skills and performances are discussed in more details.
1.1 Technical performance

1.1.1 Definitions

Technical skills refer to the manual dexterity of the physician and according to ACGME it is the ability to “…competently perform all medical, diagnostic and surgical procedures considered essential for the area of practice” ("ACGME Program Requirements for Graduate Medical Education in General Surgery," 2016). As Medical Experts, physicians perform procedures in a skillful and safe manner, adapting to unanticipated findings or changing clinical circumstances (Frank JR et al., 2015).

Technical errors are defined as the “…failure of planned actions to achieve their desired goal” (Reason, 1995) (e.g.: bowel grasper slips off of the intestine).

Technical events are defined, based on the World Health Organization guidelines, as “any deviation from usual medical care that causes an injury to the patient or poses a risk or harm” ("WHO Draft Guidelines for Adverse Event Reporting and Learning Systems.," 2005) (e.g: the slipped bowel grasper causes a serosal injury to the intestine).

Rectification of an event, or event compensation is defined as additional measures or actions required to avoid an adverse outcome (Barach et al., 2008).

1.1.2 Assessment of technical performance

The main application of technical skills assessment is in the area of surgical education and residency training programs and the purpose of testing is multifactorial, to provide constructive feedback, to use for trainee promotion and potentially to identify deficiencies in the training programs (Martin et al., 1997). However, the application of technical skills assessment in the field of patient safety and outcome research is under utilized, although human errors and adverse events have been the topic of discussion and research for many years (Bonrath, Zevin, Dedy, & Grantcharov, 2013).

Technical skills can be acquired and performance can be assessed in a variety of settings: (1) operating room (Martin et al., 1997), (2) animal laboratory ( Christopherson, Buchsbaum, Voet, & Lifshitz, 1986; Lossing, Hatswell, Gilas, Reznick, & Smith, 1992; Saifi et al., 1990) or (3)
bench models, simulators and virtual reality (Fried et al., 2004; Marescaux et al., 1998; Martin et al., 1997).

In this PhD thesis, all performances have been assessed in the operating room during real operations. Therefore, the next section discusses assessment methodologies designed to evaluate surgical technical performance in this environment, i.e. observational tools. The most widely used and well known observational tools are introduced, and for those used in this thesis, the tool itself is also presented in a table format. A number of device-based performance measures have been and are continuously being developed (e.g.: motion tracking, grasp force measurements, wearable devices, etc.) but since they have not found widespread implementation yet, they will not be discussed here.

1.1.2.1 Global Assessment Scales

Global assessment scales, tools or instruments are designed to evaluate global and generic technical skills. These scales are easy to use and the same tool can be used to assess a wide variety of procedures (Ahmed, Miskovic, Darzi, Athanasiou, & Hanna, 2011).

1.1.2.1.1 Objective Structured Assessment of Technical Skills

One of the first, well-established global rating scales was the Objective Structured Assessment of Technical Skills (OSATS) instrument, developed at the University of Toronto, and contained a global and a task-specific scoring component (Martin et al., 1997). The global scale has been used in both simulated and real operative procedures, either as live observations or review of video recordings (Aggarwal et al., 2007; Goff, Lentz, Lee, Houmard, & Mandel, 2000; Niitsu et al., 2013). The global rating scale of OSATS measures seven relevant skills on a 5-point Likert scale (Martin et al., 1997). The seven skills components measure a combination of skills, required for the “Respect for tissue”, “Time and motion”, “Instrument handling”, “Knowledge of instruments”, “Knowledge of specific procedure”, “Use of assistants” and “Flow of operation and forward planning”. The scores given ranges between the minimum of 5 and the maximum of 35. The global scale component of OSATS was widely adopted in a variety of surgical disciplines and assessment formats, both for open and minimally invasive procedures (Alici, Buerkle, & Tempfer, 2014; Birkmeyer et al., 2013; Noordin & Allana, 2015; Tsagkataki & Choudhary, 2013). Just as important, from a feasibility and practicality point of view, is the fact
that based on the inter-rater reliability in the original study (Martin et al., 1997), one rater is adequate for the assessment.

Table 1 presents the OSATS instrument.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Respect for tissue</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Careful handling of tissue but occasionally caused inadvertent damage.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistently handled tissues appropriately with minimal damage.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time and motion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Many unnecessary moves.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Efficient time/motion but some unnecessary moves.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Economy of movement and maximum efficiency.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Instrument handling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Repeatedly makes tentative or awkward moves with instruments.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Competent use of instruments although occasionally appeared stiff or awkward.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluid moves with instruments and no awkwardness.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge of instruments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently asked for the wrong instrument or used an inappropriate instrument.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Knew the name of most instruments and used appropriate instrument for the task.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obviously familiar with the instruments required and their names.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Use of assistance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistently placed assistants poorly or failed to use assistants.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Good use of assistants most of the time.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategically used assistant to the best advantage at all times.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flow of operation and forward planning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently stopped operating or needed to discuss next move.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Demonstrated ability for forward planning with steady progression of operative procedure.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obviously planned course of operation with effortless flow from one move to the next.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Knowledge of specific procedure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deficient knowledge. Needed specific instruction at most operative steps.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Knew all important aspects of the operation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstrated familiarity with all aspects of the operation.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Overall, on this task should this candidate:**  ☐ Pass  ☐ Fail?
1.1.2.1.2  Global Operative Assessment of Laparoscopic Skills

Almost a decade later, another global rating scale was developed at the University of Montreal, and that is the Global Assessment of Laparoscopic Skills (GOALS) instrument (Vassiliou et al., 2005). Similarly, to OSATS, GOALS also had a global component, a task-specific component and two additional visual analogue scales. The global component was developed based on OSATS and the purpose was to assess laparoscopic surgical performance. There are five categories in the global scale such as “Depth perception”, “Bimanual dexterity”, “Efficiency”, “Tissue handling” and finally “Autonomy”. The assigned score can vary from a minimum of 5 to a maximum of 25 (Vassiliou et al., 2005). The GOALS instrument has been also studied and found to be valid and reliable in a variety of laparoscopic procedures, such as cholecystectomy (Vassiliou et al., 2005), appendectomy (Gumbs, Hogle, & Fowler, 2007) or advanced foregut surgery (Hogle, Liu, Ogden, & Fowler, 2014).

1.1.2.1.3  Direct Observation of Procedural Skills

Developed in the UK, as its name gives it away, the Direct Observation of Procedural Skills (DOPS) rating tool aims to assess, by direct observation, the procedural skills of surgical, medical or general practice trainees at all levels (Khan MAA, Gorman M, Gwozdziewicz L, Sobani ZA, & C., 2013; Wilkinson et al., 2008). This tool is extensively used in the UK, but little is published around it (Ahmed et al., 2011). There is a generic as well as a procedure-specific DOPS assessment available. Ten categories are assessed with DOPS including perioperative preparations, technical and non-technical abilities as well (Khan MAA et al., 2013). Although it has been shown that the instrument is valid, the tool is reliable only if at least three observers assess at least two procedures each (Ahmed et al., 2011). This significantly affects its generalizability, practicability and feasibility.

1.1.2.1.4  The Ottawa Surgical Competency Operating Room Evaluation

The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) instrument was developed by an expert group of medical educators at the University of Ottawa and published in 2012. This tool was designed to be generalizable to any surgical procedure for the assessment of
surgical competence using direct OR supervision and observation (Gofton, Dudek, Wood, Balaa, & Hamstra, 2012).

The O-SCORE includes 11 items (8 items rated on a 5-point competency scale, 1 yes/no question about competency to perform the procedure independently, and 2 open-ended questions for feedback). The score received ranges from a minimum of 8 to a maximum of 40. The rated items include pre-procedure, intra- and post-procedure ratings. The O-SCORE instrument has a component to determine the relative complexity of the assessed procedure as well (Gofton et al., 2012).

There is a growing body of evidence to support the validity of the tool (MacEwan, Dudek, Wood, & Gofton, 2016), and it is the gold standard instrument at the University of Ottawa to assess surgical trainees.

1.1.2.2 Task – Specific Assessment Scales

Task-specific assessment tools are characterized by a breakdown of a procedure into tasks and these tasks are assessed individually (Ahmed et al., 2011). These instruments can be used as stand-alone tools to assess performance or can be added to global rating scales, such as OSATS or GOALS (Martin et al., 1997; Vassiliou et al., 2005). These tools are often time checklists and in general, use binary ratings, where the observer assesses whether a performance was observed or not. Others can be more sophisticated, using a defined error scoring systems (Ahmed et al., 2011). Furthermore, others are a combination of a checklist with an error scoring system. Sarker et al. (Sarker, Chang, Vincent, & Darzi, 2005) designed an instrument for laparoscopic cholecystectomy that combines task analysis with a systematic assessment of errors. The scoring system developed by Eubanks et al. (Eubanks et al., 1999) represents another well-established task-specific checklist that combines general procedure aspects with error metrics, also in the area of laparoscopic cholecystectomy. Both of these instruments had been reported to be reliable when expert raters applied the tools in a research environment (Eubanks et al., 1999; Sarker et al., 2005). Other, procedure-specific checklists have been reported to be feasible, valid and reliable in a variety of surgical specialties and sub-specialties (Ahmed et al., 2011). These include gastrointestinal endoscopy (Vassiliou et al., 2010), general surgery (Lipman et al., 2010; Van Sickle et al., 2008), thoracic surgery (Chapman, Marx, Honigman, Rosen, & Cavanaugh, 1994) and vascular surgery (Wilasrusmee, Phromsopa, Lertsitichai, & Kittur, 2007).
1.1.2.2.1  Generic Error Rating Tool

As mentioned above some task-specific assessment instruments use error analysis to evaluate surgical technical performance. One good example is the Generic Error Rating Tool or GERT, developed in 2013 by Bonrath et al. (Bonrath, Zevin, et al., 2013). Although it is a generic error rating tool, and not specific to any procedure, it is task-specific in sense that assesses the manual task that leads to an error (e.g.: a slip error made with a bowel grasper).

Four modified error modes are identified in the instrument: (1) “Too much force or distance”, (2) “Too little force or distance”, (3) “Wrong orientation” and (4) “Inadequate visualization”. Nine generic surgical tasks are defined, during which an error could be encountered: (1) “Abdominal access”; (2) “Use of retractors”; (3) “Use of energy devices”; (4) “Grasping and dissection”; (5) “Cutting, transection and stapling”; (6) “Clipping”, (7) “Suturing”; (8) “Use of suction” and (9) “Other unclassified”. In the GERT instrument, these nine surgical task groups are combined with the four error modes in a checklist format (Bonrath, Zevin, et al., 2013).

Another feature of the GERT instrument is that not only describes the errors, but the technical events as well, should any of the errors lead to an event (e.g.: slipped bowel grasper causes a serosal injury). There is no minimum and maximum score assigned. The final score is the sum of the errors, events made throughout the procedure assessed. Validity and reliability evidence for quantitative and qualitative assessment of technical errors exists in surgery and gynecology (Bonrath, Zevin, et al., 2013; Husslein, Shirreff, Shore, Lefebvre, & Grantcharov, 2015).

Table 2 presents the GERT instrument.
Table 2 - Generic Error Rating Tool (Bonrath, Zevin, et al., 2013)

<table>
<thead>
<tr>
<th>Surgical task group</th>
<th>Video Code</th>
<th>Rater Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Error mode</td>
<td>Time of observation</td>
</tr>
<tr>
<td>Abdominal access</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Use of retractors</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Use of energy devices</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Grasping and dissection</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Cutting, transection and stapling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Clipping</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Suturing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tool little force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrong orientation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Use of suction</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Too much force/distance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Inadequate visualization</td>
<td></td>
</tr>
<tr>
<td>Other unclassified</td>
<td>Description/time:</td>
<td></td>
</tr>
</tbody>
</table>

1.1.2.3 Procedure - Specific Global Assessment Scales

Procedure-specific global assessment scales combine the details of task-specific checklists with the broader scale range of global ratings. The scales are limited for use in the procedure for which they were designed; however the information provided by these scales are more granular than global ratings (Bonrath, 2015). Numerous procedures have their specific global rating scales developed, such as laparoscopic Roux-en-Y gastric bypass (LRYGB) (Zevin et al., 2013), laparoscopic Nissen fundoplication (Dath et al., 2004), laparoscopic incisional hernia repair (Ghaderi et al., 2011), open sigmoid colectomy (Lipman et al., 2010) and other colorectal surgery (Dath et al., 2004; de Montbrun et al., 2013; Palter & Grantcharov, 2012b), to name only a few. Given that more detailed information is generated from the procedure-specific rating scales, in contrast to global ratings, there is a great value of these instruments in the general surgery training settings to provide structured formative feedback (Ahmed et al., 2011; Beard, Choksy, & Khan, 2007; Bonrath, 2015).

In summary, there are a variety of instruments developed to rate technical performance. Multiple studies demonstrated that global assessment scales, used by experts, are superior to checklists or task-specific assessments for the evaluation of technical performance (Regehr, MacRae, Reznick, & Szalay, 1998; Vassiliou et al., 2005). Checklists require adherence to a particular order of steps or tasks, which makes the assessments rigid (Beard et al., 2007; Zevin et al., 2013). Furthermore, there is a limited discrimination at a senior trainee or specialist level, due to a potential ceiling effect, which is reached once the procedure has been learned (Ahmed et al., 2011; Beard et al., 2007). However, the combination of these two tools has been shown to provide an effective way of evaluation of technical performance (Naik, Perlas, Chandra, Chung, & Chan, 2007). Therefore, in this thesis, for the assessment of technical performance, both OSATS and GERT instruments were used.
1.2 Non-Technical Performance

The operating room is a unique environment, from a team dynamics point of view, as professionals from multiple disciplines are required to work together in a close, coordinated fashion to achieve a common goal, that is the successful surgery for the patient (D'Addessi et al., 2009; Flin & Mitchell, 2009). The atmosphere in this environment largely depends on human behavior interrelations and communication (D'Addessi et al., 2009).

Non-technical skills, or also known as “intrinsic” skills, of the operating room personnel are essential in avoidance, trapping and mitigating errors and events to potentially prevent more serious harm to the patient (Carthey, de Leval, Wright, Farewell, & Reason, 2003; Catchpole et al., 2007). Human factors, such as communication breakdown, excessive workload and fatigue have been identified as potential contributors to surgical errors and complications (A. A. Gawande et al., 2003; Greenberg et al., 2007). Lingard et al. (Lingard et al., 2004) have identified communication failures in over 30% of observed interactions within the OR. Mazzocco et al. (Mazzocco et al., 2009) found a correlation between poor teamwork and perioperative patient injury.

The non-technical skills of the surgeon in the operating room, e.g.: communication, teamwork, situational awareness, and leadership to name only a few, are just as important as the technical skills, however, the interaction between the two is poorly understood and under investigated. Although the main focus of this thesis is on intraoperative technical performance, study #4 (Aim #4) analyses the interaction between intraoperative technical and non-technical performances, therefore in the next section of the introduction, methods of assessments of non-technical or intrinsic skills are presented.

1.2.1 Definitions

The term “nontechnical skills” encompasses a variety of knowledge, skills, behaviors and attitudes (Dedy, 2015). In the healthcare, according to Professor Rhona Flin, of University of Aberdeen, non-technical skills are “the cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient performance” (Flin, O'Connor, & Crichton, 2008). Similarly, Dr. Morris Gordon from the UK defines non-technical skills, based on a Delphi study, as “a set of social and cognitive skills that support high quality, safe,
effective and efficient interprofessional care within the complex healthcare system” (Gordon, Baker, Catchpole, Darbyshire, & Schocken, 2015).

Based on extensive work in aviation and other high-risk industries, non-technical skills can be grouped into (1) cognitive (i.e. situation awareness and decision-making), (2) social (i.e. communication, teamwork and leadership) and personal resource skills (i.e. managing stress and coping with fatigue) (Flin et al., 2008).

1.2.2 Assessment of non-technical performance

The assessment of non-technical skills or performance was established in the aviation industry and contained the observation and assessment of the flight crews’ behaviors in the cockpit (Helmreich, Merritt, & Wilhelm, 1999). Frameworks of non-technical skills formed the basis of these assessments, while behavioral markers (i.e. negative or positive behaviors) pertaining to the skill item, facilitated the rating (Kanki, Helmreich, & Anca, 2010).

Today in medicine, the majorities of rating tools used are based on work on human factors in the aviation industry, and follow a basic framework comprising a hierarchy of categories, elements and observable behaviors (Klampfer et al., 2001). At the top of the hierarchy are the categories of different areas of skills. Each category is subdivided into a number of skill elements that conceptually belong to the respective category, but represent distinct performances within (Flin et al., 2003). Finally, the elements are described by a number of behavioral markers, which represent observable behaviors (positive or negative) that enhance or impede relevant aspects of the performance (Klampfer et al., 2001). The number and the type of categories, elements and behavioral markers can vary and depends on the conceptualization and the design of the individual rating system (Dedy, 2015). Moreover, there are rating systems developed to assess the performance of individuals separately or teams as a whole.

The most commonly used and reported instruments (tools, scales) and taxonomies of non-technical or intrinsic skills in the operating room environment are presented in the next section, focusing mainly on those used in this thesis.
1.2.2.1 Individual assessment scales

1.2.2.1.1 Non-Technical Skills for Surgeons

Non-technical skills for surgeons or NOTSS is a rating framework, designed to assess individual surgeons’ non-technical performance in the operating room environment (Yule, Flin, Paterson-Brown, Maran, & Rowley, 2006). Developed in 2006 at the University of Aberdeen, Scotland, NOTSS is geared specifically to practicing surgeons, although it has been adapted to trainees as well (Yule et al., 2006). The initial rating tool (version 1.1) comprised of 5 categories (situation awareness; decision making; task management; leadership; communication and teamwork), 14 skill elements and 2 behavioral markers indicative of good and poor performance, relevant to the surgical domain (Yule et al., 2006). The rating scale was designed as a 4-point Likert categorical scale with descriptions: “1 poor”, “2 marginal”, “3 acceptable”, “4 good”. An NO (not observed) or N/A (not applicable) rating label was also added to account for situations where a particular element or category was not required in a given situation (Yule et al., 2006). All categories are ranked on this 4-point scale to a maximum score of 16. Following a reliability study, the category of Task Management was dropped and relevant behaviors were incorporated into the other categories (Yule et al., 2008). Therefore, the revised NOTSS taxonomy (version 1.2) comprised of 4 categories, 12 elements and 2 behavioral markers. NOTSS has been validated in multiple settings and it’s being widely used to provide feedback for both trainees and surgeons (Beard, Marriott, Purdie, & Crossley, 2011; Crossley, Marriott, Purdie, & Beard, 2011; Yule et al., 2008).

Table 3 presents the NOTSS instrument.
Table 3 - Non-Technical Skills for Surgeons (Yule et al., 2008)

<table>
<thead>
<tr>
<th>Category</th>
<th>Category rating*</th>
<th>Element</th>
<th>Element rating*</th>
<th>Feedback on performance and debriefing notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation Awareness</td>
<td></td>
<td>Gathering information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Understanding information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Projecting and anticipating future state</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decision Making</td>
<td></td>
<td>Considering options</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Selecting and communicating option</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Implementing and reviewing decisions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication and Teamwork</td>
<td></td>
<td>Exchanging information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Establishing a shared understanding</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Co-ordinating team activities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leadership</td>
<td></td>
<td>Setting and maintaining standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supporting others</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coping with pressure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 1 Poor; 2 Marginal; 3 Acceptable; 4 Good; N/A Not Applicable

1 Poor - Performance endangered or potentially endangered patient safety, serious remediation is required
2 Marginal - Performance indicated cause for concern, considerable improvement is needed
3 Acceptable - Performance was of a satisfactory standard but could be improved
4 Good - Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others
N/A - Not Applicable

1.2.2.1.2 Objective Structured Assessment of Non-Technical Skills

Developed by Dedy et al., the Objective Structured Assessment of Non-Technical Skills or OSANTS is a new tool designed for in-training assessment of residents’ intraoperative non-technical performance (Dedy et al., 2015). The rating instrument is composed of seven main categories, including: (1) situation awareness; (2) decision making; (3) teamwork; (4) communication; (5) leading and directing; (6) professionalism; (7) managing and coordinating. Each of the categories is ranked on a global 5-point Likert scale with specific descriptors for each item for a combined score of a maximum 35. OSANTS is specifically designed towards trainees; therefore the skills items were selected specifically to be observable in residents and to be relevant for in training assessment. Also, the tool was designed as a global rating scale with unique descriptive anchors, allowing for a criterion-referenced assessment, without the need to interpret behaviors (Dedy et al., 2015). A recent study, also by Dedy et al., demonstrated the applicability of OSANTS as a method of feedback for surgical trainees in the real operating room environment (Dedy, Bonrath, Ahmed, & Grantcharov, 2016). The use of the OSANTS instrument significantly improved the average non-technical scores in residents, from their baseline to post-intervention performance. Also, good reliability was demonstrated in different settings, which supports the application of the OSANTS instrument in research and education (Dedy, Bonrath, et al., 2016).

1.2.2.1.3 Scrub Practitioners’ List of Intraoperative Non-Technical Skills

The operating room is a shared work environment for many professionals with various backgrounds, such as surgeons, nurses and anesthetists. These professionals work together in a coherent manner towards a common goal, which is the safe surgery of the patient (D'Addessi et al., 2009; Flin & Mitchell, 2009). Scrub practitioners (operating room nurses, operating department practitioners or instrument technicians) are key members of the operating room team, with many responsibilities, e.g. obtaining and handing in surgical instruments and supplies, ensuring that all equipment used during an operation is accounted for at the end (Mitchell et al., 2012).

Similarly to NOTSS, the Scrub Practitioners’ List of Intraoperative Non-Technical Skills or SPLINTS instrument was developed at the University of Aberdeen. The scope of SPLINTS is to provide scrub practitioners a structured method for discussing, training and evaluating non-
technical skills that are required for a safe and effective intraoperative performance (Mitchell et al., 2013). The SPLINTS taxonomy comprises of three main categories (i.e. situation awareness, communication and teamwork, and task management). Each category contains 3 elements which are the main component skills underpinning each skill category (Mitchell et al., 2013). Behavioral markers provide examples of good or poor behaviors, to which the assessor may refer to as a guide, when making ratings at the category and element levels (Flin et al., 2008). The rating scale was designed as a 4-point Likert categorical scale with descriptions: “1 poor”, “2 marginal”, “3 acceptable”, “4 good”. An N/R (not required) rating label was also added to account for situations where a particular element or category was not required in a given clinical encounter (Mitchell et al., 2013). The reliability of SPLINTS system was tested in simulated, standardized, video scenarios and being trialed in the operating rooms of four Scottish hospitals (Mitchell et al., 2012).

Table 4 presents the SPLINTS instrument.
Table 4 - Scrub Practitioners' List of Intraoperative Non-Technical Skills (Mitchell et al., 2013)

<table>
<thead>
<tr>
<th>Category</th>
<th>Element</th>
<th>Element rating*</th>
<th>Feedback on performance and debriefing notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation Awareness</td>
<td>Gathering information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recognising and understanding information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anticipating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication and Teamwork</td>
<td>Acting assertively</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exchanging information</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Co-ordinating with others</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Management</td>
<td>Planning and preparing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Providing and maintaining standards</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coping with pressure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 1 Poor; 2 Marginal; 3 Acceptable; 4 Good; N/R Not Required

1 Poor - Performance was not acceptable and could potentially have endangered patient safety, remedial action is required.
2 Marginal - Performance indicated cause for concern, considerable improvement is needed
3 Acceptable - Performance was of a satisfactory standard but could be improved
4 Good - Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others
N/R - Not Required; skill was not observed because it was not required in this case

1.2.2.1.4 Anaesthetists’ Non-Technical Skills

As discussed above, anesthetists are also part of the operating room team; therefore they are too, actively participating in the intraoperative management of the patients. Similarly to NOTSS and SPLINTS, a non-technical rating instrument was also developed at the University of Aberdeen. The Anaesthetists’ Non-Technical Skills or ANTS behavioral marker system was designed to identify non-technical skills demonstrated by anesthetists in a day-to-day practice (Fletcher et al., 2004). The ANTS system comprises of three level hierarchy. At the highest level are the four categories, namely task management, team working, situation awareness and decision-making. Beneath of the categories are a total of 15 skill elements and each of them having a definition of a good and poor performance (behavior markers) (Fletcher et al., 2004). The rating scale is also similar to NOTSS and SPLINTS on a 4-point Likert scale and an NO (not observed) rating label.

Anesthesiologists were not part of the projects included in this thesis, however without their fair share, the intraoperative non-technical assessment is incomplete, therefore including their performance is one of the future directions discussed subsequently.

1.2.2.2 Team assessment scales

In the previous section non-technical behavioral systems were presented that were designed to assess the performances on an individual level. In this section, the two widely utilized instruments designed to assess the performances of the entire surgical team as one, will be introduced.

1.2.2.2.1 Non-Technical Skills

Initially developed for the aviation industry in the early 2000s, Non-Technical Skills or NOTECHS was created as a feasible and efficient instrument to assess pilots’ non-technical skills (Flin et al., 2003). This framework was then adapted in the healthcare system, as Oxford NOTECHS, to rate the non-technical skills of the operating room teams, namely the surgical-, anesthetic-, and nursing sub-teams (Mishra, Catchpole, & McCulloch, 2009). Oxford NOTECHS has the similar, three level hierarchy, as presented in the previous instruments. The categories are: (1) Leadership and management; (2) Teamwork and cooperation; (3) Problem-
solving and decision making; and finally (4) Situation awareness (Mishra et al., 2009). Each category has several specific elements, a total of 16. In contrast to NOTSS, behavior is only rated at the level of categories, while elements are not scored (Mishra et al., 2009). Each of the categories are rated on a 4-point Likert scale: 1 – below standard; 2 – basic standard; 3 – standard; and 4 – excellent. As the complexity of the surgical teamwork became clearer, more sophisticated measurement methods followed and the original Oxford NOTECHS got restructured to Oxford NOTECHS II (Robertson et al., 2014). The main difference between the two instruments is the use of an 8-point scale instead of a four-point scale, assigning all teams a baseline score of 6, from which subsequent observations could result in upward or downward deviations. The tool was shown to be reliable and suitable for use (Robertson et al., 2014).

1.2.2.2 Observational Teamwork Assessment for Surgery

The Observational Teamwork Assessment for Surgery or OTAS tool was developed at the Imperial College London and is aimed to measure team performance in the operating room (Healey, Undre, & Vincent, 2004). The OTAS rating tool is somewhat different from the previously presented instruments, in the sense that it measures the team performance in various key phases of the surgical procedure: in the pre-, intra-, and postoperative phases (Healey et al., 2004). Hull et al. (Hull, Arora, Kassab, Kneebone, & Sevdalis, 2011) refined the tool based on expert consensus by removal and modification of behavioral markers. There are five main categories, or behavioral dimensions, that make up the tool and they include: (1) Communication; (2) Coordination; (3) Cooperation and back up behavior; (4) Leadership and (5) Team monitoring and situational awareness (Healey et al., 2004). Each of the categories is assessed on a seven-point Likert scale, ranging from 0 to 6. The lowest score (0) indicates problematic behavior and hindered team function, while the highest score (6) indicates exemplary behavior and very highly effective team function (Hull et al., 2011). OTAS generates separate behavior scores for each of the five categories, across each of the three sub-teams and also across the three key operative phases (Healey et al., 2004). Several studies have been shown the validity of the tool in a variety of clinical settings around the world (Amaya Arias et al., 2014; Hull et al., 2011; Passauer-Baierl, Chiapponi, Bruns, & Weigl, 2014).

Since in the present thesis, the intraoperative environment was assessed only (excluding pre-, and postoperative phases) and the participating teams were not all inclusive (anesthesia team did not take part), the value of team assessment scales was limited.
In summary, just like for technical performance assessment, there are a variety of tools designed to rate non-technical or intrinsic performances in the operating room, as well. In this thesis, the NOTSS and SPLINTS instruments were used to rate the surgical and nursing non-technical performances. Although NOTSS and SPLINTS assess non-technical performance on an individual level (surgeon or scrub practitioner) and OTAS, NOTECHS on a team level, the interaction between individuals, i.e.: the scrub nurse and the surgeon, remains under investigated. Aim # 4 of this thesis is to investigate this relationship in more details and highlight some of the nuances of this collaboration, using a novel approach.
1.3 Validity of Performance Assessments

Assessment instruments are designed to measure a particular “construct” (e.g. performance, attitude, behavior or skill) and it is important to ensure that these measurements are valid (they measure what they were designed to measure) and reliable (the measurements are accurate) (Bonrath, 2015). In the previous sections, validity evidence for the tools used in this thesis has been presented. However, the framework surrounding the validity evidence has undergone a shift and in the current, state-of-the-art conceptual framework, proposed by Samuel Messick, validity is defined as the appropriate interpretation of the test results (Ghaderi et al., 2015; Messick, 1995). In such, validity applies to the scores or interpretation of the results in a specific context, and the term commonly used as “valid instrument” is imprecise (Cook & Beckman, 2006; Messick, 1989; Schmitz, 2006).

In the traditional framework, validity consisted of three types: content, criterion (including concurrent and predictive validity) and construct (Cronbach & Meehl, 1955; Downing, 2003; Ghaderi et al., 2015). In the contemporary, Messick’s unitary framework, construct is the only form of validity, with five sources to it (Downing, 2003). The five sources are (1) content, (2) response process, (3) internal structure, (4) relationship to other variables and (5) consequences of testing (Ghaderi et al., 2015). The definition of content is the “relationship between a test’s content and the construct it is intended to measure” (e.g.: test blueprint) (Beckman, Cook, & Mandrekar, 2005; Downing & R., 2009). The response process it entails analysis of responses and accuracy of scoring (e.g.: rater training) (Beckman et al., 2005; Downing & R., 2009; Ghaderi et al., 2015). Internal structure is usually referred to as “reliability” and includes reproducibility of the results (e.g.: generalizability) (Downing & R., 2009; Ghaderi et al., 2015). Relationship to other variables means the correlation between the instrument assessment scores and other variables or scores on other performance assessments (e.g.: correlation between scores and outcomes) (Beckman et al., 2005; Downing & R., 2009). Finally, the consequence of testing is a validity source more relevant to formative (feedback) or summative (pass/fail) assessments (e.g.: determining pass-fail score) (Ghaderi et al., 2015).

Although Messick’s framework represents the standard for test and tool design, it has not found wide adoption in the area of surgical education yet, and recent reviews demonstrated that the
majority of the studies are still reporting according to the older framework (Ahmed et al., 2011; Ghaderi et al., 2015).

For the purpose of this thesis and the projects included in this PhD work, all technical and non-technical assessments used Messick’s framework, as a guide. Table 5 shows the validity evidence to support the use of these rating tools, namely for OSATS, GERT, NOTSS and SPLINTS. The projects in this thesis aim to add to this validity evidence as well.
### Table 5 - Validity Evidence based on Messick's Framework

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Sources of validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sources of validity</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OSATS</strong></td>
<td>+ (Anastakis et al., 2003; Martin et al., 1997)</td>
</tr>
<tr>
<td></td>
<td>+ (R. Reznick, Regehr, MacRae, Martin, &amp; McCulloch, 1997; Shah, Munz, Manson, Moorthy, &amp; Darzi, 2006)</td>
</tr>
<tr>
<td></td>
<td>+ (Martin et al., 1997; R. Reznick et al., 1997; Shah et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>+ (Birkmeyer et al., 2013; Shah et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>+ (Anastakis et al., 2003; Chipman &amp; Schmitz, 2009; Shah et al., 2006)</td>
</tr>
<tr>
<td><strong>GERT</strong></td>
<td>+ (Bonrath, Zevin, et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>+ (Bonrath, Zevin, et al., 2013; Husslein et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>+ (Bonrath, Zevin, et al., 2013; Husslein et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>+ (Bonrath, Zevin, et al., 2013; Husslein et al., 2015)</td>
</tr>
<tr>
<td></td>
<td>+ (Bonrath, Zevin, et al., 2013; Husslein et al., 2015)</td>
</tr>
<tr>
<td><strong>NOTSS</strong></td>
<td>+ (Yule et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>+ (Crossley et al., 2011; J. Y. Lee, Mucksavage, Canales, McDougall, &amp; Lin, 2012; Yule et al., 2008; Yule et al., 2006; Yule et al., 2009)</td>
</tr>
<tr>
<td></td>
<td>+ (Crossley et al., 2011; Yule et al., 2008)</td>
</tr>
<tr>
<td></td>
<td>+ (Crossley et al., 2011; J. Y. Lee et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>+ (Yule et al., 2009)</td>
</tr>
<tr>
<td><strong>SPLINTS</strong></td>
<td>+ (Mitchell et al., 2013)</td>
</tr>
<tr>
<td></td>
<td>+ (Mitchell et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>+ (Mitchell et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>−</td>
</tr>
</tbody>
</table>

OSATS = Objective Structured Assessment of Technical Skills; GERT = Generic Error Rating Tool; NOTSS = Non-Technical Skills for Surgeons; SPLINTS = Scrub Practitioners' List of Intraoperative Non-Technical Skills
In summary, in the above sections, technical and non-technical performances in the operating room and the main instruments and rating tools used for their assessments were introduced. In this thesis, the technical performance of the operating surgeon is assessed using OSATS and GERT, while the non-technical or intrinsic performance of the operating room staff (surgical and nursing) is assessed by SPLINTS and NOTSS scales. All of these instruments have evidence in the literature to support their use.

Although, as previously stated, the main focus of this thesis is on how technical performance affects patient outcomes in surgery, even the most skillfully performed operation sometimes can result in complications and adverse outcomes, suggesting that this is a more complex subject with multiple potentially contributing variables. Therefore in the next sections patient outcomes, with the focus on complications, are introduced and the main variables that have been shown to affect them are discussed as well.
1.4 Patient Outcomes

At the beginning of 20th century, Dr. Ernst A. Codman, a surgeon and a champion in the field, recognized the value of patient outcomes and their measurements as an indicator of quality and proposed that all hospitals should produce a report to show the results of their treatments, to make comparisons amongst other hospitals and to assure treatment benefits are conferred (Codman, 1914; S. J. Lee, Earle, & Weeks, 2000).

1.4.1 Definitions

According to the Canadian Institute for Health Information (CIHI), patient outcomes or health outcomes are defined as changes in a health status of a person (patient) as a result of the interventions received within the healthcare system (CIHI, 2012). A result of any medical and surgical intervention can be divided into complicated or uncomplicated.

An adverse outcome or a complication is defined as

“... an unintended and unwanted event or state occurring during or following medical care, that is so harmful to a patient’s health that (adjustment of) treatment is required or that permanent damage results. The adverse outcome may be noted during hospitalization until 30 days after or transferal to another department. The intended result of treatment, the likelihood of the adverse outcome occurring, and the presence or absence of medical error causing it, are irrelevant in identifying an adverse outcome”

(de Mheen, van Duijn-Bakker, & Kievit, 2007; Kievit, Jeekel, & Sanders, 1999; Marang-van de Mheen, van Hanegem, & Kievit, 2005).

Adverse outcomes or complications are relevant to both the physician and the patient. They are relevant for the physician, because they may decrease the quality of care and for the patient because they might increase the burden of the disease (de Mheen et al., 2007).

1.4.2 Adverse outcomes in surgery

For many years, mortality (death of the patient) has been the exclusive outcome measure to assess the risk of the majority of surgical operations (Clavien, Sanabria, & Strasberg, 1992; Pearse et al., 2012; Vonlanthen & Clavien, 2012). With the dramatic decrease in mortality after
major operations (M. F. Brennan, Radzyner, & Rubin, 2009), the focus has shifted towards other endpoints, such as morbidity, quality of life and cost (Schneider, 2002; Slankamenac, Graf, Barkun, Puhan, & Clavien, 2013). The consequence of a complication, also known as morbidity, has emerged as one of the key parameters to measure the harmful outcome in most studies evaluating invasive interventions, such as surgery (Slankamenac et al., 2013). Furthermore, besides morbidity and mortality as “hard endpoints” of outcome variables, more and more emphasis is being placed on patient satisfaction and is increasingly considered to be an important indicator of quality of care (Delbanco, 2001; Jenkinson, Coulter, Bruster, Richards, & Chandola, 2002). Patient reported outcome measures (PROM) (e.g. functional status; health related quality of life) are being recognized as equally important as morbidity and mortality (Soreide & Soreide, 2013).

Over the last decades, as mentioned above, a steady decrease in morbidity and mortality was noted, however surgical outcomes and their optimization are still in increased focus, not only for humanitarian reasons but also as fundamentals and potential sources to decrease healthcare costs (Kehlet & Jørgensen, 2016; Kehlet & Mythen, 2011). Since Dr. Codman’s contribution, the field evolved immensely over the next century and in 2008, the Institute for Healthcare Improvement (IHI) developed a framework that describes an approach to optimize healthcare outcomes, “The IHI Triple Aim”. The Triple Aim includes (1) the improvement of patient experience and care, (2) the improvement of the health of the population and (3) the reduction of the cost of the healthcare (IHI, 2008). This thesis aims to target the first aim of the IHI, namely to improve patient care by identifying variables that potentially contribute to adverse outcomes and potentially are modifiable.

1.4.3 Classification of adverse outcomes

Every publication on results of a particular surgery reports complications, because their number and severity are important measures not only for that particular intervention, but also to compare results in individual centers and among centers. However, the lack of exact definition, uniform reporting and classification systems of surgical complications has hindered proper interpretation and comparison of surgical data and literature for a long time (Clavien et al., 1992).

The first severity grading system for postoperative complications were developed in 1992 in Toronto, by Clavien et al. (Clavien et al., 1992). The core concept was to objectively grade the
complications, by the treatments they provoked or whether they resulted in patient mortality (Strasberg & Hall, 2011). The system was revised multiple times leading to the Clavien-Dindo classification in 2004 (Dindo, Demartines, & Clavien, 2004) and more recently, in 2013, to the Comprehensive Complication Index (Slankamenac et al., 2013). The main difference between the Clavien-Dindo (CD) classification and the Comprehensive Complication Index is that CD is based on an ordinal scale and selects the single most severe complication after a procedure, while the Comprehensive Complication Index incorporates all postoperative complications and assigns particular weights to them. These weights (“operation risk index”) are mathematical calculations to reflect the perspectives of both the patients and physicians (Slankamenac et al., 2013).

Several other systems were developed based on CD, such as the Memorial Sloan Kettering (MSKCC) severity grading system and the Accordion severity grading system for surgical complications (Strasberg, Linehan, & Hawkins, 2009), however, the most extensively used is the Clavien-Dindo classification. There is strong evidence that this classification is valid and applicable worldwide in a variety of surgical specialties (Clavien et al., 2009) and this is the system that was used in this thesis, as well.

The CD classification system has five grades and the therapy that is needed to correct a specific complication is the cornerstone in ranking them (Dindo et al., 2004). In this classification system, grades I and II represent minor complications or deviation from the normal postoperative course, while grade III and above represent major complications, requiring invasive interventions and intensive therapies (Casadei et al., 2011; Dindo et al., 2004).

Table 6 presents the Clavien-Dindo classification system.
### Table 6 - Classification of Surgical Complications (Clavien Dindo system)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade I</td>
<td>Any deviation from the normal postoperative course without the need for pharmacological treatment or surgical, endoscopic, and radiological interventions. Allowed therapeutic regimens are: drugs as antiemetics, antipyretics, analgetics, diuretics, electrolytes, and physiotherapy. This grade also includes wound infections opened at the bedside.</td>
</tr>
<tr>
<td>Grade II</td>
<td>Requiring pharmacological treatment with drugs other than such allowed for grade I complications. Blood transfusions and total parenteral nutrition are also included.</td>
</tr>
<tr>
<td>Grade III</td>
<td>Requiring surgical, endoscopic or radiological intervention</td>
</tr>
<tr>
<td>Grade IIIa</td>
<td>Intervention not under general anesthesia</td>
</tr>
<tr>
<td>Grade IIIb</td>
<td>Intervention under general anesthesia</td>
</tr>
<tr>
<td>Grade IV</td>
<td>Life-threatening complication (including CNS complications)* requiring IC/ICU management</td>
</tr>
<tr>
<td>Grade IVa</td>
<td>Single organ dysfunction (including dialysis)</td>
</tr>
<tr>
<td>Grade IVb</td>
<td>Multiorgan dysfunction</td>
</tr>
<tr>
<td>Grade V</td>
<td>Death of a patient</td>
</tr>
<tr>
<td>Suffix &quot;d&quot;</td>
<td>If the patient suffers from a complication at the time of discharge, the suffix “d” (for “disability”) is added to the respective grade of complication. This label indicates the need for a follow-up to fully evaluate the complication.</td>
</tr>
</tbody>
</table>

*Brain haemorrhage, ischemic stroke, subarachnoidal bleeding, but excluding transient ischemic attacks. CNS, central nervous system; IC, intermediate care; ICU, intensive care unit.

1.4.4 Contributing factors to adverse outcomes

The incidence of overall complications in surgery can be up to 4-5 times greater than in general medicine (Baines et al., 2013; T. A. Brennan et al., 2004; de Vries et al., 2008). Multiple variables can potentially contribute to adverse outcomes, therefore in this section the main categories are discussed.

1.4.4.1 Patient factors

A large body of evidence exists in the literature that demonstrates the relationship between patient factors and adverse outcomes. Undeniably, regardless of surgical subspecialty, certain patient characteristics will determine the outcomes of the surgery, and even the most skillfully executed operation sometimes will result in complications. A recent systematic review by Visser et al. (Visser, Geboers, Gouma, Goslings, & Ubbink, 2015) summarized a number of patient factors that have been shown to predict outcomes the most. Naturally, this is not an exclusive list given the fact that only general-, gastrointestinal-, and vascular surgery specialties were included, however it gives a detailed overview of some of the most commonly used and studied patient variables. These are age, sex, body mass index (BMI), preoperative laboratory values, functional status, past medical and social history as well as preoperative medications of the patient (Visser et al., 2015).

To account and grade for a variety of these patient variables, several scoring systems were developed and subsequently modified, e.g. the American Society of Anesthesiologists (ASA) classification (Saklad, 1941), the Acute Physiology and Chronic Health Evaluation (APACHE) system (Knaus, Zimmerman, Wagner, Draper, & Lawrence, 1981), the Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM) scoring system (Copeland, Jones, & Walters, 1991) or the Charlson Comorbidity Index (CCI) (M. Charlson, Szatrowski, Peterson, & Gold, 1994), to name only a few. In this thesis the ASA and the CCI classification systems were used.
### 1.4.4.1.1 American Society of Anesthesiologists classification

One of the most widely used systems is the American Society of Anesthesiologists or ASA classification. Developed in 1941 by the American Society of Anesthetists, Inc., by Drs. E. A. Rovenstine, I. B. Taylor and M. Saklad, the ASA score meant to classify patients with reference to their “Physical State”, to their ability to withstand surgery (Saklad, 1941). Over the years, ASA underwent modifications and in the most current version it’s called the ASA-Physical Status Classification System and contains six classes with definitions and examples to increase the appropriateness of the assignments among both anesthesiologists and other clinicians (Hurwitz et al., 2017).

Table 7 presents the current ASA-Physical Status Classification System.
<table>
<thead>
<tr>
<th>ASA-Physical Status Class</th>
<th>Definition</th>
<th>Examples, Including, but Not Limited to</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A normal healthy patient</td>
<td>Healthy, non-smoking, no or minimal alcohol use</td>
</tr>
<tr>
<td>II</td>
<td>A patient with mild systemic disease</td>
<td>Mild diseases only without substantive functional limitations. Examples include (but not limited to) current smoker, social alcohol drinker, pregnancy, obesity (30 &lt; BMI &lt; 40), well-controlled DM/HTN, mild lung disease</td>
</tr>
<tr>
<td>III</td>
<td>A patient with severe systemic disease</td>
<td>Substantive functional limitations; one or more moderate to severe diseases. Examples include (but not limited to) poorly controlled DM or HTN, COPD, morbid obesity (BMI ≥ 40), active hepatitis, alcohol dependence or abuse, implanted pacemaker, moderate reduction of ejection fraction, ESRD undergoing regularly scheduled dialysis, premature infant PCA &lt; 60 weeks, history (&gt; 3 months) of MI, CVA, TIA, or CAD/stents</td>
</tr>
<tr>
<td>IV</td>
<td>A patient with severe systemic disease that is a constant threat to life</td>
<td>Examples include (but not limited to) recent (&lt; 3 months) MI, CVA, TIA, or CAD/stents, ongoing cardiac ischemia or severe valve dysfunction, severe reduction of ejection fraction, sepsis, DIC, ARDS, or ESRD not undergoing regularly scheduled dialysis</td>
</tr>
<tr>
<td>V</td>
<td>A moribund patient who is not expected to survive without the operation</td>
<td>Examples include (but not limited to) ruptured abdominal/thoracic aneurysm, massive trauma, intracranial bleed with mass effect, ischemic bowel in the face of significant cardiac pathology or multiple organ/system dysfunction</td>
</tr>
<tr>
<td>VI</td>
<td>A declared brain-dead patient whose organs are being removed for donor purposes</td>
<td></td>
</tr>
</tbody>
</table>

The addition of “E” denoted emergency surgery: an emergency is defined as existing when delay in treatment of the patient would lead to a significant increase in the threat to life or body part. ARDS = acute respiratory distress syndrome; BMI = body mass index; CAD = coronary artery disease; COPD = chronic obstructive pulmonary disease; CVA = cerebrovascular accident; DIC = disseminated intravascular coagulation; DM = diabetes mellitus; ESRD = end-stage renal disease; HTN = hypertension; MI = myocardial infarction; PCA = post conceptual age; TIA = transient ischemic attack.

Reproduced with kind permission from Wolters Kluwer Health Inc., Anesthesiology 2017; 126; 614-22.
1.4.4.1.2 Charlson Comorbidity Index

Developed in 1986, the Charlson Comorbidity Index or CCI is a valid prognostic taxonomy for patient comorbidities that might have an impact on postoperative mortality (M. E. Charlson, Pompei, Ales, & MacKenzie, 1987). In this index 19 different comorbidities are listed and different weights are employed for different comorbid diseases. The age of the patient, at the time of surgery, is also factored in, and each decade of age over 40 would add 1 point to the risk (e.g. 50 yrs, 1; 60 yrs, 2; etc.) (M. E. Charlson et al., 1987). The CCI has been widely used in the oncology literature and demonstrated utility for most major cancers (Hall, Ramachandran, Narayan, Jani, & Vijayakumar, 2004).

Table 8 presents the detailed Charlson Comorbidity Index.
### Table 8 - Weighted index of comorbidity (M. E. Charlson et al., 1987)

<table>
<thead>
<tr>
<th>Assigned weights for diseases</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Myocardial infarct</td>
</tr>
<tr>
<td></td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td></td>
<td>Peripheral vascular disease</td>
</tr>
<tr>
<td></td>
<td>Cerebrovascular disease</td>
</tr>
<tr>
<td></td>
<td>Dementia</td>
</tr>
<tr>
<td></td>
<td>Chronic pulmonary disease</td>
</tr>
<tr>
<td></td>
<td>Connective tissue disorder</td>
</tr>
<tr>
<td></td>
<td>Ulcer disease</td>
</tr>
<tr>
<td></td>
<td>Mild liver disease</td>
</tr>
<tr>
<td></td>
<td>Diabetes</td>
</tr>
<tr>
<td>2</td>
<td>Hemiplegia</td>
</tr>
<tr>
<td></td>
<td>Moderate to severe renal disease</td>
</tr>
<tr>
<td></td>
<td>Diabetes with end organ damage</td>
</tr>
<tr>
<td></td>
<td>Any tumor</td>
</tr>
<tr>
<td></td>
<td>Leukemia</td>
</tr>
<tr>
<td></td>
<td>Lymphoma</td>
</tr>
<tr>
<td>3</td>
<td>Moderate or severe liver disease</td>
</tr>
<tr>
<td>6</td>
<td>Metastatic solid tumor</td>
</tr>
<tr>
<td></td>
<td>AIDS</td>
</tr>
</tbody>
</table>

Assigned weights for each condition that a patient has. The total equals the score. Example: chronic pulmonary (1) and lymphoma (2) = total score (3).

Reproduced with kind permission from Elsevier Inc., Journal of Chronic Diseases 1987; 40(5); 373-383.
1.4.4.2 Disease factors

Disease factors are somewhat less straightforward to point out, since many of these characteristics are often accounted for under hospital or surgical variables. The acuteness of the disease is one of these variables represented as an emergency operation (e.g.: a ruptured aortic aneurysm requires emergency intervention mainly because of the progression or an acute change in the disease itself). Emergency surgery was a predictor for an adverse outcome in a variety of surgical specialties, such as vascular surgery (Crawford et al., 2010; Davenport et al., 2007; Neumayer et al., 2007), general surgery (Davenport et al., 2007; Neumayer et al., 2007) or other non-cardiac surgery (Glance et al., 2012; Visser et al., 2015). In oncology, it seems more straightforward to factor in some of the variables related to the disease process itself. Some of them are the stage of the malignancy (Henderson, Pitcher, Steenkamp, Fowler, & Keeley, 2016), or the extent of resection and type of reconstruction (Kodera et al., 2005; Park et al., 2005) required by the extent or nature of the disease.

1.4.4.3 Hospital factors

Great variations between hospitals have been noticed and could not be explained only by the patient-, treatment-, or disease variables (Nordback, Parviainen, Raty, Kuivanen, & Sand, 2002). Therefore, hospital factors have been in the focus of research, especially in the area of high risk and complex diseases. Higher hospital volumes were linked to lower mortality in many of the surgical specialties including major operations, such as urology, general-, thoracic-, vascular-, and cardiac surgery (Begg, Cramer, Hoskins, & Brennan, 1998; Birkmeyer et al., 2002). The level and type of hospitals (Gali, Madziga, Na'aya, & Yawe, 2007; Holm, Johansson, Cedermark, Ekelund, & Rutqvist, 1997), the number of intensive care unit (ICU) beds (Pearse et al., 2012; Vonlanthen & Clavien, 2012) have been the topic of discussion as well and been shown to affect postoperative adverse outcomes.

Besides the level and type of hospital, the length of recovery and stay of the patient has been recognized to contribute to the outcomes as well (Chen et al., 2016; Mari et al., 2016). The fast – track recovery program or also know as Enhanced Recovery After Surgery (ERAS) is a multimodal, multidisciplinary approach, with a combination of optimized evidence-based perioperative management principles, improved surgical and anesthetic techniques, that has led
to improved surgical outcomes (Kehlet, 2015; Kehlet & Jorgensen, 2016; Kehlet & Wilmore, 2008).

1.4.4.4 Surgical factors

The role of the surgeon in the success of an operation as well as a potential contributor to adverse outcomes has been studied extensively. In many of the published articles surgical and surgeon factors have been studied by analyzing the level of training of the surgeon (Sethi, Tay, Scally, & Sood, 2014); surgeon’s experience (Biancari et al., 2012; M. C. Kim et al., 2008); surgeon’s case volume (Bianco, Riedel, Begg, Kattan, & Scardino, 2005; Elsayed, McShane, & Shackcloth, 2012; Schrag et al., 2002), learning curve (Chung et al., 2012; Woelk et al., 2013); operative time (Campbell et al., 2008; Chan, Hamza, & Ammori, 2013; Kessler, Kinkel, Kafer, Puhl, & Schochat, 2003; Kheterpal et al., 2009; Leong, Wilson, & Charlett, 2006; Procter, Davenport, Bernard, & Zwischenberger, 2010; Tan et al., 2012); perioperative blood loss (Liang, Bai, Li, & Zheng, 2007; Pratt, Callery, & Vollmer, 2008; Regenbogen et al., 2008) and perioperative blood transfusion (Elmi et al., 2016; Kheterpal et al., 2009), to name only a few. Surgical skills and performances were inferred from the variables mentioned above, however, when these surrogate variables were used to indirectly measure skills, the evidence was conflicting. Individual performance did not necessarily correlate with surgeon’s experience (Duclos et al., 2012), the level of training and experience did not improve the rate of postoperative complications (Patil, Patel, Mistry, Deshpande, & Desai, 1992; Sethi et al., 2014; Woelk et al., 2013) and surgeon’s volume was a poor surrogate for technical skills (Rutegard, Lagergren, Rouvelas, & Lagergren, 2009).

In summary, patient outcomes are important measurement of the healthcare system and can be affected by a number of different patient-, disease-, hospital-, and surgical variables. Since the operative complications are the ones that occur more often, have more severe consequences and are more often preventable (Zegers et al., 2011), this is the area that needs great attention and the assessment of surgical performance should be part of it. However, technical skills assessment has a wide application mainly in surgical education for summative and formative assessments of trainees, and only a few studies have directly assessed these skills in practicing surgeons yet and even fewer ones have linked them to postoperative patient outcomes (Birkmeyer et al., 2013). To systematically summarize the literature, to present the existent evidence and to highlight the current gaps in knowledge on how technical performance affects
patient outcomes, a systematic review was conducted, which is presented in the next section of the introduction. This review addresses Aim #1 of this thesis.
1.5 The effect of technical performance on patient outcomes in surgery: a systematic review

This section follows a paper format and the content has been published in Annals of Surgery and reprinted with kind permission from Wolters Kluwer Health, Inc. entitled “The Effect of Technical Performance on Patient Outcomes in Surgery: A Systematic Review” by Fecso AB, Szasz P, Kerezov G and Grantcharov TP, Annals of Surgery 2017 Mar; 265(3): 492-501. Formatting has been adapted for the purpose of this thesis.

1.5.1 Abstract

1.5.1.1 Objective

Systematic review of the effect of intraoperative technical performance on patient outcomes.

1.5.1.2 Summary Background Data

The operating room is a high-stakes, high-risk environment. As a result, the quality of surgical interventions affecting patient outcomes has been the subject of discussion and research for years.

1.5.1.3 Methods

MEDLINE, EMBASE, PsycINFO and Cochrane databases were searched. All surgical specialties were eligible for inclusion. Data were reviewed in regards to 1) the methods by which technical performance was measured, 2) what patient outcomes were assessed and 3) how intraoperative technical performance affected patient outcomes. Quality of evidence was assessed using the Medical Education Research Study Quality Instrument (MERSQI).

1.5.1.4 Results

Of the 12758 studies initially identified, 24 articles (7775 total participants) were ultimately included in this review. Seventeen studies assessed the performance of the faculty alone, two assessed both the faculty and trainees, one assessed trainees alone and in four studies, the level
of the operating surgeon was not specified. In 18 studies, a performance assessment tool was used. Patient outcomes were evaluated using: intraoperative complications, short-term morbidity, long-term morbidity, short-term mortality and long-term mortality. The average MERSQI score was 11.67 (Range: 9.5-14.5). Twenty-one studies demonstrated that superior technical performance was related to improved patient outcomes.

1.5.1.5 Conclusions

The results of this systematic review demonstrated that superior technical performance positively affects patient outcomes. Despite this initial evidence, more robust research is needed to directly assess intraoperative technical performance and its effect on postoperative patient outcomes using meaningful assessment instruments and reliable processes.
1.5.2 Introduction

The operating room is a high stakes, high-risk environment (Kohn et al., 2000). Previous research has shown that 39.6 to 54.2 % of adverse events occur in the operating room and that one third to one half of these adverse events are potentially avoidable (de Vries et al., 2008; A. A. Gawande et al., 1999; Kable et al., 2002; Zegers et al., 2009). Adverse events are defined as unexpected consequences caused to the patient by the treatment provided and not the underlying disease (Kohn et al., 2000; Leape et al., 1991; Rebasa et al., 2009). Surgical errors are defined as adverse events directly related to the manual errors of the surgeon (e.g. tissue damage by inappropriate handling) (Bonrath, Dedy, Zevin, & Grantcharov, 2013; Kohn et al., 2000).

Although it has been suggested that cognitive skills outweigh technical skills with regards to the success of a surgical procedure (Satava et al., 2003), there are multiple reasons why it is necessary to investigate elements (i.e., technical skills and errors) that reflect technical performance. Firstly, it has been shown that in many procedures the technical skill of the surgeon may be more important for a successful outcome, than the perioperative care the patient receives (Birkmeyer et al., 2013). Secondly, it has been advocated that surgeons should reflect on their own performance and analyze their errors in order to improve their learning and patient care (Rebasa et al., 2009). Lastly, the objective assessment and subsequent improvement of technical skills, based on feedback and deliberate practice, is important for all surgeons, irrespective of their experience and rank (Hu et al., 2012).

Given the importance of technical errors and adverse events on quality of care, the objective of this study was to systematically review the literature, identify and summarize the existing evidence regarding how intraoperative technical performance affects patients’ outcomes.

1.5.3 Methods

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Moher, Liberati, Tetzlaff, & Altman, 2009) was followed to create the protocol for this systematic review (Figure 1).
1.5.3.1 Search strategy

One author and an information specialist conducted a systematic search of the literature. The following databases were searched: Ovid MEDLINE (R) 1946 to August Week 4 2014; Embase Classic + Embase 1947 to 2014 Week 34; PsycINFO 1806 to August Week 4 2014; EBM Reviews (Cochrane Central Register of Controlled Trials July 2014, Cochrane Database of Systematic Reviews 2005 to July 2014, Database of Abstracts of Reviews of Effects 3rd Quarter 2014). The medical subject headings (MeSH) used included: “surgeons”, “surgical procedure, operative”, “intraoperative procedures”, “perioperative procedures”, “skill, clinical”, “psychomotor performance”, “postoperative complications”, “treatment outcome”. Key terms included “intraoperative”, “technical”, “performance”, “proficient”, “competent”, “ability”, ”postoperative”, “mortality” and “morbidity”. Appropriate variations were also used to account for plurals and other alternatives. The Boolean operator AND was used, to link the above-mentioned terms. Titles and abstracts were initially screened. The full text manuscript of the publications considered relevant were retrieved and reviewed in their entirety. The reference lists of the included publications were manually searched as well, to include any potentially missed publications through the above mentioned database search.

1.5.3.2 Study inclusion criteria

Studies included in this systematic review were limited to original, full-text articles, published in the English literature. Articles were included if intraoperative technical performance was assessed AND if this assessment was linked to patient outcome. Technical performance assessment was defined as any method used within the studies that evaluated the quality of a procedure, rather than the cognitive knowledge and decision-making. Patient outcomes were defined and categorized into intraoperative-, short- and long-term postoperative morbidity and mortality. The short-term postoperative period was defined as the first 30 days following the operation. Long-term postoperative period was defined as the period after the first 30 days following the operation. Outcomes were grouped into 5 groups: (1) intraoperative complications; (2) short-term morbidity; (3) short-term mortality; (4) long-term morbidity and (5) long-term mortality. All operating surgeons (attending and trainees) and surgical specialties were eligible for inclusion. Direct measurement of technical performance included assessment methods with or without the usage of a specific rating tool.
1.5.3.3 Study exclusion criteria

Articles not representing original research, such as letters, commentaries, reviews, or opinion papers were excluded. Studies assessing technical performance, but not patient outcomes, or evaluating patient outcomes without assessing technical performance, were also excluded. Studies without accessible and retrievable full-text manuscripts, such as conference abstracts or posters only, were also excluded.

1.5.3.4 Extracted data and quality assessment of the included studies

Data was reviewed in regards to the methods by which technical performance was measured, what patient outcomes were assessed and how intraoperative technical performance affected these outcomes. Two authors assessed the quality of the evidence using the MERSQI (Medical Education Research Study Quality Instrument) rating system (Cook & Reed, 2015; Lin, Lin, Auditore, & Fanning, 2016; Reed et al., 2008; Reed et al., 2007). The MERSQI rating system was designed to measure the methodological quality of studies within the field of medical education (Cook & Reed, 2015; Reed et al., 2007). The instrument contains 10 separate items reflecting 6 domains of study quality: study design, sampling, type of data (subjective or objective), validity, data analysis, and outcomes (Reed et al., 2007). The study quality refers to research rather than reporting quality (Reed et al., 2007). The MERSQI score ranges from a minimum of 5 to a maximum of 18 (Reed et al., 2007). Any discrepancies in the scores were resolved via consensus.

1.5.4 Results

1.5.4.1 Included studies

The initial search, after removing duplicates of non-human and non-English articles, yielded 12758 articles. After the title and abstract review, using predefined exclusion and inclusion criteria, 114 articles underwent full-text review of which 21 were ultimately included. A review of references of the included articles yielded three more relevant publications for inclusion, resulting in a total of 24 articles in this systematic review. These 24 articles had a total of 7775 participants.
Figure 1 illustrates the review algorithm in the PRISMA format.

Of the 24 included studies, 23 were observational studies and one was a randomized controlled trial. The following specialties/sub-specialties were represented: Cardiac Surgery (nine studies); Orthopedic Surgery (five studies); General Surgery (four studies); Neurosurgery (one study); Ophthalmology (one study); Vascular Surgery (one study); Urology (one study); Multiple surgical specialties (two studies). The latter two studies analyzed surgical errors and performance in closed malpractice cases.

Table 9 shows an overview and the characteristics of the included 24 articles.
Publications identified through all database searches; n=23713

Number of publications after duplicate, non-human and non-English articles removed: n=12758

Studies excluded after screening titles and abstracts: n=12644

Full text articles excluded based on eligibility criteria: n=93

Specific rationale for exclusion:
- No objective assessment of technical performance
- No outcome assessed
- Technical performance not linked to outcome
- Not original research (review, book chapter)

Number of articles undergoing full text review: n=114

Studies identified via review of the lists of references and included in the systematic review: n=3

Number of articles included in systematic review: n=24
Table 9 - Overview of the studies

<table>
<thead>
<tr>
<th>Study No.</th>
<th>Author</th>
<th>Year</th>
<th>Design</th>
<th>Specialty</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bacha et al.</td>
<td>2008</td>
<td>R. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>2</td>
<td>Karamichalis et al.</td>
<td>2010</td>
<td>R. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>3</td>
<td>Karamichalis et al.</td>
<td>2012</td>
<td>P. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>4</td>
<td>Larrazabal et al.</td>
<td>2007</td>
<td>R. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>5</td>
<td>Nathan et al.</td>
<td>2011</td>
<td>P. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>6</td>
<td>Nathan et al.</td>
<td>2012</td>
<td>P. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>7</td>
<td>Nathan et al.</td>
<td>2013</td>
<td>R. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>8</td>
<td>Nathan et al.</td>
<td>2014</td>
<td>P. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>9</td>
<td>Shuhaiber et al.</td>
<td>2012</td>
<td>R. Observational study</td>
<td>Cardiac Surgery</td>
</tr>
<tr>
<td>10</td>
<td>Boriani et al.</td>
<td>1991</td>
<td>R. Observational study</td>
<td>Orthopedic Surgery</td>
</tr>
<tr>
<td>11</td>
<td>Docquier et al.</td>
<td>2002</td>
<td>R. Observational study</td>
<td>Orthopedic Surgery</td>
</tr>
<tr>
<td>12</td>
<td>Frank et al.</td>
<td>2012</td>
<td>R. Observational study</td>
<td>Orthopedic Surgery</td>
</tr>
<tr>
<td>13</td>
<td>Koprowski et al.</td>
<td>2007</td>
<td>R. Observational study</td>
<td>Orthopedic Surgery</td>
</tr>
<tr>
<td>14</td>
<td>Pascarella et al.</td>
<td>2008</td>
<td>R. Observational study</td>
<td>Orthopedic Surgery</td>
</tr>
<tr>
<td>15</td>
<td>Arvidsson et al.</td>
<td>2005</td>
<td>P. Randomized control trial</td>
<td>General Surgery</td>
</tr>
<tr>
<td>16</td>
<td>Birkmeyer et al.</td>
<td>2013</td>
<td>R. Observational study</td>
<td>General Surgery</td>
</tr>
<tr>
<td>17</td>
<td>Scheyer et al.</td>
<td>1996</td>
<td>R. Observational study</td>
<td>General Surgery</td>
</tr>
<tr>
<td>18</td>
<td>Way et al.</td>
<td>2003</td>
<td>R. Observational study</td>
<td>General Surgery</td>
</tr>
<tr>
<td>19</td>
<td>Amato et al.</td>
<td>2010</td>
<td>R. Observational study</td>
<td>Neurosurgery</td>
</tr>
<tr>
<td>20</td>
<td>Corey et al.</td>
<td>1998</td>
<td>R. Observational study</td>
<td>Ophtalmology</td>
</tr>
<tr>
<td>21</td>
<td>Katz et al.</td>
<td>1994</td>
<td>R. Observational study</td>
<td>Vascular surgery</td>
</tr>
<tr>
<td>22</td>
<td>Stern et al.</td>
<td>2009</td>
<td>P. Observational study</td>
<td>Urology</td>
</tr>
<tr>
<td>23</td>
<td>Rogers et al.</td>
<td>2006</td>
<td>R. Observational study</td>
<td>Multiple surgical specialities</td>
</tr>
<tr>
<td>24</td>
<td>Somville et al.</td>
<td>2010</td>
<td>R. Observational study</td>
<td>Multiple surgical specialities</td>
</tr>
</tbody>
</table>

R = Retrospective; P= Prospective
1.5.4.2 Assessment of technical performance

Seventeen studies assessed the performance of the faculty alone, two studies assessed both the faculty and trainees, one assessed trainees alone, and in four studies the level of the surgeon assessed was not specified.

Eighteen studies used a performance assessment tool. In the remaining six studies the tool was not specified. Given a variety of surgical specialties included, the assessment tools varied from specialty to specialty. Out of those studies, using a performance assessment method, 14 were based on postoperative data and four were based on intraoperative data. For example, in Cardiac Surgery, the parameters for the score assessment were based on postoperative clinical, echocardiographic and cardiac catheterization data (Bacha et al., 2008; Karamichalis et al., 2012; Karamichalis et al., 2010; Larrazabal et al., 2007; Nathan et al., 2014; Nathan et al., 2011; Nathan et al., 2012; Nathan et al., 2013; Shuhaiber et al., 2012). In Orthopedic Surgery (Docquier, Manche, Autrique, & Geulette, 2002; Frank et al., 2012) and Neurosurgery (Amato, Giannachi, Irace, & Corona, 2010) the parameters were based on the accurate placement of the implant or fixation material, determined on postoperative radiographs or computerized tomography (CT) scans. In two studies, looking at multiple surgical specialties, the parameters were collected based on postoperative chart reviews (Rogers et al., 2006; Somville, van Sprundel, & Somville, 2010). In General Surgery (Arvidsson et al., 2005; Birkmeyer et al., 2013; Way et al., 2003) and Urology (Stern et al., 2011), the parameters for the assessment method were based on intraoperative data, either from direct observations or review of the video recorded procedure using various assessment tools such as the Objective Structured Assessment of Technical Skills (OSATS). In the remaining six studies, where the use of the specific performance metric was not specified, four studies used postoperative data [three Orthopedic Surgery (Boriani et al., 1991; Koprowski et al., 2007; Pascarella et al., 2008); one Vascular Surgery (Katz & Kohl, 1994)] and two used intraoperative data [one General Surgery (Scheyer & Zimmermann, 1996); one Ophthalmology (Corey & Olson, 1998)] to determine performance.

In 20 studies, the raters were the authors themselves, in three studies the raters were independent observers, and in one study both the authors and an independent observer rated the performance (Table 10). In only one study, there was data showing that the raters were trained to analyze the
performance (Rogers et al., 2006). In another study the raters were not trained (Birkmeyer et al., 2013), whereas in the remaining studies there was no data or discussion regarding rater training.

1.5.4.3 Assessment of patient outcomes

Outcomes were divided into five groups: (1) intraoperative complications; (2) short-term morbidity; (3) short-term mortality; (4) long-term morbidity and (5) long-term mortality. Nineteen studies looked at more than one, whereas five studies looked at a single outcome measure. Intraoperative complications were the outcome of interest in eight studies, short-term morbidity in 17, short-term mortality in 11, long-term morbidity in 11 and long-term mortality in four studies (Table 10).

1.5.4.4 Relationship between technical performance and patient outcomes

Twenty-one studies showed that superior technical performance had a positive effect on patient outcomes. The remaining three studies did not show this association (Larrazabal et al., 2007; Stern et al., 2011; Way et al., 2003).

Table 10 summarizes these relationships and the conclusion of each study.
## Study conclusion

Inadequate performance increases risk for all clinical outcomes. Optimal technical performance results in good outcomes and attenuates the effect of illness severity or case complexity. Less than optimal performance increases mortality in high risk groups and increases resources for lower risk groups.

### Table 10 - Technical performance and patient outcomes assessed

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of Assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacha et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✓</td>
<td>Inadequate performance increases risk for all clinical outcomes</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✓</td>
<td>Optimal technical performance results in good outcomes and attenuates the effect of illness severity or case complexity</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✓</td>
<td>Less than optimal performance increases mortality in high risk groups and increases resources for lower risk groups</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
Study conclusion

Technical performance is not associated with postoperative outcomes

Optimal technical performance results in best outcome and attenuates the effects of poor preoperative physiology

Technical performance predicts mid-term mortality and morbidity

Table 10 - (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Larrazabal et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔ ✔</td>
<td>Technical performance is not associated with postoperative outcomes</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔ ✔ ✔</td>
<td>Optimal technical performance results in best outcome and attenuates the effects of poor preoperative physiology</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔ ✔</td>
<td>Technical performance predicts mid-term mortality and morbidity</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
Table 10 - (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nathan et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔️</td>
<td>Technical performance is strongly associated with late mortality and late morbidity</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔️</td>
<td>Technical performance is strongly associated with early morbidity and mortality</td>
</tr>
<tr>
<td>Shuhaiber et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Discharge technical score</td>
<td>Authors</td>
<td>✔️</td>
<td>Inadequate technical performance is associated with short-term mortality</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
### Table 10 - (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boriani et al.</td>
<td>Staff</td>
<td>Not specified</td>
<td>Accuracy of nail and screw placement on postoperative radiographs</td>
<td>Authors</td>
<td>✓ ✓</td>
<td>Good technical performance is associated with good functional outcomes</td>
</tr>
<tr>
<td>Docquier et al.</td>
<td>Staff and trainees</td>
<td>Yes</td>
<td>Accuracy of nail and screw placement on postoperative radiographs</td>
<td>Authors</td>
<td>✓ ✓</td>
<td>Good technical performance and knowledge of the procedure can decrease intra-and postoperative complications</td>
</tr>
<tr>
<td>Frank et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Accuracy of fixation and tunnel positioning on radiographic measurements and operative reports</td>
<td>Authors</td>
<td>✓</td>
<td>Traumatic reinjury, inadequate technical performance and biologic failure does influence outcomes</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koprowski et al.</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Accuracy of fixation on postoperative radiographs</td>
<td>Authors</td>
<td>✔</td>
<td>Incorrect technical performance is associated with long term morbidity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pascarella et al.</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Accuracy of nail and screw placement on postoperative radiographs</td>
<td>Authors</td>
<td>✔ ✔ ✔</td>
<td>Incorrect technical performance leads to major complications</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arvidsson et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Intraoperative observations and review of video-recorded operations</td>
<td>Independent observer</td>
<td>✔</td>
<td>Technical performance is inversely correlated with long-term morbidity (recurrence)</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
Study conclusion

Technical performance is inversely associated with postoperative morbidity and mortality

Technical performance influences short and long-term morbidity

Misperception, rather than technical performance, is associated with intraoperative complications

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of Assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birkmeyer et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Review of video-recorded operations</td>
<td>Independent observers</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Technical performance is inversely associated with postoperative morbidity and mortality</td>
</tr>
<tr>
<td>Scheyer et al.</td>
<td>Staff</td>
<td>Not specified</td>
<td>Review of video-recorded operations</td>
<td>Authors</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Technical performance influences short and long-term morbidity</td>
</tr>
<tr>
<td>Way et al.</td>
<td>Not specified</td>
<td>Yes</td>
<td>Review of operative-, pathology-, radiology reports and video-recorded operations</td>
<td>Authors</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Misperception, rather than technical performance, is associated with intraoperative complications</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)  

Continues
Table 10 - (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amato et al.</td>
<td>Staff</td>
<td>Yes</td>
<td>Accuracy of implant placement, on postoperative CT</td>
<td>Authors and an independent observer</td>
<td>✓</td>
<td>✓ Technical performance influences short and long-term morbidity</td>
</tr>
<tr>
<td>Corey et al.</td>
<td>Trainees</td>
<td>Not specified</td>
<td>Intraoperative observations</td>
<td>Independent observers</td>
<td>✓</td>
<td>✓ Technical performance is an important factor in the complication rate</td>
</tr>
<tr>
<td>Katz et al.</td>
<td>Staff</td>
<td>Not specified</td>
<td>Review of operative records</td>
<td>Authors</td>
<td>✓</td>
<td>✓ Technical performance influences short-term morbidity and mortality</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
### Table 10 - (Continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of performer assessed</th>
<th>Performance assessment tool</th>
<th>Method of assessment</th>
<th>Assessors of performance</th>
<th>Outcomes</th>
<th>Study conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stern <em>et al.</em></td>
<td>Staff</td>
<td>Yes</td>
<td>Intraoperative self assessment</td>
<td>Author</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Self-assessed technical performance is not a good predictor of short and long term outcomes</td>
</tr>
<tr>
<td>Rogers <em>et al.</em></td>
<td>Staff and trainees</td>
<td>Yes</td>
<td>Review of medical records</td>
<td>Authors</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Somville <em>et al.</em></td>
<td>Not specified</td>
<td>Yes</td>
<td>Review of medical records</td>
<td>Authors</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

IO= Intraoperative; STMB= Short-term morbidity (<POD#30); STMT= Short-term mortality (<POD#30); LTMB=Long-term morbidity (>POD#30); LTMT=Long-term mortality (>POD#30)
1.5.4.5 Quality of Evidence

Table 11 shows the overall MERSQI score as well as the components of the score for all 24 studies. For the included articles in this review, the average MERSQI score was 11.67 (Range: 9.5-14.5). In medical education, the average MERSQI score, for the assessment of the quality of research studies, is 9.6 out of the maximum of 18 (Lin et al., 2016; Reed et al., 2008).

In the 18 studies, where a performance assessment tool was utilized, the validity of those tools was determined using the current, conceptual framework, proposed by Messick (Ghaderi et al., 2015; Messick, 1995). Also, as part of the MERSQI score, which was used to assess the quality of the studies as described above, three of the five elements of this conceptual framework were analyzed and these included content, internal structure and relationships to other variables sources (Table 11). Seventeen studies had content evidence. None of the studies, included in this systematic review, contained data to demonstrate all five sources of validity.

Table 12 shows the different sources of validity included in Messick’s framework.
<table>
<thead>
<tr>
<th>Author</th>
<th>Overall score</th>
<th>Study design</th>
<th>Sampling</th>
<th>Data type</th>
<th>Data analysis</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Institutions</td>
<td>Response rate</td>
<td>Content</td>
<td>Internal structure</td>
</tr>
<tr>
<td>Bacha et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Larrazabal et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Shuhaiberg et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Boriani et al.</td>
<td>11</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Docquier et al.</td>
<td>10.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Frank et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Koprowski et al.</td>
<td>11.5</td>
<td>2</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Pascarella et al.</td>
<td>9.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Arvidsson et al.</td>
<td>14.5</td>
<td>3</td>
<td>1.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Birkmeyer et al.</td>
<td>14.5</td>
<td>1</td>
<td>1.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Scheyer et al.</td>
<td>9.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Way et al.</td>
<td>10.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Amato et al.</td>
<td>12.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Corey et al.</td>
<td>11.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Katz et al.</td>
<td>12.5</td>
<td>1</td>
<td>1.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Stern et al.</td>
<td>9.5</td>
<td>1</td>
<td>0.5</td>
<td>N/A</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rogers et al.</td>
<td>14.5</td>
<td>1</td>
<td>1.5</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Somville et al.</td>
<td>13</td>
<td>1</td>
<td>1</td>
<td>N/A</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

MERSQI = Medical Education Research Study Quality Instrument; N/A = Not available
<table>
<thead>
<tr>
<th>Author</th>
<th>Content Evidence</th>
<th>Response Process</th>
<th>Internal Structure</th>
<th>Relationship to Other Variables</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacha et al.</td>
<td>✓</td>
<td>N/A</td>
<td>□</td>
<td>⌂</td>
<td>N/A</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Karamichalis et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Larrazabale et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Nathan et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Shuhaiber et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Docquier et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Frank et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Arvidsson et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Birmeyer et al.</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Way et al.</td>
<td>X</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Amato et al.</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>X</td>
<td>N/A</td>
</tr>
<tr>
<td>Stern et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>X</td>
<td>✓</td>
</tr>
<tr>
<td>Rogers et al.</td>
<td>✓</td>
<td>N/A</td>
<td>✓</td>
<td>✓</td>
<td>N/A</td>
</tr>
<tr>
<td>Somville et al.</td>
<td>✓</td>
<td>N/A</td>
<td>X</td>
<td>✓</td>
<td>N/A</td>
</tr>
</tbody>
</table>

N/A = Information was not available in regards to these categories within the studies
1.5.5 Discussion

Using a systematic approach, this review aimed to determine the methods by which intraoperative technical performance is assessed and whether it has an effect on patient outcomes. This is the first review demonstrating that superior technical performance is related to improved patient outcomes (intraoperative, postoperative short- and long-term). In this systematic review all but three studies showed that superior technical performance positively affected patient outcomes (Amato et al., 2010; Arvidsson et al., 2005; Bacha et al., 2008; Birkmeyer et al., 2013; Boriani et al., 1991; Corey & Olson, 1998; Docquier et al., 2002; Frank et al., 2012; Karamichalis et al., 2012; Karamichalis et al., 2010; Katz & Kohl, 1994; Koprowski et al., 2007; Nathan et al., 2014; Nathan et al., 2011; Nathan et al., 2012; Nathan et al., 2013; Pascarella et al., 2008; Rogers et al., 2006; Scheyer & Zimmermann, 1996; Shuhaiber et al., 2012; Somville et al., 2010). The remaining three studies did not show a positive relationship (Larrazabal et al., 2007; Stern et al., 2011; Way et al., 2003). This could have been the result of uneven distribution of data or insufficient variability of the data, as stipulated by the authors of these studies (Larrazabal et al., 2007; Stern et al., 2011). In one of these studies, although a deficiency was noted in technical performance, and this did have an effect on patient outcomes, other variables were shown to impact outcomes more than performance (e.g. misperception of the surgical anatomy) (Way et al., 2003).

Assessment of technical performance needs to be done in a direct manner, rather than using surrogate variables. This will allow a better understanding of the technical deficiencies and development of targeted educational intervention designed to address these deficiencies. In much of the published literature, across various surgical (sub)specialties, technical performance was assessed indirectly, using various surrogate variables, such as surgeon’s experience (Biancari et al., 2012; Chestnut, Eden, Gall, & Parker, 1985; Duclos et al., 2012; Friend, Scott, Everett, & Scott, 1990; Gabrielli et al., 2001; Holm et al., 1997; M. C. Kim et al., 2008; Magheli et al., 2014; Patil et al., 1992; Plecha, Seabrook, Bandyk, & Towne, 1993), operative time and surgeon’s case volume (Bianco et al., 2005; Chung et al., 2012; Elsayed et al., 2012; Pratt et al., 2008; Rutegard et al., 2009; Woelk et al., 2013), level of hospital (Gali et al., 2007; Holm et al., 1997), case complexity and survival (Kang, Tsang, Elliott, de Leval, & Cole, 2006; Photiadis et al., 2011), postoperative bleeding (Liang et al., 2007), hospital volume (Nordback et al., 2002), estimated intraoperative blood loss (Pratt et al., 2008; Regenbogen et al., 2008) and level of
training (Sethi et al., 2014). However, when performance was assessed using surrogate variables, the results of the published studies are conflicting. These studies have shown that individual performance does not necessarily correlate with surgeon experience (Duclos et al., 2012), and there is no difference in the rate of postoperative complications based on the level of training (Patil et al., 1992; Sethi et al., 2014). Moreover, surgeon’s volume has been shown to be a poor surrogate for skill (Rutegard et al., 2009), and although operative time and length of stay might decrease as experience increases, complication rate may not (Woelk et al., 2013).

Majority of the studies in this review assessed the performance of faculty with various experience and number of years in practice. However none of the studies examined the possibility of learning curve potentially contributing to the technical performance. The effect of learning curve on performance was outside of the scope of this review.

The studies included in this systematic review measured technical performance, however, most of them did it via indirect observations, by using postoperative data, looking at the “end result” to determine the performance leading to that result. This approach is not ideal, due to the possible impact of other confounding factors, related to patient- and peri-operative characteristics. For example, in Cardiac Surgery, using data from a postoperative echocardiography has its own limitations - that of being operator dependent. (Larrazabal et al., 2007). In Orthopedic Surgery, not completing radiographs for each operation performed in order to determine the accuracy of the fixation, might falsely increase/decrease the number of negative outcomes as a result of technical error (Frank et al., 2012). Finally, performance analysis based on retrospective chart reviews, (i.e. malpractice cases), might introduce a hindsight bias, with the reviewers having knowledge of patient outcomes when evaluating such claims (Rogers et al., 2006; Somville et al., 2010). Similarly to Moorthy et al. (Moorthy, Munz, Sarker, & Darzi, 2003), the authors of this systematic review believe that direct observation in the operating room, or review of a video recorded procedure and the use of specific assessment tools/criteria are seen as a more reliable and valid means of assessing technical performance. In this review, only seven studies used such direct observations to evaluate intraoperative data (either via direct assessment of live observations or review of video-recordings), alone (Arvidsson et al., 2005; Birkmeyer et al., 2013; Corey & Olson, 1998; Scheyer & Zimmermann, 1996; Stern et al., 2011) or in combination with postoperative data (Nathan et al., 2011; Way et al., 2003), to measure technical performance.
Although this approach is valid as discussed above, there are other limitations in some of these studies, which may preclude the accurate assessment of technical performance. These include the use of self-selected (on the part of the participant being assessed) and edited videos, and the use of untrained raters (Birkmeyer et al., 2013), all of which can introduce a bias, by excluding relevant data or educated opinions. Furthermore, the lack of a valid and reliable rating tool used (Scheyer & Zimmermann, 1996) might question the validity of the rating itself. On the other hand, just because a rating tool was used, does not necessarily mean that the provided results are valid, as was demonstrated in a study where self-assessment was not a good predictor of perioperative outcomes (Stern et al., 2011). This questions the validity of the assessment method and highlights the importance of using independent, trained observers with content expertise.

Assessment of specific technical errors, e.g. instrument handling, time and motion, tissue handling, knot tying, etc., need to be part of the rating of technical performance. This detailed knowledge of performance is extremely relevant in order to design effective educational interventions that can result in improved quality and safety. Most of the studies included in this review did not provide information about the specific type of technical errors that could potentially lead to complications. Only one study, from Scheyer et al. (Scheyer & Zimmermann, 1996) discussed these aspects in some details, e.g. a misplaced clip causing paraesthesia or pain in the postoperative phase.

The process of assessing technical performance has to produce results that have evidence to support their use (Ghaderi et al., 2015; Messick, 1995). In the traditional framework, validity consisted of separate domains such as – content, criterion and construct- types of validity (Cronbach & Meehl, 1955; Downing, 2003; Ghaderi et al., 2015). In the current and conceptual (Messick’s) framework, validity is defined as appropriately interpreting assessment results (Ghaderi et al., 2015). Therefore, the validity applies only to the scores or their interpretation, and not to the instrument (rating tool) itself (Cook & Reed, 2015; Schmitz, 2006). In this contemporary framework, construct validity is the only form of validity (Downing, 2003; Ghaderi et al., 2015). The validation process requires identification of relevant sources of validity, such as: content, response process, internal structure, relationships to other variables, and consequences of the ratings (Ghaderi et al., 2015). The majority of the studies in this review displayed content evidence only. Fewer studies displayed additional sources, such as - in descending order - relationship to other variables, internal structure, consequences and response
process. None of them displayed evidence in all five sources of validity. Interestingly, despite the trend to utilize this conceptual framework, most of the studies still adhered to the traditional concepts of validity. Although, the detailed discussion of these concepts is beyond the scope of this review, it is important to acknowledge that future assessments should follow the modern framework, and rather than developing new assessment tools, work needs to be directed towards the improvement of the existing tools as well as the assessment process (Ghaderi et al., 2015).

Most of the studies included in this review had a higher MERSQI score than the previously reported averages in the field of medical education research (Lin et al., 2016; Reed et al., 2008). This could be explained by the fact, that all of the studies included in this review, evaluated patient outcomes (the highest outcome level assessed using the MERSQI scoring system), therefore automatically scoring a maximum of three points in this category. Birkmeyer et al. (Birkmeyer et al., 2013) published a landmark study, which should be used as a methodological model for future, high quality research in the area.

Postoperative outcomes are used for multiple purposes. The healthcare system and healthcare providers use this information as part of their quality metrics and improvement protocols, and patients use this information to get a clear picture of the potential benefits and harms of a particular procedure (Birkmeyer, Dimick, & Birkmeyer, 2004). When evaluating postoperative outcomes, both mortality and morbidity are measured (Birkmeyer et al., 2004). However, these traditional, “hard endpoints in mortality and morbidity” are no longer viewed as the sole measurements of success of a given treatment (Soreide & Soreide, 2013). More emphasis is being attributed to patient reported outcome measures (PROM) (e.g. functional status; health related quality of life, etc.) (Soreide & Soreide, 2013). These two aspects of postoperative outcomes (both morbidity/mortality and PROMs) are considered equally important and they are jointly able to measure more comprehensively the success of a treatment (Soreide & Soreide, 2013). In this systematic review all included articles looked at postoperative outcomes (morbidity and mortality), as this was one of the study inclusion criteria. However, only two articles looked at PROMs as well. Arvidsson et al. (Arvidsson et al., 2005) assessed postoperative pain (using a visual analog scale), duration of sick leave and patient complaints at three months follow-up. Boriani et al. (Boriani et al., 1991) used postoperative walking ability and presence of pain as patient reported outcome measure. Going forward, for comprehensive
and accurate outcome measurements, both, traditional and patient reported outcomes, should be collected and analyzed together.

There are some limitations of this review that require discussion. Since studies were included across all surgical (sub) specialties, the data is heterogeneous, most studies are low quality and the methods of assessment utilized in these various disciplines are, and rightly so, different; therefore the results need to be interpreted with this in mind. A “one size fits all” approach might not be feasible across all specialties. We do, however, think, and agree with previously published literature, that the future of assessing performance lies in utilizing trained raters assessing the entire length of a procedure, regardless of surgical specialty (Szasz, Louridas, Harris, Aggarwal, & Grantcharov, 2015).

1.5.6 Conclusion

The results of this systematic review demonstrate that superior technical performance positively affects patient outcomes. The present study provides this initial evidence, however the majority of the studies have several limitations and are low quality, as discussed above. Future research endeavors should adopt a more robust methodological approach to error and performance analysis and directly assess intraoperative technical performance and its effect on postoperative patient outcomes. The authors believe that trained observers with content expertise should directly observe an operation or review video recorded procedures to provide an accurate assessment of the individual’s technical performance, using existing rating instruments with evidence to support their use, according to Messick’s framework. The lack of PROMs identifies a void currently present in evaluating postoperative outcomes and highlights an area for future impactful research endeavors.

1.6 Summary

In summary, Chapter 1 provided a relevant background introduction and overview of the relevant topics included in this PhD thesis, namely surgical performances and patient outcomes. It also summarized the current evidence on how technical performance affected patient outcomes and more importantly highlighted the deficiencies in the published studies. As demonstrated, intraoperative technical performance of the surgeon does affect patient outcomes. Since the operative complications are the ones that occur more often, have more severe
consequences and are more often preventable (Zegers et al., 2011) and most of them are technical in nature, it is important to optimize this environment in order to improve patient safety and outcomes.

The following chapters in this thesis aim to strengthen this body of evidence and narrow the current gap in knowledge.
CHAPTER 2

The purpose of this chapter is to:

I. Describe the purpose statement
II. Provide the research hypotheses
III. List the specific aims and objectives
2 Hypotheses, Aims, Objectives

2.1 Purpose statement

The main purpose of this thesis is to determine the effects intraoperative technical performances have on short-term postoperative patient outcomes.

2.2 Hypotheses

2.2.1 Main hypothesis

Intraoperative technical performance of the surgeon affects short-term postoperative patient outcomes and superior skills lead to decreased complications in laparoscopic gastric cancer operations.

2.2.2 Secondary hypotheses

Employing a systematic review strategy, a summary of the current evidence can be provided on how technical performance affects patient outcomes as well as future research directions can be highlighted.

Trainees (residents and fellows) do not routinely perform the major operative steps of a laparoscopic gastrectomy for cancer at teaching hospitals within a large university center.

Intraoperative team (surgical and nursing) non-technical (intrinsic) performances and individual surgeon’s technical performance are correlated.

2.3 Aims and objectives

The overall aims are presented along with their objectives right below.

Aim 1:
To systematically review the literature evaluating how technical performance affects patient outcomes in the broad specialty of surgery.
Objectives 1:
1. Review of data in regards to the methods by which technical performance was assessed.
2. Determine the type of patient outcomes assessed.
3. Establish the effects technical performance had on patient outcomes.

Aim 2:
To assess the relationship between technical performance and short-term patient outcomes in laparoscopic gastric cancer operations

Objectives 2:
1. Assessment of intraoperative technical skills using unedited, full length, de-identified video recordings.
3. To determine the relationship between performance and outcomes.

Aim 3:
To determine the level of involvement of residents and fellows in laparoscopic gastric cancer operations across teaching hospitals within a large university center.

Objectives 3:
1. Design of a survey instrument to represent the main steps of a laparoscopic gastric cancer operation.
2. Administration and analysis of the survey results

Aim 4:
To define the interaction between team members in the operating room, how non-technical performances of the team and the surgeons technical performances are correlated.

Objectives 4:
1. Assessment of intraoperative technical performance of the surgeon.
2. Assessment of intraoperative non-technical (intrinsic) performances of the surgical and nursing team
3. Establishing the relationship between technical and non-technical performances.
CHAPTER 3

The purpose of this chapter is to:

I. Design a methodology to address the main study limitations identified in Chapter 1
II. Demonstrate the effects of technical performances on patient outcomes

The study in this chapter is being considered for publication at Annals of Surgery (ANNSURG-D-17-0200) with the title “Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery” by Fecso AB, Bhatti JA, Stotland PK, Quereshy FA and Grantcharov TP.
Formatting has been adapted for the purpose of this thesis.
3 Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery

3.1 Abstract

3.1.1 Introduction

Laparoscopic gastrectomy for cancer is an advanced intra-abdominal procedure with a high rate of postoperative morbidity and mortality. Many variables including patient-, disease- and perioperative management factors have been shown to impact postoperative outcomes; however, the role of surgical performance is insufficiently investigated. The purpose of this study was to evaluate the relationship between technical performance and patient outcomes in laparoscopic gastric cancer surgery.

3.1.2 Methods

A retrospective review was performed for all patients who had undergone laparoscopic gastrectomy for cancer at three teaching institutions between 2009 - 2015. Patients with available, unedited videorecording of their procedure were included in the study. Video files were rated for technical performance, using OSATS (Objective Structured Assessments of Technical Skills) and GERT (Generic Error Rating Tool) instruments. The main outcome variable was short-term outcomes. The effect of technical performance on patient outcomes was assessed using logistic regression analysis with backward selection strategy.

3.1.3 Results

Sixty-one patients with available video recordings were included in the study. The overall complication rate was 29.5%. The mean Charlson Comorbidity Index, type of procedure and the global OSATS score were included in the final predictive model. Lower performance score (OSATS≤29) remained an independent predictor for short-term outcomes [Odds Ratio (OR): 6.49], while adjusting for comorbidities and type of procedure.
3.1.4 Conclusion

3.2 Introduction

Multiple reports suggest that a successful perioperative course of a surgical patient is multifactorial. In this process, the intraoperative performance of the surgical team is important, however, the evidence for the impact of surgeon’s technical skills on clinical outcomes is still limited (Fecso, Szasz, Kerezov, & Grantcharov, 2017).

Despite the notion that the cognitive skills might outweigh other aspects of surgical performance (Satava et al., 2003), technical performance remains an important component and it might contribute to a successful outcome more than the perioperative care (Birkmeyer et al., 2013). A recently published systematic review supported the importance of technical skills and suggested that superior performance positively affected patient outcomes (Fecso, Szasz, et al., 2017).

Laparoscopic gastrectomy for cancer is an advanced intra-abdominal procedure that requires the highest level of technical skills and the postoperative surgical morbidity and mortality can be substantial (Li et al., 2011; Orsenigo, Di Palo, Tamburini, & Staudacher, 2011; Stotland, Chia, Cyriac, Hagen, & Klein, 2009).

In gastric surgery, similarly to other procedures, multiple patient, disease and procedure related factors have been linked to patient outcomes. Some of the patient factors studied in this context are age, sex, the presence of comorbidities, body mass index (BMI) and visceral fat areas (VFA) (M. C. Kim et al., 2008; W. Kim et al., 2008; Kodera et al., 2005; Kunisaki et al., 2009; Park et al., 2005). Factors related to the surgical procedure are the extent of lymph node dissection, type of resection and reconstruction, multivisceral resection, operative time and surgeon’s experience (M. C. Kim et al., 2008; Kodera et al., 2005; Park et al., 2005; Ryu et al., 2008). To date, there are no reports investigating the effect of surgical performance on patient outcomes.

The purpose of this study was to directly assess intraoperative technical performance in laparoscopic gastrectomy for cancer and explore its effect on postoperative, short-term outcomes.
3.3 Methods

Institutional ethics approval was obtained prior to the start of this retrospective, multi-center, multi-surgeon cohort study at the University of Toronto.

3.3.1 Patients

All adult patients who underwent a laparoscopic gastrectomy (total or partial) for gastric adenocarcinoma in the period of January 1st, 2009 and December 31st, 2015 were eligible for inclusion in the study. Those patients whose procedure was video-recorded and available for review in an unedited fashion, were included. Patients’ charts were reviewed for patient-, disease-, treatment characteristics and postoperative short-term outcomes.

3.3.2 Measures

The main outcome variable was short-term complications. For the purpose of this study two patient groups were defined - no complications and complications groups. Clavien-Dindo classification of surgical complications was used to grade postoperative morbidity and mortality (Dindo et al., 2004) and the complications group was defined as Clavien Dindo Grade ≥III. In the no complications groups, minor complications (Clavien Dindo I &II) were also included.

Short-term period was defined as the first 30 days after the initial elective operation as long as the patient was discharged prior to this. If the patient was admitted longer, then the last day of the admission was the end of the postoperative study period. For in-hospital deaths, the last day of the postoperative period was the day of death.

Within the University of Toronto, all surgeons who perform laparoscopic gastrectomy for cancer on a regular basis and record their procedure, were invited to participate. Three surgeons from three teaching institutions participated. All three surgeons were fellowship trained (surgical oncology or minimally invasive surgery), and were practicing at academic or university affiliated community hospitals. Surgeons previous experience was defined as at least 30 performed procedures prior to submission of the first video. All participating centers were teaching institutions and all procedures involved participation of surgical residents and fellows.
3.3.3 Surgical Technique

The extent of the lymphadenectomy was determined using the Japanese gastric cancer treatment guidelines 2014 (ver.4) ("Japanese gastric cancer treatment guidelines 2014 (ver. 4)," 2017). The quality of the lymphadenectomy was assessed using the KLASS-02-QC study’s evaluation criteria for completeness of D2 lymphadenectomy (H. I. Kim et al., 2014).

The surgical procedure was divided into three stages: (1) dissection; (2) resection and dissection; (3) reconstruction. Part 1 (Dissection) started at the beginning of the operation and ended at the beginning of the distal transection of the specimen (i.e. duodenum or stomach). Part 2 (Resection and Dissection) began with the distal transection and concluded with the resection of the entire specimen. This step also included the remaining of the dissection to include the necessary lymph node stations. Part 3 (Reconstruction) had either one step, in the case of a Billroth II (gastro-jejunostomy in distal gastrectomies) or EG (esophago-gastrostomy in proximal gastrectomies) anastomosis or had 2 steps in the case of Roux-En-Y reconstruction (jejuno-jejunostomy and esophago-jejunostomy or gastro-jejunostomy) in total and subtotal gastrectomies.

All transections were performed with endo-staplers. At the discretion of the surgeon, the transection of the duodenum, was performed with or without a reinforced cartridges. All anastomoses were performed with either a circular (EJ and EG anastomosis) or linear (JJ and GJ anastomosis) staplers. All linear anastomoses were closed with hand-sewn technique, using absorbable suture (polyglycolic acid) in one or two layers. Jackson-Pratt drains were not used routinely. Specimens were extracted through extension of one of the port sites or Pfannenstiel incision.

3.3.4 Data extraction

The assessment process in this study followed the current, conceptual Messick’s framework of validity, as a guide (Ghaderi et al., 2015). One expert rater reviewed all the video files. The expert rater, prior to reviewing the videos had undergone a structured program in video analysis. This process included a training curriculum, certification and calibration using the applied measurement instruments. The instruments used in this study, were the Objective Structured Assessments of Technical Skills (OSATS) (Martin et al., 1997) and the Generic Error Rating
Tool (GERT) (Bonrath, Zevin, et al., 2013). Both instruments have extensive evidence in the literature to support their use.

The video analysis was performed in a second to second time fashion, using the Operating Room Black Box® analysis platform (SST Inc., Toronto, ON). Video files contained laparoscopic view only, without any audio, patient or surgeon information. The rater was blinded to the origin of the video, the level of the operating surgeon, as well as to the patient outcome. For technical performance assessment, only the operating surgeon was rated.

Errors, events and their rectifications were recorded as they occurred in the procedure. An error was defined as a failure of a planned action (e.g.: slip of a bowel grasper) (Bonrath, Zevin, et al., 2013; Reason, 1995). An event was defined as any deviation that caused injury (e.g.: a slipped bowel grasper causing a serosal tear) (Bonrath, Zevin, et al., 2013; "WHO Draft Guidelines for Adverse Event Reporting and Learning Systems.," 2005). In the GERT instrument there are 9 different error groups (e.g. grasping/dissection, suturing) and 30 different error modes (e.g. too much force/distance, wrong orientation) (Bonrath, Zevin, et al., 2013). Intraoperative events are another parameter, identified and recorded using the GERT instrument (Bonrath, Zevin, et al., 2013). A rectification of an event was defined as an additional action or measure to prevent an adverse outcome (e.g., repair of the serosal tear, to prevent a potential leak) (Barach et al., 2008; Bonrath, Zevin, et al., 2013). A rectification of an event is not part of the GERT instrument and is recorded separately.

Separate OSATS scores were given for each stage of the procedure. A global OSATS score for the procedure was calculated based on the mean of these scores and was used to create high-performance (>29/35) and low-performance groups (≤ 29/35). The maximum possible score on OSATS is 35. The cut-off value (29/35) for the high- and low-performance groups was chosen based on previous reports, indicating that expert surgeons score at least 80% on the global OSATS scale in a various operations (Beard et al., 2007; Bonrath, Zevin, et al., 2013; Curry et al., 2012; Kumar et al., 2012).

### 3.3.5 Statistical analysis

All statistical analyses were performed using SPSS version 24.0 (IBM SPSS Statistics, IBM Corp., Armonk, NY, USA).
Normality of data distribution was assessed using the Shapiro-Wilk test. Descriptive statistics were performed.

Patients were dichotomized based on the outcome into i) no complications or ii) complications groups (Clavien-Dindo Grade ≥ III). Continuous variables were compared using either Student’s T or Mann-Whitney-U test, based on the data distribution. Categorical variables were compared using Chi-square ($\chi^2$) test of association. When cell size was <5, Fischer’s exact test was used.

Variables from the univariate analysis, with a significance of P<0.1, were included in a multivariate logistic regression analysis with backward selection strategy where the main outcome variable was short-term complications. Variables with P<0.05 were kept in the final model. Hosmer-Lemeshow goodness-of-fit test (P>0.05) was used to assess the fit of the model. Correlations between OSATS, GERT and events were calculated by Spearman’s correlation.

### 3.4 Results

#### 3.4.1 Patient factors

In the period of January 1st, 2009 and December 31st, 2015, 172 patients underwent laparoscopic gastrectomy for gastric cancer, at the three participating institutions, by the three surgeons.

A total of 61 (35.5%) patients had their procedure video recorded, stored in an unedited fashion and were available for review. Patient and disease characteristics in the two groups are compared in Table 13.

The pathological diagnosis of these 61 cases was gastric adenocarcinoma, intestinal type in 30 (49.2%); diffuse type in 16 (26.2%); mixed type in 5 (8.2%); type not specified in 6 (9.83%) and high grade dysplasia in 4 (6.6%) cases.

In 8 cases (13.1%) the recording was incomplete. In 6 cases, part of the procedure was performed open and in 2 cases the recording was started late. However, in all these 8 cases the video files contained sufficient data, for a meaningful evaluation, and were therefore included in the cohort.
Table 13 - Patient Factors

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No complications</th>
<th>Complications</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall N (%)</td>
<td>43 (70.5)</td>
<td>18 (29.5)</td>
<td></td>
</tr>
<tr>
<td>Age (y); mean [SD]</td>
<td>67.16 [12.83]</td>
<td>71.28 [9.86]</td>
<td>0.229</td>
</tr>
<tr>
<td>Male Sex; N (%)</td>
<td>21 (48.8)</td>
<td>13 (72.2)</td>
<td>0.094</td>
</tr>
<tr>
<td>BMI (kg/m$^2$); mean [SD]</td>
<td>25.05 [4.47]</td>
<td>25.71 [4.52]</td>
<td>0.646</td>
</tr>
<tr>
<td>CCI; mean [SD]</td>
<td>5.14 [1.96]</td>
<td>6.28 [1.84]</td>
<td>0.04*</td>
</tr>
<tr>
<td>Previous abdominal surgery; N (%)</td>
<td>16 (37.2)</td>
<td>9 (50)</td>
<td>0.354</td>
</tr>
<tr>
<td>Neoadjuvant chemotherapy; N (%)</td>
<td>14 (32.6)</td>
<td>3 (16.7)</td>
<td>0.207</td>
</tr>
<tr>
<td>Pathological stage; N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3 (7)</td>
<td>1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>3 (7)</td>
<td>2 (11.1)</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>11 (25.6)</td>
<td>7 (38.9)</td>
<td></td>
</tr>
<tr>
<td>IIA</td>
<td>7 (16.3)</td>
<td>1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>IIB</td>
<td>5 (11.6)</td>
<td>4 (22.2)</td>
<td>0.479</td>
</tr>
<tr>
<td>IIIA</td>
<td>5 (11.6)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>IIIB</td>
<td>3 (7)</td>
<td>1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>IIIC</td>
<td>3 (7)</td>
<td>1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>1 (2.3)</td>
<td>1 (5.6)</td>
<td></td>
</tr>
<tr>
<td>Stage n/a</td>
<td>2 (4.7)</td>
<td>0 (0)</td>
<td></td>
</tr>
</tbody>
</table>

SD = standard deviation; BMI = Body Mass Index; CCI = Charlson Comorbidity Index; *=statistically significant result
3.4.2 Surgical factors

The majority of the operations were partial (distal, subtotal, proximal) gastrectomies (N=37, 60.7%). One patient had a total gastrectomy and right hemicolectomy during the same operation for Lynch syndrome. The extent of the lymphadenectomies was D1+ in 40 cases (65.6%). The resection was R0 (negative margins) in 54 cases (88.5%). In partial gastrectomies, the majority of the reconstruction was Roux-en-Y (N=17, 45.9%), followed by Billroth II (N=14, 37.8%). The median duration of a surgery for partial gastrectomy was 3 hours and 33 minutes [interquartile range (IQR): 2 hrs 50 mins–4 hrs 18 mins]. For total gastrectomy, the median duration was 4 hours and 7 minutes (IQR: 3 hrs 16 mins–4 hrs 29 mins).

The mean global OSATS for the cohort was 28.4 (SD 1.7). There was statistically significant difference in the OSATS scores between the two groups of patients (p=0.02) (Table 14). A total of 3814 technical errors were observed in the 61 operations, with a median of 60 errors/case [IQR: 45-74]. The most common errors (N=1651, 43.3%) was slipping of tissue due to insufficient grasping force or undershooting with the instrument due to challenge in depth perception. This was followed by too much grasping force of the tissue (N=841, 22.1%) and "Wrong Orientation" (N=644, 16.9%). Finally "Inadequate Visualization" of the working instrument or operative field accounted for the remaining observations (N=617, 16.2%). Approximately half of the errors (51.9%) led to an event. A total of 1980 events were recorded, with a median of 30 events/case [IQR: 22-42]. The most common event was bleeding (N=1105, 55.9%), followed by thermal or mechanical tissue injury (N=844, 42.6%). Almost three-quarter (73.5%) of the events were rectified. Rectification happened 812 times, with a median of 11 rectifications/case [IQR: 6-20]. The most common rectification was hemostasis (N=749, 92.2%), followed by tissue repair (N=52, 6.4%). In the 2 patient groups there was no statistically significant difference between errors (p=0.331), events (p=0.758) and rectifications (p=0.433). There was a statistically significant strong correlation between OSATS and Errors (r_s= - 0.665, p<0.001) and moderate between OSATS and Events (r_s= - 0.401, p=0.001).
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>No complications</th>
<th>Complications</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall N (%)</td>
<td>43 (70.5)</td>
<td>18 (29.5)</td>
<td></td>
</tr>
<tr>
<td>Surgeon's prior experience; N (%)</td>
<td>36 (83.7)</td>
<td>16 (88.9)</td>
<td>0.713</td>
</tr>
<tr>
<td>Type of procedure; N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>14 (32.6)</td>
<td>10 (55.6)</td>
<td>0.094</td>
</tr>
<tr>
<td>Subtotal</td>
<td>18 (41.9)</td>
<td>2 (11.1)</td>
<td></td>
</tr>
<tr>
<td>Distal</td>
<td>7 (16.3)</td>
<td>4 (22.2)</td>
<td></td>
</tr>
<tr>
<td>Proximal</td>
<td>4 (9.3)</td>
<td>2 (11.1)</td>
<td></td>
</tr>
<tr>
<td>R0 resection; N (%)</td>
<td>37 (86.4)</td>
<td>17 (94.4)</td>
<td>0.318</td>
</tr>
<tr>
<td>Number of harvested lymph nodes; mean [SD]</td>
<td>20.28 [10.55]</td>
<td>24.78 [10.07]</td>
<td>0.129</td>
</tr>
<tr>
<td>Extent of lymphadenectomy, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D0</td>
<td>5 (11.6)</td>
<td>2 (11.1)</td>
<td></td>
</tr>
<tr>
<td>D1</td>
<td>9 (20.93)</td>
<td>3 (16.7)</td>
<td>0.805</td>
</tr>
<tr>
<td>D1+</td>
<td>27 (64.3)</td>
<td>13 (72.2)</td>
<td></td>
</tr>
<tr>
<td>D2</td>
<td>1 (2.3)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Type of reconstruction; N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B2</td>
<td>12 (27.9)</td>
<td>2 (11.1)</td>
<td></td>
</tr>
<tr>
<td>EG</td>
<td>4 (9.3)</td>
<td>2 (11.1)</td>
<td>0.291</td>
</tr>
<tr>
<td>RnYEJ</td>
<td>14 (32.6)</td>
<td>10 (55.6)</td>
<td></td>
</tr>
<tr>
<td>RnYGJ</td>
<td>13 (30.2)</td>
<td>4 (22.2)</td>
<td></td>
</tr>
<tr>
<td>Conversion to open; N (%)</td>
<td>4 (9.3)</td>
<td>2 (11.1)</td>
<td>1</td>
</tr>
<tr>
<td>Operative time (mins); median [IQR]</td>
<td>230 [185-265]</td>
<td>240 [189-261]</td>
<td>0.806</td>
</tr>
<tr>
<td>Global OSATS ≤ 29; N (%)</td>
<td>17 (39.5)</td>
<td>13 (72.2)</td>
<td>0.02*</td>
</tr>
</tbody>
</table>

Surgeon's prior experience = >30 cases; R0 resection = negative margins; SD = standard deviation; B2 = Billroth II; EG = Esophago-Gastrostomy; RnYEJ = Roux-en-Y Esophago-Jejunostomy; RnYGJ = Roux-en-Y Gastro-Jejunostomy; OSATS = Objective Structured Assessment of Technical Skills; *=statistically significant result
3.4.3 Postoperative complications

Eighteen patients (29.5%) developed serious complications (Clavien-Dindo Grade III-V). The mortality rate was 8.2 % (N=5). Four patients died after the initial operation, without being discharged from the hospital and one patient died upon readmission within 30 days. The morbidity rate was 21.3% (N=13). There were four reoperations (6.5%) due to anastomotic leak and three (5%) drain insertions by interventional radiology, two for intra-abdominal abscess and one for a pleural effusion.

In the no complications group there were 11 (25.6%) patients who developed transient postoperative nausea and vomiting (5 patients), acute kidney injury (one patient), urinary retention (one patient), pneumonia (two patients), urinary tract infection (one patient) and transient ischemic attack (one patient).

The median length of stay was 6 days [IQR: 5-12]. The median length of stay of patients with complications was significantly higher compared to patients with an uneventful postoperative recovery (5 days vs. 25 days, p<0.001).

Table 15 details the postoperative complications.
### Table 15 - Postoperative Complications

<table>
<thead>
<tr>
<th>Complications</th>
<th>N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mortality</strong></td>
<td></td>
</tr>
<tr>
<td>Anastomotic leak and intra-abdominal sepsis</td>
<td>3 (4.92)</td>
</tr>
<tr>
<td>Internal hernia and intestinal obstruction</td>
<td>1 (1.64)</td>
</tr>
<tr>
<td>Pulmonary embolism and anastomotic leak</td>
<td>1 (1.64)</td>
</tr>
<tr>
<td><strong>Morbidity</strong></td>
<td></td>
</tr>
<tr>
<td>Anastomotic leak</td>
<td>7 (11.5)</td>
</tr>
<tr>
<td>Intra-abdominal abscess</td>
<td>2 (3.28)</td>
</tr>
<tr>
<td>Single organ dysfunction (Respiratory)</td>
<td>3 (4.92)</td>
</tr>
<tr>
<td>Intra-abdominal bleeding</td>
<td>1 (1.64)</td>
</tr>
</tbody>
</table>
3.4.4 Predictors for postoperative complications

Univariate analysis showed that four variables (p<0.1) were candidates to enter the binomial logistic regression analysis, and these were Charlson Comorbidity Index (CCI), gender, type of procedure and the global OSATS score.

The mean CCI, type of procedure and the global OSATS score (≤29) were kept in the final model. Global OSATS score remained an independent predictor for postoperative complications (Odds Ratio = 6.49; 95% CI = 1.60-26.34, p= 0.009), while adjusting for comorbidities and type of procedure. The Hosmer and Lemeshow goodness-of-fit Test’s significance was p=0.396.

Table 16 presents the final predictive model.
<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>OR</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low OSATS</td>
<td>6.49</td>
<td>1.6</td>
<td>26.39</td>
</tr>
<tr>
<td>Total gastrectomy</td>
<td>6.67</td>
<td>1.47</td>
<td>30.36</td>
</tr>
<tr>
<td>Comorbidity (CCI)</td>
<td>1.64</td>
<td>1.1</td>
<td>2.45</td>
</tr>
</tbody>
</table>

OR = Odds Ratio; CI = Confidence Interval; OSATS = Objective Structured Assessment of Technical Skills; Low OSATS: ≤ 29; CCI = Charlson Comorbidity Index; *=statistically significant result
3.5 Discussion

The present study showed that technical performance predicted short-term outcomes in laparoscopic gastrectomy for cancer. Lower technical performance was an independent predictor for increased short-term postoperative complications.

Laparoscopic gastrectomy is a complex procedure, and patient outcomes are multifactorial. In the predictive model of the present study, besides technical performance scores, the comorbidities of the patient and the type of procedure performed were also significant predictors of short-term postoperative complications. These finding are in keeping with previous reports (M. C. Kim et al., 2008; W. Kim et al., 2008). Efforts aiming to improve outcomes should attempt to enhance all contributing factors, including pre-operative optimization, intra-operative surgical performance and evidence-based enhanced recovery programs (Wong-Chong, Kehlet, & Grantcharov, 2016).

There is a significant body of evidence suggesting that surgeons’ operative volume predicts patient outcomes in gastric cancer. In the Western series, surgeons’ volume is considered “medium” for between 9 – 13 cases/year and “high” between14 – 21 cases/year are performed (Birkmeyer et al., 2002; Enzinger et al., 2007). In the present study, all three participating surgeons were considered medium – high volume.

Surgeon’s experience, although intuitive, is not always synonymous with technical expertise. In fact, in the published literature, across various surgical subspecialties and procedures, the individual surgeons’ technical performance did not necessarily correlate with their experience (Duclos et al., 2012). Moreover, surgeon’s volume was a poor surrogate for technical skills (Rutegard et al., 2009). Also, as experience increased, operative time and length of stay might have decreased, but complication rate might have not (Woelk et al., 2013).

In laparoscopic gastric cancer surgery, previous reports suggest that surgical expertise is reached after 30 – 50 procedures (H. I. Kim et al., 2014; M. C. Kim et al., 2008; W. Kim et al., 2008; J. H. Lee et al., 2006; Ryu et al., 2008). In the present study two surgeons had the previous experience of >30 cases, and this variable was not a significant contributor to outcomes. The present study demonstrated that surgeons’ volume and previous experience did not have an effect on patient outcomes, emphasizing the fact that the use of surrogate variables might not be
accurate and the direct assessment of technical skills might be more valuable and warranted. Furthermore, the direct and objective assessment of technical performance and its improvement is important and necessary for all surgeons, irrespective of their level of training or years of practice (Hu et al., 2012). Introducing concepts that facilitate critical reflection on individual performance will serve as a powerful tool for quality improvement and will have a positive effect on patient care and safety (Rebasa et al., 2009). This approach has the potential to impact the clinical outcomes after laparoscopic gastrectomy for cancer, where the rate of postoperative complications is still substantial despite significant advancements in preoperative optimization and the introduction of evidence–based regimen for enhanced recovery (Li et al., 2011; Orsenigo et al., 2011; Stotland et al., 2009; Wong-Chong et al., 2016).

In the present study, technical performance was measured using OSATS (Martin et al., 1997) and GERT (Bonrath, Zevin, et al., 2013) instruments. GERT and OSATS measure different aspects of intraoperative performance and can serve as adjuncts to each other. This was demonstrated by the strong negative correlation between the two variables, which replicates the findings in the study by Bonrath et al. (Bonrath, Zevin, et al., 2013). OSATS provides a global performance analysis of the operating surgeon including respect for tissue, knowledge of instrument, efficiency and forward planning, knowledge of the procedure and the use of assistant. GERT is powerful tool for identifying intermediate safety outcomes such as type and number of errors made, and using the information for specific, individualized feedback. Currently, there is paucity of data to create normative standards for errors and adverse events in complex laparoscopic surgery. Previous literature suggests that for a simple procedure (e.g. laparoscopic cholecystectomy) experts make between 2-11 errors, whereas in a more complex procedure (e.g. laparoscopic gastric bypass operation) the number of errors could be 2-3 times higher (Bonrath, Zevin, et al., 2013; Seymour et al., 2004; Tang, Hanna, Joice, & Cuschieri, 2004). In the present study, the median number of errors was 60. Although there was no statistically significant difference between the two groups in terms of number of errors, events and rectification, the trend was favorable towards higher number of errors, events and rectification in the complication group. A possible explanation for these findings could be that at the present time all errors and events are weighted equally and the original GERT framework does not take into account the severity of the observed errors and events. This is a topic with potential for future research and investigations and studies are underway.
The results of this study have to be interpreted while keeping several limitations in mind. First, this was a retrospective design with limited number of patients. This could be another explanation why the trend for some of the variables (i.e. errors, events and rectifications) were favorable, but did not reach statistical significance. Second, not all consecutive cases were recorded and therefore not analyzed. This might have introduced a selection bias and skewed the results, however every effort was made to include all available cases. Previous study based on a similar cohort, but consecutive patients, at one of the participating sites reported similar overall major complication rates (27.9 %), however lower mortality rates (2.3%) (Wong-Chong et al., 2016). An explanation for this discrepancy could be that, although there were not specific guidelines for video-recording, perhaps it was more likely to record and store procedures where the surgeon expected technical challenges or difficulties within the case and wanted to save the procedure for future educational purposes. Third, it is impossible to assess the level of the operating surgeon when looking at laparoscopic view only. In this study this information remained unknown (staff vs. fellow vs. senior resident), since the recordings were stored without it. Therefore, it is possible that fellows and residents performed steps of the operations as well and this might have influenced the global OSATS score. However, a recent survey done at the same university, where this study was conducted, showed that trainees do not routinely perform the major steps of these operations (Fecso, Bonrath, & Grantcharov, 2016). Moreover, the main aim of the current study was to assess technical performance itself and not necessarily the level of the operator. Finally, in this environment not only the technical performance of the individual surgeon plays a role in a patient outcome, but the non-technical performance of the OR (operating room) team (A. A. Gawande et al., 2003; Greenberg et al., 2007; Sevdalis et al., 2014). In order to assess all potentially contributing variables and determine their relationships, the recordings or live observations have to be more detailed, not just limited to a laparoscopic view. This could be achieved with a comprehensive video and audio analysis of the entire operating room environment. Studies to address these research questions are currently underway.

Despite the limitations, this study performed a robust, methodologically sound and detailed assessment of technical performance, and identified one of the missing variables that can potentially affect patient outcomes and more importantly can potentially be modified and improved. Future directions include (1) the ongoing assessment of the intraoperative environment, individuals and teams and their interaction and performances using a
comprehensive and synchronized surgical data capture platform; (2) the enhancement of surgical skills, using modern strategies such as structured feedback, deliberate practice, coaching etc., at all levels with the ultimate goal to improve patient outcomes.

3.6 Conclusion

In conclusion, intraoperative technical performance predicts short-term complications in laparoscopic gastrectomy for cancer. The direct measurement of the variable of interest might be more valuable than the measurement of its surrogates. Future work should focus on assessing the entire operating room environment and studying the effectiveness of individualized, targeted educational interventions in laparoscopic gastric cancer surgery.
CHAPTER 4

The purpose of this chapter is to:

I. Explore the current intraoperative educational environment at a large university center
II. Determine how surgeons distribute, amongst trainees, the main steps of a laparoscopic gastric cancer operation
III. Identify if the opinions of trainees and faculty match in terms of operative experience

The study in this chapter was published as an article, entitled “Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges, and Potential Opportunities at a Large University Centre” by Fecso AB, Bonrath EM and Grantcharov TP, Journal of Surgical Education 2016 Jul-Aug; 73(4): 749-55. Reprinted with kind permission from Elsevier. Formatting has been adapted for the purpose of this thesis.
4 Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges and Potential Opportunities at a Large University Centre

4.1 Abstract

4.1.1 Objective

The purpose of this study was to explore and understand how surgeons distribute tasks during a laparoscopic gastrectomy for gastric cancer in an academic teaching environment.

4.1.2 Design

An anonymous, cross-sectional, census survey was used to poll trainees’ and staff members’ opinions pertaining to laparoscopic gastrectomy.

4.1.3 Setting

Academic and community tertiary teaching hospitals, affiliated with the University of Toronto

4.1.4 Participants

All surgeons, within the Department of General Surgery at the University of Toronto, who practice laparoscopic gastrectomy for gastric cancer, were invited to participate. All general surgery residents, postgraduate year 1-5, minimally invasive surgery and surgical oncology fellows at the University of Toronto were invited to participate. Overall response rate was 74.35% (n=87/117).

4.1.5 Results

The results suggested that trainees do not routinely perform the major operative steps. Trainees agreed with faculty in this regard, however there was a statistically significant difference in
opinions, related to the degree of the perceived active operating of the trainees. There was also a difference in opinion, between trainees and faculty, regarding the common reasons for take over.

4.1.6 Conclusions

The present survey highlights that current level of active exposure of surgical trainees to laparoscopic gastric surgery might be insufficient. A lack of role clarity may further hinder an optimal educational experience during these cases. Adopting a step-wise approach, with task deconstruction, could optimize training. Additional training modalities may be required to ensure technical proficiency is acquired prior to independent practice.
4.2 Introduction

Work hour restrictions ("ACGME Highlights Its Standards on Resident Duty Hours ", 2001; "Council Directive 93/104/EC.," 1993; Leach, 2004), constant demand for increased operating room efficiency, medico-legal concerns (K. R. Reznick & MacRae, 2006) have resulted in reduction in operative exposure of residents and brought significant challenges in training in complex and uncommon procedures. These pressures have lead to an increased emphasis on virtual education (Vollmer et al., 2011). While simulation and didactics can enhance the learning of surgical skills, the operating room remains a critical and unique environment, where trainees acquire not only their skills, but their judgment and knowledge (Butvidas, Anderson, Balogh, & Basson, 2011; Procter et al., 2010; K. R. Reznick & MacRae, 2006). Therefore, it is paramount that the intraoperative learning is as efficient as possible and that all cases are included in the trainees’ learning portfolio. There are multiple training models described in the literature that are used to enhance residents’ intraoperative learning (DaRosa et al., 2013; Marangoni, Morris-Stiff, Deshmukh, Hakeem, & Smith, 2012), however the translation of this evidence into clinical practice is still suboptimal.

Laparoscopic gastrectomy for gastric cancer is considered a complex surgical procedure that requires high level of skill and knowledge in order to achieve a successful short- and long-term outcome. The experience of the Western surgeon and the exposure of trainees are impacted by a number of factors, such as lower incidence of gastric cancer in the Western hemisphere (Rosin et al., 2009), different body habitus and associated comorbidities of the western patients (H. H. Kim & Ahn, 2011; Lianos et al., 2014), the majority of gastrectomies being performed at non-referral centers (Birkmeyer et al., 2002; Smith, Elting, Learn, Raut, & Mansfield, 2007; Yoon & Yang, 2009) and frequently patients present initially with more advanced disease (Lianos et al., 2014). Nevertheless, as experience in advanced laparoscopy continues to grow (Katsios, Baltogiannis, & Roukos, 2010), laparoscopic treatment of gastric cancer is here to stay. Therefore the importance of standardized and structured laparoscopic training and competence are imperative (H. H. Kim & Ahn, 2011).

The first step towards standardizing the teaching of this complex procedure is to explore and understand the current learning environment and how this complex procedure is integrated in educational concepts and models.
Therefore, the objectives of the current study were (1) to determine how surgeons distribute tasks during a laparoscopic gastrectomy for gastric cancer in an academic teaching environment; (2) to explore how trainees perceive their involvement in this particular procedure and (3) to assess whether staff and trainees perceptions are similar.

4.3 Methods

Institutional ethics approval was obtained prior to starting this anonymous, cross-sectional, census survey study.

4.3.1 Survey Instrument

To successfully complete a laparoscopic gastrectomy, a series of operative sub-steps need to be completed by the operating surgeon. A survey instrument, containing these operative sub-steps (14 in total), was designed using consensus by 2 expert surgeons in laparoscopic foregut surgery. These 14 operative sub-steps formed the core part of the survey. A set of questions were designed, to explore how often trainees perform these sub-steps. The participants were asked to answer the questions using a 4-point scale (1=Yes, all the time; 2=Yes, most of the time; 3=Yes, sometime; 4=No). The answer choices were the same for all the 14 core questions. A further set of questions polled opinions regarding role during surgery and perceived reasons for surgical take-over. The rest of the questionnaire was modified to fit the profile of the participants (staff vs. trainee). A free-text comment field was also incorporated at the end of the survey.

4.3.2 Study population

All surgeons, within the Department of General Surgery at the University of Toronto (UofT), who practice laparoscopic gastrectomy for gastric cancer, were invited to participate (Staff). Surgeons represented both university and university-affiliated, teaching community hospitals. All general surgery residents, postgraduate year (PGY) 1-5, minimally invasive surgery (MIS) and surgical oncology fellows at the University of Toronto were invited to participate (Trainees).
4.3.3 Survey process

An e-mail invitation was sent out to all potential participants. A link to the online platform (SurveyMonkey, Menlo Park, CA) was included in the e-mail. The participants’ responses were collected anonymously. Reminder e-mails were sent to all participants, on a weekly basis, for 3 consecutive weeks. To assure that every trainee had the opportunity to participate in this type of procedure, they were surveyed in June, at the end of the academic year.

4.3.4 Data analysis

The results were analyzed using SPSS software v. 22.0 (IBM SPSS Statistics, IBM Corp., Armonk, NY, USA). Descriptive statistical analysis was performed. Results are shown as frequency counts (%) and medians (range) due to the ordinal nature of the data. Group comparisons were performed, using Mann-Whitney U test. Statistical significance between the groups was set to a $P < .05$.

4.4 Results

4.4.1 Respondents

A total of 117 participants were invited, 20 Staff members and 97 Trainees (84 residents and 13 fellows). The overall response rate was n=87 (74.35%). Nineteen surgeons (95%) and 68 trainees (70.1%) returned the survey. Staff surgeons’ responses revealed that the majority of respondents routinely supervised trainees at all levels. Seventeen staff surgeons (89.5%) reported routinely supervising fellows, 19 (100%) reported supervising seniors (PGY: 3-5), and 15 (78.9%) routinely worked with juniors (PGY: 1-2). Three surgeons (15.8%) reported that they did not routinely work with juniors whereas one (5.3%) reported never working with junior trainees. For questions pertaining to supervising a specific trainee cohort, responses from the staff member that stated never to work with juniors were excluded from analysis of the subsection on junior trainees.

Out of the 68 trainee responses, 2 participants returned incomplete surveys, missing more than 80% of the questions; therefore, although they were included in the response rate, they were
excluded from final analysis (n=66 trainees included). Thirty-four trainees (51.5%) reported having participated in a laparoscopic gastrectomy for gastric cancer during their training. The distribution of trainee respondents according to their training level was as follows: three PGY 1 (33.3%), three PGY 2 (42.9%), eight PGY 3 (36.4%), six PGY 4 (50%), eight PGY 5 (88.9%) residents and six fellows (85.7%). There was a gradual increase in cumulative operative exposure as residents progressed through their training. Trainees participated in laparoscopic gastrectomy at various levels of their training. As expected all of the juniors (n=6; 100%) participated as juniors, however three of the seniors (13.6%) participated in their junior years only, and three of the fellows (50%) participated in their senior years only.

### 4.4.2 Intraoperative roles

For intraoperative responsibility as well as for group comparisons, in terms of operative sub-steps, trainees' reported scores were adjusted for the highest level of training at which they reported having participated in the case, adjusting the number of juniors to 9, seniors to 22 and fellows to 3.

All junior trainees (100%), 15 of the senior residents (68.2%) and 1 fellow (33%) reported having participated in less than 5 cases. Five seniors (22.7%) and 1 fellow (33.3%) reported having participated in 5-10 cases and only a small portion of seniors (n=2; 9.1%) and 1 fellow (33.3%) reported participation in 11-20 cases.

Perceived intraoperative responsibility varied based on the level of training. Juniors reported having played a minimal role, i.e.: during access and closure (n=5, 55.5%); camera driving (n=2, 22.2%); and assisting (n=2, 22.2%). Senior residents stated that they observed (n=1; 4.5%) while scrubbed; had minimal role, i.e.: access and closure (n=2, 9.1%); drove the camera (n=3, 13.6%); mainly assisted (n=6, 27.2%); had a mix of operating and assisting (n=7, 31.8%); or mainly operated (n=3, 13.6%). One of the fellows (33.3%) reported mainly assisting, while the other two (66.7%) reported a mix of operating and assisting.

Level of active participation, as viewed and reported by the trainee participants, was compared to the views of staff members for each operative sub-step (Table 17-Table 19). Significant discrepancies in opinions were noted for both the junior and senior cohorts.
Data reported by participants at fellow level (Table 19) were too few (n=3) to conduct a meaningful, between groups, analysis.
<table>
<thead>
<tr>
<th>Steps</th>
<th>Junior Resident level Median (range)</th>
<th>Staff Median (range)</th>
<th>P-Value (Mann-Whitney U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>2 (1.5-3)</td>
<td>2 (1-3)</td>
<td>1</td>
</tr>
<tr>
<td>Exposure</td>
<td>3 (3-4)</td>
<td>3 (3-4)</td>
<td>0.820</td>
</tr>
<tr>
<td>Adhesiolysis</td>
<td>3 (3-4)</td>
<td>3 (3-3)</td>
<td>0.128</td>
</tr>
<tr>
<td>Exploration</td>
<td>4 (3.5-4)</td>
<td>3 (2.75-3)</td>
<td><strong>0.004</strong></td>
</tr>
<tr>
<td>Dissection</td>
<td>4 (3.5-4)</td>
<td>4 (3-4)</td>
<td>0.653</td>
</tr>
<tr>
<td>Vascular control</td>
<td>4 (3.5-4)</td>
<td>4 (4-4)</td>
<td>0.405</td>
</tr>
<tr>
<td>Lymph node dissection</td>
<td>4 (4-4)</td>
<td>4 (4-4)</td>
<td>0.480</td>
</tr>
<tr>
<td>Resection</td>
<td>4 (4-4)</td>
<td>4 (3-4)</td>
<td>0.223</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>4 (4-4)</td>
<td>4 (3-4)</td>
<td>0.320</td>
</tr>
<tr>
<td>Inspection</td>
<td>4 (3.5-4)</td>
<td>3 (3-4)</td>
<td>0.180</td>
</tr>
<tr>
<td>Hemostasis</td>
<td>4 (4-4)</td>
<td>4 (3-4)</td>
<td>0.434</td>
</tr>
<tr>
<td>Removal of Specimen</td>
<td>4 (3-4)</td>
<td>3 (2.75-4)</td>
<td>0.152</td>
</tr>
<tr>
<td>Final Exploration</td>
<td>4 (4-4)</td>
<td>3 (3-4)</td>
<td><strong>0.038</strong></td>
</tr>
<tr>
<td>Closure</td>
<td>2 (1-3)</td>
<td>2 (2-3)</td>
<td>0.498</td>
</tr>
</tbody>
</table>

Response options on a 4 point scale: 1=Yes, all the time; 2=Yes, most of the time; 3=Yes, sometime; 4=No
Table 18 - Senior Residents and Staff responses on case sharing

<table>
<thead>
<tr>
<th>Steps</th>
<th>Senior Resident level Median (range)</th>
<th>Staff Median (range)</th>
<th>P-Value (Mann-Whitney U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>1 (1-2.25)</td>
<td>1 (1-2)</td>
<td>0.885</td>
</tr>
<tr>
<td>Exposure</td>
<td>2.5 (2-3.25)</td>
<td>2 (2-3)</td>
<td>0.225</td>
</tr>
<tr>
<td>Adhesiolysis</td>
<td>2 (2-3)</td>
<td>2 (2-3)</td>
<td>0.281</td>
</tr>
<tr>
<td>Exploration</td>
<td>2.5 (1-4)</td>
<td>2 (2-3)</td>
<td>0.308</td>
</tr>
<tr>
<td>Dissection</td>
<td>3 (2-4)</td>
<td>3 (2-3)</td>
<td>0.207</td>
</tr>
<tr>
<td>Vascular control</td>
<td>4 (2.75-4)</td>
<td>3 (2-3)</td>
<td>0.019</td>
</tr>
<tr>
<td>Lymph node dissection</td>
<td>4 (3-4)</td>
<td>3 (3-4)</td>
<td>0.103</td>
</tr>
<tr>
<td>Resection</td>
<td>3 (3-4)</td>
<td>3 (2-3)</td>
<td>0.012</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>3.5 (2-4)</td>
<td>3 (2-3)</td>
<td>0.011</td>
</tr>
<tr>
<td>Inspection</td>
<td>3 (2-4)</td>
<td>2 (2-3)</td>
<td>0.028</td>
</tr>
<tr>
<td>Hemostasis</td>
<td>3 (2-4)</td>
<td>2 (2-3)</td>
<td>0.017</td>
</tr>
<tr>
<td>Removal of Specimen</td>
<td>2 (2-3.25)</td>
<td>2 (1-2)</td>
<td>0.031</td>
</tr>
<tr>
<td>Final Exploration</td>
<td>2.5 (2-3.25)</td>
<td>2 (1-3)</td>
<td>0.133</td>
</tr>
<tr>
<td>Closure</td>
<td>1 (1-2)</td>
<td>2 (1-2)</td>
<td>0.667</td>
</tr>
</tbody>
</table>

Response options on a 4 point scale: 1=Yes, all the time; 2=Yes, most of the time; 3=Yes, sometime; 4=No
<table>
<thead>
<tr>
<th>Steps</th>
<th>Fellow level Median (range)</th>
<th>Staff Median (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>1 (1-1)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>Exposure</td>
<td>1 (1-1)</td>
<td>2 (1-2)</td>
</tr>
<tr>
<td>Adhesiolysis</td>
<td>1 (1-1)</td>
<td>2 (1-2)</td>
</tr>
<tr>
<td>Exploration</td>
<td>1 (1-1)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>Dissection</td>
<td>2 (1-2)</td>
<td>2 (2-3)</td>
</tr>
<tr>
<td>Vascular control</td>
<td>2 (2-2)</td>
<td>2 (2-3)</td>
</tr>
<tr>
<td>Lymph node dissection</td>
<td>3 (2-3)</td>
<td>3 (2-3)</td>
</tr>
<tr>
<td>Resection</td>
<td>2 (2-2)</td>
<td>2 (2-3)</td>
</tr>
<tr>
<td>Reconstruction</td>
<td>2 (2-2)</td>
<td>2 (2-3)</td>
</tr>
<tr>
<td>Inspection</td>
<td>1 (1-1)</td>
<td>2 (1-2)</td>
</tr>
<tr>
<td>Hemostasis</td>
<td>1 (1-1)</td>
<td>2 (2-2)</td>
</tr>
<tr>
<td>Removal of Specimen</td>
<td>1 (1-1)</td>
<td>1 (1-2)</td>
</tr>
<tr>
<td>Final Exploration</td>
<td>1 (1-1)</td>
<td>2 (1-2)</td>
</tr>
<tr>
<td>Closure</td>
<td>1 (1-1)</td>
<td>1 (1-2)</td>
</tr>
</tbody>
</table>

Response options on a 4 point scale: 1=Yes, all the time; 2=Yes, most of the time; 3=Yes, sometime; 4=No
4.4.3 Reasons for take-over

Eight of the respondents reporting participation as juniors (88.9%) completed this part of the survey and 6 of them (75%) stated that in their view, the main reason for staff taking over the case was their technical skills. Nineteen respondents reporting participation as seniors (86.36%) completed this part of the survey. Their answers were distributed between the options (skill level, case difficulty, time restraints or other reasons) and only one senior thought (5.3%) that it was only due to lacking technical skills that staff would take over the case. Of the respondents reporting participation as fellows, technical skills (n=1), time constraint (n=1) and a combination of case difficulty and time constraint (n=1) accounted for reasons of take-over.

More than half of the surgeons (n=11, 57.9%) reported the main reason for taking over the case, regardless of training level, was a combination of technical skills, case difficulty, and time constraints. Other (free text entries) perceived reasons trainees reported for not operating were the presence of fellows, staff inexperience and the fact that juniors were not afforded a chance to try the operation. Other reasons reported by staff surgeons included their own perceived inexperience in this type of procedure, patient safety issues, previous complications, to avoid problems, and that juniors were felt not be at a level to perform any of the steps of the operation, or that they didn’t even scrub most of the time.
4.5 Discussion

This study reports the results of a survey among faculty and trainees at the University of Toronto to evaluate the current practices for trainee exposure to laparoscopic gastrectomy for gastric cancer. The results suggest that trainees do not routinely perform the major steps of the surgery and both faculty and trainees tend to agree on this fact. However, when it comes to the level of active involvement, there are statistically significant differences in opinions.

The good response rate (74.35%), which is in accordance with the recommendations in the literature (Asch, Jedrziewski, & Christakis, 1997; Creswell, 2015: 378-425; Price, Murnan, Dake, Dimming, & Hayes, 2004), confirms the relevance of the topic and the need for a constructive discussion both from faculty and trainees’ perspective.

More than half of the trainees were exposed to laparoscopic gastrectomy for gastric cancer, at various stages throughout their training, therefore it is imperative that even these complex operations to be included in the portfolio of teaching procedures. Multiple studies have demonstrated that laparoscopic gastrectomy for gastric cancer is a safe and effective procedure in specialized centers (Galvani, Choh, & Gorodner, 2010; H. H. Kim et al., 2010; Rosin et al., 2009; Sokolich, Galanopoulos, Dunn, Linder, & Jeyarajah, 2009). At the same time, the topic of the impact of resident involvement in complex cases on postoperative patient outcomes has been investigated in several reports with conflicting results. Some studies have shown an increase in the operating time (Iannuzzi et al., 2013; Tokunga et al., 2011), increase of postoperative morbidity (Castleberry et al., 2013; Iannuzzi et al., 2013). Others have reported no difference in postoperative morbidity (Paisley, Madhavan, Paterson-Brown, Prassdom, & Garden, 1999; Tokunga et al., 2011), an improved rescue rate with corresponding decrease in mortality (Castleberry et al., 2013).

Regardless, the surgical workplace remains a valuable learning environment and should be used efficiently for delivery of safe and relevant education (Monkhouse, 2010). In this vein, clinical practice should enable residents to actively participate in the appropriate steps of any complex surgical intervention. Previous concerns about patient consent have been dismissed in reports demonstrating that most patients understand the need to facilitate trainee exposure and are not against resident participation in a supervised clinical setting (Berg, Engel, Saba, & Hatton, 2011;
In the present study, a substantial number of trainees reported that they have been exposed to laparoscopic gastrectomy for cancer at a junior training level, despite its complexity. As expected, there was a gradual increase in cumulative operative exposure as residents progressed through their training. Interestingly, approximately one third of the senior residents who participated in laparoscopic gastrectomies reported having done so already as juniors, but none of the fellows polled reported participation in junior years of training. This finding could have been due to either a culture shift over the last five training years, where surgeon-educators may have made a greater conscious effort to involve junior residents in the operating room, or explained by the fact, that some of the fellows might have come from different residency programs, where these procedures were distributed differently.

Several approaches have been described in an attempt to increase the active participation in the operating room for trainees at all stages (DaRosa et al., 2013; Marangoni et al., 2012). One of these approaches is the graded training approach, including task deconstruction. Marangoni et al. (Marangoni et al., 2012) showed that, although pancreatoduodenectomy is considered as one of the most technically challenging abdominal procedures, if deconstructed into sub-steps of different complexity, than it could offer a learning opportunity to trainees of various levels and allow a more efficient use of the OR (operating room) environment for education. Another approach is the "Zwisch model", which follows the principles of graded responsibilities and was developed, to teach cardiothoracic residents to operate independently (DaRosa et al., 2013).

Junior residents reported that they had minimal active participation during the operations. This is in agreement with the opinion of the faculty members and confirmed that both staff and juniors had a similar perception of the role junior trainees should play during these cases. From the survey data, it seems that the junior cohort is involved in steps, appropriate to their training level. However, the majority of the junior residents did not feel routinely involved in active teaching/learning process during the case. The survey did not include questions related to briefing, intraoperative communication and debriefing, but perhaps extra effort needs to be invested from the staff to engage them in the intra-operative planning and decision making even when they are not involved in the execution. This passive role is an important step in resident learning, titled – “show and tell” in the Zwisch model (DaRosa et al., 2013), that represents the first step in learning an operation. This first step is of particular importance in cases that are not performed in high volumes. A trainee may be exposed to a particular procedure during one
specific rotation and then not have a further opportunity for several rotations, thus each procedure must be used as an appropriate learning venue. More than 70% of the senior residents reported performing more active tasks, such as assisting, operating and a mix of the two. This level of activity would correspond to the “smart help and dumb help” stages, in the Zwisch model (DaRosa et al., 2013).

Senior residents also agreed with the faculty, that they are not routinely completing the "critical steps" of the procedure, such as dissection, vascular control, lymph node dissection, resection and reconstruction. However, there was a statistically significant difference in opinions, in between senior residents and faculty members, related to the degree of the perceived active operating of the trainees. Similar discrepancies in opinions, pertaining to the quality of the preoperative preparation, intraoperative teaching and postoperative feedback have also been described in the literature (Butvidas et al., 2011; Levinson, Barlin, Altman, & Satin, 2010; Rose, Waibel, & Schenarts, 2011; Vollmer et al., 2011). Surprisingly, none of the fellows reported mainly participating in these procedures as the operating surgeon. At the same time, senior residents reported a common perception that fellows take valuable training opportunities away from them and that they limit their operative exposure. Yet, even on the fellowship level, trainees did not report routinely performing the critical steps of the surgery, such as dissection, vascular control, lymph node dissection and resection. This discrepancy in views of operative opportunity between trainees of differing levels reflects a lack of clear agreement of roles and responsibilities and highlights the need for an open discussion on operative roles and expectations during the pre-operative briefing. Roberts et al. (Roberts, Brenner, Williams, Kim, & Dunnington, 2012) suggests that having a structured approach to an operation, starting with preoperative task deconstruction with clear communication to clarify roles, objectives and expectations, followed by intraoperative teaching and postoperative feedback could potentially maximize education at all levels.

Surgeon educators have a responsibility towards the trainee to provide education and towards the patient to provide high quality and safe care and keeping a healthy balance between the two might be occasionally challenging. For steps that are not routinely performed by trainees, alternative training opportunities, such as international collaborations, simulation training including virtual reality, animal or cadaveric surgical models (H. H. Kim & Ahn, 2011) should
be utilized as there is sufficient evidence that these tools are effective in achieving psychomotor proficiency prior to OR involvement (Bagai et al., 2012; Palter & Grantcharov, 2012a).

Staff respondents felt that, regardless of the level of the trainee, the main reason for take-over during a case was a combination of the trainee’s skill level, case difficulty and time constraints, however trainees perceived this differently. Junior trainees perceived their skills level as the most common reason for a take-over. More senior trainees recognized that other factors seemed to play a role including case difficulty and time constraint or other factors.

Several limitations require discussion in the context of the present work. First, only one training program was surveyed; therefore, the data might not be generalizable to other programs due to potential different educational and clinical practices. Second, although the overall response rate was high, the fellow group was under represented as well as only a small number of them reported completing the procedure during their fellowship, limiting any statistical analysis for this cohort. Third, the present study looked at self-reported data, subjective measurements based on recent or distant memory recollection. Future work will focus on objective recording of the level of participation, performance and educational challenges and opportunities using comprehensive video analysis.

4.6 Conclusion

The present survey highlights that current level of active exposure of surgical trainees to laparoscopic gastric surgery might be insufficient. A lack of role clarity may further hinder an optimal educational experience during these cases. Adopting a step-wise approach, with task deconstruction, could optimize training. Additional training modalities may be required to ensure technical proficiency is acquired prior to independent practice.
CHAPTER 5

The purpose of this chapter is to:

I. Investigate the intraoperative environment using a novel approach
II. Determine how intraoperative technical and non-technical (intrinsic) performances are related

The study in this chapter has been accepted for publication in British Journal of Surgery (BJS-1409-Oct-17) with the title “The Relationship Between Intraoperative Non-Technical Performances and Technical Events in Bariatric Surgery” by Fecso AB, Kuzulgil SS, Babaoglu C, Bener AB and Grantcharov TP.

Formatting has been adapted for the purpose of this thesis.
5 The Relationship Between Intraoperative Non-Technical Performances and Technical Events in Bariatric Surgery

5.1 Abstract

5.1.1 Introduction

The operating room is a unique environment with complex team interactions, where technical and non-technical performance affects patient outcomes. The correlation between technical and non-technical performance however remains under-investigated. The purpose of this study was to explore these interactions in the operating room.

5.1.2 Methods

A prospective, single center observational study was conducted at a tertiary academic medical center. One surgeon and three fellows participated as main operators. All patients who underwent a Laparoscopic Roux-en-Y Gastric Bypass operation and had their procedures captured using the OR Black Box (ORBB) platform were included. Technical assessment was performed using OSATS (Objective Structured Assessments of Technical Skills) and GERT (Generic Error Rating Tool) instruments. For non-technical assessment, the NOTSS (Non-Technical Skills for Surgeons) and SPLINTS (Scrub Practitioners’ List of Intraoperative Non-Technical Skills) tools were used. Spearman’s rank-order correlation and N-gram statistics were conducted.

5.1.3 Results

Fifty-six patients were included in the study and 90 procedural steps (Gastro-Jejunostomy and Jejuno-Jejunostomy) were analyzed. There was a moderate to strong correlation between technical adverse events and non-technical performance of the surgical and nursing teams (intraoperative events \( r_s = 0.417 – 0.687 \), rectifications \( r_s = 0.380 – 0.768 \) and NOTSS and SPLINTS). N-gram statistics showed that following technical errors, events, and prior
rectifications, the staff surgeon and the scrub nurse exhibit the most positive non-technical behaviours, regardless of the operator (Staff or Fellow).

5.1.4 Conclusions

This study demonstrated that technical and non-technical performances are related, both on individual and team level. Valuable data can be obtained around intraoperative errors, events and rectifications.
5.2 Introduction

Team dynamics make the operating room a unique environment as professionals from multiple disciplines work together in a closely coordinated fashion, with one common goal (D'Addessi et al., 2009).

A successful surgical outcome is a result of many individual, team and organizational factors. At the same time there is emerging evidence emphasizing the importance of both technical (Birkmeyer et al., 2013; Fecso, Szasz, et al., 2017) and non-technical performance (Mazzocco et al., 2009; Mishra, Catchpole, Dale, & McCulloch, 2008; Weldon, Korkiakangas, Bezemer, & Kneebone, 2013). Reports have indicated that most errors in the operating room are technical in nature (A. A. Gawande et al., 1999; A. A. Gawande et al., 2003; Rogers et al., 2006) and technical performance might affect patient outcomes more than the perioperative management (Birkmeyer et al., 2013). Furthermore, human factors and non-technical performance, such as communication and teamwork, distractions, excessive workload and fatigue, have been identified as potential contributors to surgical errors, complications and patient harm (A. A. Gawande et al., 2003; Greenberg et al., 2007; Lingard et al., 2004; Mazzocco et al., 2009; Sevdalis et al., 2014). Although both technical and non-technical performances are important contributors to a successful patient outcome, the interaction between the two remains under-investigated.

The purpose of this study was to explore the relationship between technical and non-technical performance of the surgical and nursing teams during surgical procedures.

5.3 Methods

Institutional ethics approval was obtained prior to commencing the study.

5.3.1 Patients

This was a prospective, observational, single-center cohort study at a tertiary academic medical center. All adult patients who underwent a Laparoscopic Roux-En-Y Gastric Bypass (LRYGB) between April 1st 2014 and Dec 31st 2015 were eligible for inclusion. All operations with valid
consents were recorded and analyzed for both technical and non-technical performance. Patients’ charts were reviewed for patient demographics and comorbidities.

### 5.3.2 Operating Room Participants

One staff surgeon (S) and three fellows (F) participated in the study as the main operating physicians. Residents’ performance was not included in this study. The nursing team participated as well and consisted of experienced, staff Scrub (SN) and Circulating Nurses (CN).

### 5.3.3 Surgical Technique

All LRYGB operations were performed in a standardized manner. All procedures were divided into 2 phases: jejuno-jejunostomy (JJ) and gastro-jejunostomy (GJ). The JJ part started with the initiation of the procedure and ended with the closure of the JJ anastomosis. The GJ part started after the closure of the JJ anastomosis and ended with the conclusion of the operation. To be included in the study, at least one part had to be performed either by the surgeon or one of the fellows. For the purpose of this study, every part (JJ vs. GJ) was considered and analyzed as a separate observation.

### 5.3.4 Data Collection and Analysis

Technical and non-technical performance of the surgical and nursing teams were evaluated using panoramic view of the operating room including high-definition audio and a laparoscopic camera view. The video analysis was performed in a second to second time fashion on synchronized timelines, using the Operating Room Black Box® (ORBB) platform (SST Inc., Toronto, ON).

The assessment method in this study used the Messick’s conceptual framework, as a guide (Ghaderi et al., 2015). Four expert raters participated in the video reviews. Prior to starting the review process, all raters underwent detailed training in the use of the instruments and were calibrated to another expert rater as well as to each other. To keep the inter-rater reliability constant, an ongoing recalibration process was performed.

For rating technical performance the Objective Structured Assessments of Technical Skills (OSATS) (e.g. instrument handling, flow of operation, etc.) (Martin et al., 1997) and the Generic
Error Rating Tool (GERT) (e.g. slip of an instrument, inadequate visualization, etc.) (Bonrath, Zevin, et al., 2013) instruments were used. For technical performance, the operator received a score out of a maximum 35, as per OSATS. Separate global OSATS scores were generated for every JJ and GJ parts. The errors, events and rectifications were recorded and time-stamped as they occurred. An error was defined as a failure of a planned action (e.g.: manipulation of non-target tissue or without adequate visualization) (Bonrath, Zevin, et al., 2013; Reason, 1995). An event was defined as any deviation that caused injury (e.g.: mechanical or thermal tissue injury) (Bonrath, Zevin, et al., 2013; "WHO Draft Guidelines for Adverse Event Reporting and Learning Systems.," 2005). A rectification of an event, or event compensation was defined as an additional action or measure to prevent an adverse outcome (e.g: repair of a tissue injury) (Barach et al., 2008). In order to minimize bias, the first step was the assessment of the technical performance (laparoscopic view only without audio, blinded for the identity and rank of the operating surgeon), followed by the non-technical assessment (panoramic view of the OR and audio).

For non-technical performance, the Non-Technical Skills for Surgeons (NOTSS) (e.g. situation awareness, decision making, etc.) (Yule et al., 2006) and the Scrub Practitioners’ List of Intraoperative Non-Technical Skills (e.g. task management, communication and teamwork, etc.) (SPLINTS) (Mitchell et al., 2013) tools were used. Positive (desirable) (i.e.: surgeon anticipating future state, when asking for an upcoming instrument well in advance) and negative (undesirable) (i.e.: lack of situation awareness when the anaesthesia monitor alarms are going off) behaviours in the categories of the aforementioned instruments were captured as they occurred, and also time stamped. For non-technical performances global scores were not given. The purpose for this approach was to determine the relationship between technical and non-technical performances in a temporal fashion.

The operating surgeon (staff or fellow) was assessed for technical and non-technical performance, while the rest of the participants (first assistant, nurses) were only assessed for non-technical performances.

On the ORBB synchronized timelines, “hot-zones” and “cold-zones” were created. A “hot-zone” was determined as 1-minute periods before and after an error, event and rectification. The “cold-zone” was determined as the routine flow of a procedure without any technical or non-technical
occurrences. The frequencies of the non-technical observations (NTO/minute) were calculated in both zones, and graphically illustrated. The “cold-zone” frequencies were considered baseline.

Non-technical behaviours of the surgical and nursing teams were evaluated after errors, events and prior to rectifications.

5.3.5 Statistical Analysis

Statistical analysis was performed using SPSS version 24.0 (IBM SPSS Statistics, IBM Corp., Armonk, NY, USA) and R Core Team 2016 (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). Normality of data distribution was assessed using the Shapiro-Wilk test. Descriptive statistics were performed of the patients’ characteristics. Technical and non-technical characteristics were compared in the Surgeon and Fellow groups. To compare the three fellows, one-way ANOVA was performed. For continuous variables, either Student’s T or Mann-Whitney-U test was used, based on the data distribution. For categorical variables, Chi-square ($\chi^2$) test of association was performed. When cell size was <5, Fisher’s exact test was used. Associations between technical and non-technical performances were calculated using Spearman’s rank-order correlation.

To evaluate the sequences of non-technical behaviours after an error, event and before a rectification, N-gram statistics was used. N-gram statistics is a model that assigns probabilities to sequences of occurrences. An N-gram is a sequence of N occurrences. In this study a 2-gram (bigram), or a two-occurrence sequence model was used, where one occurrence is the technical performance (error, event and rectification) and the other is the non-technical performance (NOTSS, SPLINTS).

5.4 Results

5.4.1 Patient characteristics

Between April 1st 2014 and Dec 31st 2015, 200 patients underwent laparoscopic Roux-En-Y Gastric Bypass (LRYGB) operation. A total of 63 (31.5%) operations were recorded and 56 (88.9%) of them fulfilled the inclusion criteria and were included in the study. The main reason
for the discrepancy between performed and recorded operations was incomplete data capture during the introduction of the technology.

The 56 operations contained 112 (56 JJ and 56 GJ) operative steps. Ninety steps (80.35%) were performed either by the surgeon (54 – 60%) or one of the fellows (36 – 40%).

There was no difference in performance between the three fellows (OSATS \( p=0.205 \); GERT \( p=1.000 \); Events \( p=0.298 \); Rectification of Events \( p= 0.457 \); Positive NOTSS \( p = 0.088 \); Negative NOTSS \( p= 0.271 \)) and therefore they were grouped.

### 5.4.2 Technical Performance

The mean duration was 40 minutes (SD: 11) for a GJ and 31 minutes (SD: 12) for the JJ. There were differences between surgeon and fellows in the mean operative time (GJ group: 36 vs. 48 mins, \( p<0.001 \); JJ group: 23.5 vs. 41.5 mins, \( p<0.001 \)), median OSATS (GJ group: 33 vs. 29.5, \( p<0.001 \); JJ group: 32 vs. 29, \( p<0.001 \)) and median GERT (GJ group: 8 vs 12 \( p=0.006 \), JJ group: 5 vs. 8, \( p=0.039 \)).

A total of 828 errors were made in the 90 steps, with a median of 8 errors/step [IQR: 5 - 11]. The most common errors were related to inadequate grasping force and depth perception (N= 416, 50.2% and N=205, 24.7% respectively), followed by wrong orientation (N=145, 17.5%) and inadequate visualization of the working instrument (N= 62, 7.5%).

A total of 222 events were recorded, with a median of 2 events/step [IQR: 1 - 4]. The most common event in both groups was bleeding (N=195, 87.8%), followed by tissue injury (N=25, 11.3%).

More than half of the events were rectified (N=149, 67.1%), with a median of 1 rectification/step [IQR: 0 - 3]. The most common rectification was hemostasis (N=136, 91.3%), followed by tissue repair (N=11, 7.4%).
5.4.3 Non-Technical Performance

A total of 1605 non-technical behaviour instances were recorded. Approximately half (815) were recorded for the surgical team (NOTSS), with a median of 7 observations/operative step [IQR: 4 – 12.3]. The most common behaviour exhibited was situational awareness (N= 436, 54.1%), followed by leadership (N=181, 22.2%), communication and teamwork (N=102; 12.5%) and finally decision making (N=91, 11.2%). Most of these instances were positive (N=786, 96.4%) and performed by the surgeon (N=635, 78%). Twenty-nine (3.6%) negative instances (e.g. inadequate communication, teamwork or situation awareness, etc.) were recorded and most of them (N=17, 58.6%) were exhibited by the surgeon.

For the nursing team (SPLINTS), a total of 790 non-technical observations were recorded with a median of 8 observations/operative step [IQR: 3 - 13]. The most common behaviour was task management (N=410, 51.9%), followed by situational awareness (N=240, 30.4%), and finally communication and teamwork (N=140, 17.7%). The majority of these instances were positive (N= 747, 94.6%) and were performed by the scrub nurse (N=675, 85.4%). Forty-three (5.4%) negative behaviours were exhibited most of them by the scrub nurse (N=34, 79%) as well.

5.4.4 Interaction Between Technical and Non-Technical Performances

Table 20 shows the Spearman correlation matrix for the surgeon and the fellow group. In both groups (S and F) there were moderate to strong correlations between steps of the procedure, duration, technical (Events, Rectification) and non-technical (NOTSS, SPLINTS) aspects of the operation.

Figures 2 and 3 show the frequency of non-technical observations/minute (NTO/min) at the baseline, around an error, event and rectification.

Bigrams (context bigrams starting with the term in question) show that the surgeon and the scrub nurse exhibit the most frequent non-technical behaviours after an error, an event and before a rectification (Figures 4 and 5).
### Table 20 - Spearman's Correlation Matrix for Surgeon and Fellow Groups

<table>
<thead>
<tr>
<th>Surgeon</th>
<th>Fellow</th>
<th>Step (JJ vs GJ)</th>
<th>Duration</th>
<th>OSATS</th>
<th>GERT</th>
<th>Events</th>
<th>Rectification</th>
<th>S NOTSS+ve</th>
<th>F NOTSS+ve</th>
<th>SN SPLINTS+ve</th>
<th>CN SPLINTS+ve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step (JJ vs GJ)</td>
<td>0.720**</td>
<td>0.168</td>
<td>0.316*</td>
<td>0.463**</td>
<td>0.553**</td>
<td>0.616**</td>
<td>-0.215</td>
<td>0.346*</td>
<td>-0.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration</td>
<td>0.292</td>
<td>0.130</td>
<td>0.395**</td>
<td>0.396**</td>
<td>0.454**</td>
<td>0.511**</td>
<td>-0.016</td>
<td>0.222</td>
<td>0.144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSATS</td>
<td>-0.008</td>
<td>-0.462**</td>
<td>-0.212</td>
<td>-0.067</td>
<td>-0.016</td>
<td>0.113</td>
<td>-0.23</td>
<td>-0.144</td>
<td>-0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GERT</td>
<td>0.246</td>
<td>0.643**</td>
<td>-0.495**</td>
<td>0.374**</td>
<td>0.271*</td>
<td>0.180</td>
<td>-0.082</td>
<td>0.187</td>
<td>0.137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events</td>
<td>0.276</td>
<td>0.179</td>
<td>-0.144</td>
<td>0.205</td>
<td>0.880**</td>
<td>0.679**</td>
<td>0.144</td>
<td>0.687**</td>
<td>-0.038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectification</td>
<td>0.222</td>
<td>0.137</td>
<td>-0.018</td>
<td>0.192</td>
<td>0.886**</td>
<td>0.672**</td>
<td>0.115</td>
<td>0.768**</td>
<td>-0.076</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S NOTSS+ve</td>
<td>-0.025</td>
<td>0.462**</td>
<td>-0.286</td>
<td>0.263</td>
<td>0.417*</td>
<td>0.380*</td>
<td>0.072</td>
<td>0.546**</td>
<td>0.193</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F NOTSS+ve</td>
<td>0.396*</td>
<td>0.476**</td>
<td>0.063</td>
<td>0.205</td>
<td>0.304</td>
<td>0.287</td>
<td>0.227</td>
<td>-0.007</td>
<td>0.084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SN SPLINTS+ve</td>
<td>-0.003</td>
<td>0.285</td>
<td>0.008</td>
<td>0.101</td>
<td>0.538**</td>
<td>0.625**</td>
<td>0.641**</td>
<td>0.371*</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN SPLINTS+ve</td>
<td>0.285</td>
<td>0.249</td>
<td>0.131</td>
<td>0.236</td>
<td>0.455**</td>
<td>0.466**</td>
<td>0.392*</td>
<td>0.147</td>
<td>0.396*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Left – bottom half of the Matrix illustrates the relationships in the Fellow group. Top – right half of the Matrix illustrates the relationships in the Surgeon group. The variables on the X and Y axis are similar for both groups. JJ = Jejuno-Jejunostomy; GI = Gastro- Jejunostomy; GERT = Generic Error Rating Tool; S = Surgeon; F = Fellow; NOTSS = Non-Technical Skills for Surgeon; +ve = positive behaviour; SN = Scrub Nurse; CN = Circulating Nurse; SPLINTS = Scrub Practitioners’ List of Intraoperative Non-Technical Skills; **Correlation is significant at the 0.01 level (2-tailed); *Correlation is significant at the 0.05 level (2-tailed)
Figure 2: S = Surgeon; F = Fellow; NOTSS = Non-Technical Skills for Surgeons; SN = Scrub Nurse; CN = Circulating Nurse; SPLINTS = Scrub Practitioners’ List of Intraoperative Non-Technical Skills, NTO/min = Non-Technical Observations / minute
Figure 3: S = Surgeon; F = Fellow; NOTSS = Non-Technical Skills for Surgeons; SN = Scrub Nurse; CN = Circulating Nurse; SPLINTS = Scrub Practitioners' List of Intraoperative Non-Technical Skills, NTO/min = Non-Technical Observations / minute
**Figure 4** - Bigrams for Surgeon Group

![Bar chart with categories: 1 phase post Error, 1 phase post Event, 1 phase prior Rectification, showing percentages for S NOTSS+ve, F NOTSS+ve, SN SPLINTS+ve, CN SPLINTS+ve](image)

Figure 4: S = Surgeon; F = Fellow; NOTSS = Non-Technical Skills for Surgeons; SN = Scrub Nurse; CN = Circulating Nurse; SPLINTS = Scrub Practitioners' List of Intraoperative Non-Technical Skills
Figure 5: S = Surgeon; F = Fellow; NOTSS = Non-Technical Skills for Surgeons; SN = Scrub Nurse; CN = Circulating Nurse; SPLINTS = Scrub Practitioners' List of Intraoperative Non-Technical Skills
5.5 Discussion

This study demonstrated the correlation between technical and non-technical performance on individual and team levels and presented a novel approach in studying the intraoperative environment. Similar findings have been previously published (Crossley et al., 2011), however on an individual level and with global scores. The current study focused on intraoperative observations as an intermediate safety outcome rather than global scores and found unique relationships between technical (events, rectifications) and non-technical performance (NOTSS, SPLINTS). This is the first paper analyzing the OR environment and these interactions using intermediate safety outcome parameters.

One of the unique aspects of the present study was showing the relationship between individual NOTSS and SPLINTS. There are multiple instruments in the literature designed to rate non-technical skills on a team level, such as OTAS (Observational Teamwork Assessment for Surgery) (Hull et al., 2011) or the Oxford NOTECHS II (Non-Technical Skills) (Robertson et al., 2014). Other instruments are designed to provide assessment on an individual level, such as NOTSS (Yule et al., 2006), SPLINTS (Mitchell et al., 2013), ANTS (Anaesthetists’ Non-Technical Skills) (Flin & Patey, 2011) or OSANTS (Dedy et al., 2015). However these instruments generate final global scores, which limits the opportunities to explore the relationships between sub-teams or individuals. The correlation seen in this study between the operating physicians (S and F), and nurses (SN and CN), addresses the limitations of global assessments, and further emphasizes the importance of teamwork in the operating room. The level of granularity of this approach provides opportunities for inter-disciplinary education of the OR team, using a deliberate practice model to address specific areas for performance improvement of the participating individuals.

As expected, more non-technical behaviours were exhibited around events and their rectification. This is intuitive as after an error causing an event, the surgeon needs to exhibit adequate non-technical performance (e.g. situational awareness, communication and teamwork) in order to accurately and promptly identify and rectify it, to avoid a potential postoperative adverse outcome. As demonstrated by Siu et al. (Siu, Maran, & Paterson-Brown, 2016) negative team behaviours likely precede technical occurrences, while more positive behaviours follow them.
The moderate to strong correlations between events, rectifications and positive non-technical performance (Table 20), as well as the increased rate of observed positive team interactions (Figures 2 and 3) is consistent with the above-mentioned previous report. Due to the low number of negative instances, no meaningful statistical analysis could be performed with them. Nevertheless, there was a decrease in the rate of positive non-technical observations around an error in both groups (Figures 2 and 3). There are a few possible explanations for this finding. Any particular non-technical occurrence might have been linked to more than one error happening consequently. However, instead of clustering these errors as one, they were accounted for individually and this might have overinflated the denominator, thus reducing the rate of positive non-technical observations around errors. This finding may also reflect the effect of non-technical performance on technical occurrences and highlight that there may be a true decrease in positive and increase in negative non-technical behaviours, potentially leading to an error. This hypothesis warrants further investigation.

Another interesting and unique aspect of this study was the finding that regardless who was the operating physician (Surgeon vs. Fellow) most of the non-technical behaviours were performed by the same two contributors to the team function - the Surgeon and the Scrub Nurse. There are two potential explanations for this finding. It may represent proactive behaviour by the surgeon, given his rank and experience in directing the scrub nurse. It may also represent a reactive (back up) behaviour to the fellow’s inexperience in non-technical performance. The stronger relationship between the Surgeon, the Scrub Nurse and the Circulating Nurse, while the fellow was the operating physician, might indicate a supportive “back up” behaviour and represent a resilience mechanism in “crisis”. This finding emphasizes the need for comprehensive educational interventions for trainees, the continuous assessment and improvement of both technical and non-technical skills as one, rather than separate entities (Abdelshehid et al., 2013; Phitayakorn, Minehart, Hemingway, Pian-Smith, & Petrusa, 2015), using coaching and deliberate practice principles.

The results of this study need to be interpreted with some limitations in mind. This study included only elective, standardized procedures, and involved a familiar operating team; hence the low number of negative non-technical behaviours. Future studies should focus on increasing and diversifying the number of participants and the operating settings to an acute, less familiar environment. Also, the anaesthesia team was not part of this initiative. The interactions between
surgeons and anaesthesiologists are indispensable for a successful patient outcome. Therefore the current assessment of the intraoperative environment is incomplete and future studies should be inclusive. Finally, the NOTSS and SPLINTS instruments were used to identify individual behaviours rather than global scores.

5.6 Conclusion

In conclusion, this present study demonstrated that technical and non-technical performances are related, both on individual and team level. Valuable data can be obtained around intraoperative errors, events and rectifications. This information could be used for future research endeavours in the area of surgical education and patient safety.
CHAPTER 6

The purpose of this chapter is to:

I. Provide summary of the main findings of the studies
II. Provide a summary for the thesis
6 General Discussion

6.1 Summary of Study Findings

Each study is briefly summarized below.

6.1.1 Study No. 1 (Chapter 1)

The systematic review in Chapter 1 (Study No.1) summarized the current evidence on how technical performance affected patient outcomes in surgery (Fecso, Szasz, et al., 2017). By doing so, it also underlined the deficiencies and weaknesses in the current evidence and highlighted opportunities for future work to strengthen this area of research.

This study addressed Aim #1 of the thesis.

This was the first review demonstrating, that regardless of the surgical sub/specialty, superior technical performance positively affected both intra- and postoperative patient outcomes.

Although this is exciting news, the lower quality and the deficiency in methodologies of included studies warranted further investigations. Some of the key points raised by this review that need to be addressed in future research are:

1. The direct assessment of technical performance;
2. Independent, trained observers;
3. Valid assessments using appropriate rating tools;
4. Detailed assessment of technical skills and errors to design educational interventions;
5. Appropriate patient outcome measures, from both, physician (M&M – morbidity and mortality) and patient aspects (PROM – patient reported outcome measures).
6.1.2 Study No. 2 (Chapter 3)

The retrospective multi-center, multi-surgeon cohort study in Chapter 3 (Study No. 2) investigated how intraoperative technical performance of the surgeon affected postoperative short-term patient outcomes in laparoscopic gastrectomy for cancer (Fecso, Bhatti, Stotland, Quereshy, & Grantcharov, 2017).

This study addressed Aim #2 of the thesis.

The methodology in this study was designed with a specific purpose in mind, and that is to address the main deficiencies highlighted in the systematic review. Independent, trained observer with the use of appropriate rating tools, performed a direct and detailed assessment of intraoperative technical performance, error and events and determined that lower technical performance was an independent predictor of higher postoperative complications. This study also demonstrated that surrogate variables for technical performance, such as surgeon volume or experience did not differentiate between the two groups, emphasizing the need for direct assessments of the variable of interest.

Besides technical performance, the patient comorbidities and the type of procedure were also found to be independent predictors of outcomes, which is in accordance with previous publications, however this study emphasizes the fact that the focus should be on variables that can be potentially enhanced and that is technical performance.

Another key point in this study was the detailed analysis of technical skills, with the use of error and event analysis and their negative correlation to the performance score.
6.1.3 Study No. 3 (Chapter 4)

This cross-sectional, census survey study in Chapter 4 (Study No. 3) investigated how surgeons generally involve trainees in a laparoscopic gastric cancer operation and how they distributed the main operative steps amongst them (Fecso et al., 2016).

This study addressed Aim #3 of the thesis.

Since the previous study demonstrated that technical performance affected outcomes, it was important to know how trainees are generally involved in the operating room, especially in a complex and uncommon procedure, like laparoscopic gastrectomy for cancer, for two main reasons: education and evaluation.

This study suggested that trainees do not routinely perform the main steps of these operations, and there were significant discrepancies in opinions between trainees and staff. Several other good discussion points were brought up that could provide the base for future research ideas. These are:

1. The lack of role clarity amongst trainees;
2. How to increase the level of active participation;
3. Alternate teaching opportunities, and finally;
4. The need for objective measurements, i.e. comprehensive recording of the entire operating room environment and video analysis are needed.
6.1.4 Study No. 4 (Chapter 5)

This prospective, observational, single-center, pilot study in Chapter 5 (Study No. 4) investigated the intraoperative environment in great detail, using a novel approach (Fecso, Kuzulugil, Babaoglu, Bener, & Grantcharov, 2017).

This study addressed Aim #4 of the thesis.

The study explored and revealed unique aspects of the interactions between technical and non-technical performances, on individual as well as team level. One aspect was the positive correlation between events, rectifications and non-technical behaviors. Although this is intuitive, it was interesting to find that regardless who was the operating physician (staff vs. fellow), the two main contributors to the non-technical performance were the staff and the scrub-nurse. Another important aspect was the stronger correlation between the surgeon, the scrub- and circulating nurses, while the fellow was operating. This could be exhibited back-up behavior and certainly warrant further investigations. Finally, in surgical education, perhaps technical and non-technical, or intrinsic performances should not be separated, but taught and assessed as one.
6.2 Thesis Summary

This thesis studied the intraoperative environment and some of the participants within. One of the participants of this environment is the surgeon whose performance (cognitive, technical and non-technical) are indispensable for a successful patient outcome.

The first two studies in this thesis support the importance of technical skills. To the authors best knowledge the systematic review was the first of its kind, in summarizing this evidence and providing guidance for future research (Fecso, Szasz, et al., 2017). The collective summary in the review, evaluating a heterogeneous group of surgical specialties, strengthened the validity and generalizability of the results. Although, the assessment instruments across different surgical sub/specialties might be rightfully different, the assessment framework suggested in the review might be applicable to all, where technical expertise is key to success. This framework consists of independent, trained observer(s), with content expertise to directly assess performance for the entire procedure, either in the operating room or based on video recordings, while using rating instruments with evidence to supports their use (Moorthy et al., 2003; Szasz et al., 2015). Due to the limited and low quality evidence more research was warranted, therefore the second study provided this evidence with specific methodological design in mind, to address most of the deficiencies reported in previous studies and by doing that, strengthened the body of knowledge in this specific area of research.

6.2.1 Strategies to enhance technical skills

Technical skills are important, since in many procedures the performance might contribute more to a successful outcome, than the perioperative care the patient receives (Birkmeyer et al., 2013). The mastery of these skills starts in the residency training programs, along with the mastery of other competencies such as patient care, medical knowledge, practice-based learning and improvement, interpersonal skills, professionalism and finally system-based practice. These are the competencies in general surgery ("ACGME Program Requirements for Graduate Medical Education in General Surgery," 2016; Frank JR et al., 2015). Therefore, for a surgeon to be competent needs to be proficient in all of the aforementioned categories (Satava et al., 2003).
In surgical education, technical skills can be acquired and performance can be assessed in a variety of settings: (1) operating room (Martin et al., 1997), (2) animal laboratory (Christopherson et al., 1986; Lossing et al., 1992; Saifi et al., 1990) or (3) bench models, simulators and virtual reality (Fried et al., 2004; Marescaux et al., 1998; Martin et al., 1997). However, in the era of work hour restrictions for trainees ("ACGME Highlights Its Standards on Resident Duty Hours ", 2001; "Council Directive 93/104/EC.," 1993; Leach, 2004), constant demand to increase efficiency and to decrease costs as well as medico-legal concerns (K. R. Reznick & MacRae, 2006), the operative exposure and hands-on training of residents got significantly reduced, therefore it is paramount that the remaining intraoperative learning opportunities are as efficient as possible, that all cases are included in the trainees portfolio and “outside of the OR” training possibilities exist.

In modern surgical education training programs, simulation-based curricula have been shown to deliver superior training results when compared to no intervention, and besides offering a safe environment for practice; it can also incorporate established instructional approaches to training, such as feedback and deliberate practice (Bonrath, 2015; Zendejas, Brydges, Wang, & Cook, 2013). Training programs and faculty to provide additional, “outside of the OR” learning opportunities and trainees to take “advantage” of them, especially in the cases of complex and uncommon procedures is paramount in order to ensure competence prior to start of independent clinical practice.

Such a complex and uncommon procedure is a laparoscopic gastrectomy for cancer and as the survey study (Study No.3) in this thesis revealed, keeping its limitations in mind, trainees do not routinely perform the major steps of these operations (Fecso et al., 2016). Since more than half of the trainees were exposed to this procedure at various stages throughout their training, it is imperative that even these complex operations to be included in the portfolio of teaching procedures. The survey also revealed, that the intraoperative learning opportunities are not utilized efficiently, especially on a junior trainee level (Fecso et al., 2016). Therefore, in order to improve the exposure and learning of trainees, a number of measures can be taken. First of all, improving the current use of existent intraoperative opportunities, by either using the graded responsibility approach, also known as the “Zwish model” (DaRosa et al., 2013) or the task deconstruction model (Marangoni et al., 2012). These training opportunities are described in more details in the discussion part of Chapter 4. To best utilize these opportunities, there is a
need for good communication between faculty and trainees, to clearly assign intraoperative roles and to clarify responsibilities in the briefing part of the pre-operative phase.

Combining the two models could maximize the benefits and could actively involve all members of the operating team from the most junior to the most senior. As the survey revealed, junior trainees have minimal active participation in these complex cases, however actively engaging them in the discussion around the case, the “show and tell” stage in the Zwish model, would benefit them the most, while the senior resident and the fellow could perform the operation using the task deconstruction model. The faculty potentially could supervise all the trainees by providing either smart-, dumb or no help at all. This approach is supported by Roberts et al. as well, suggesting that a structured approach to an operation, starting with preoperative task deconstruction with clear communication to clarify roles, objectives and expectations, followed by intraoperative teaching and postoperative feedback could potentially maximize education at all levels (Roberts et al., 2012).

Patient safety is above all, and based on the principle of “Primum non nocere”, surgeon educators have the responsibility to balance high quality and safe care with high quality education and occasionally this balance might be challenging. Alternatives to learning in the operating room are opportunities that include international collaborations, simulation training with virtual reality, animal or cadaveric surgical models (H. H. Kim & Ahn, 2011). A structured, foregut curriculum to include lectures, videos and hands-on models for trainees would be of great value, therefore future research should include the design, evaluation and implementation of such teaching interventions. These interventions could advance trainees in their learning curve, prepare them to be “pre-trained novices” as they enter the operating room (Gallagher et al., 2005). Recent systematic reviews provided evidence, that skills are transferable between simulation and the real operating room in a number of different laparoscopic procedures (e.g.: Nissen fundoplication, cholecystectomy, right hemicolecotomy, total extraperitoneal inguinal hernia repair, bilateral tubal ligation and salpingectomy) and endoscopic interventions (e.g.: endoscopic sinus surgery, fiberoptic laryngoscopy, esophagastroduodenoscopy, colonoscopy, transurethral resection of the prostate and flexible cystourethroscopy) (Dawe, Pena, et al., 2014; Dawe, Windsor, et al., 2014). Whether this holds true in the case of laparoscopic gastric cancer operations, remains to be determined.
While the above mentioned approach, with simulation and didactics can enhance the learning of surgical skills, the operating room remains a critical and unique environment, where trainees acquire not only their skills, but also their judgment and knowledge (Butvidas et al., 2011; Procter et al., 2010; K. R. Reznick & MacRae, 2006). As previously stated the efficient use of this environment has to be maximized and all operations could be potentially used for educational purposes, naturally keeping patient safety first and above all. Since uncommon procedures are uncommon, other similar operations could be used to acquire the necessary skills and subsequently transfer skills amongst similar procedures. The transfer of proficiency has been shown in some laparoscopic procedures in a simulated setting (laparoscopic appendectomy and laparoscopic salpingectomy) (Bjerrum et al., 2016). In the case of laparoscopic foregut surgery, one good example with a potential could be the transfer of skills between bariatric (e.g.: Roux-en-Y gastric bypass or sleeve gastrectomy) and oncology (gastric cancer) procedures. Whether this holds true, needs to be determined. However, this approach could be a feasible one if future research proves its value, since the number of laparoscopic bariatric procedures in the western world is far higher than the number of oncology procedures for gastric cancer, therefore trainees exposure and active involvement would be also much greater.

Interestingly though, skills might be transferable between procedures, however their evaluation might not be. A recent study done by Varban et al. (Varban et al., 2016) showed that in bariatric surgery, the same set of skills that predicted complications in laparoscopic Roux-en-Y gastric bypass failed to predict complications for laparoscopic sleeve gastrectomy, suggesting that the assessment of surgical skills has to be done independently for every procedure.

Although working hour restrictions have resulted in reduced operative exposure for trainees, there are a number of potential opportunities, discussed above, that needs further exploration and could potentially compensate for the “lost hours”.

Teaching and assessment of technical skills of surgical trainees is structured in training programs, and their goal is to assure competency and to create a fit independent practitioner at the end. However, does the learning and improvement of these skills end with the successful completion of training programs and passing board certification exams? Should technical skills and performance of practicing surgeons be assessed and enhanced on an ongoing basis? In the world of sports, even the elite athletes have ongoing coaching to provide a longitudinal, self-
directed improvement (Hu et al., 2012). In the world of surgery, although critical, no effective system exists for such coaching and most of the educational interventions for surgeons incorporate ineffective principles, e.g. teacher-driven rather than learner, didactic rather than interactive and amassed rather than distributed (Hu et al., 2012). Research has also suggested that the surgical learning curve even for more routine operations, such as a reduction mammoplasty, persists for several years after the completion of training (Carty, Chan, Huckman, Snow, & Orgill, 2009; Hu et al., 2012). It also has been advocated that surgeons should reflect on their own performance and analyze their errors to improve their skills and patient care (Rebasa et al., 2009). Furthermore, the objective assessment and subsequent feedback is important for all surgeons, regardless of their level of training or experience (Hu et al., 2012). Physician self-assessments are known to be unreliable (Davis et al., 2006), therefore there is a need to use an objective, independent observer to provided perspective, targeted feedback (Hu et al., 2012).

Thus far, only a few studies have assessed the technical skills of practicing surgeons and even fewer ones have linked skill levels to patient outcomes. One of the first studies in this area was published in 2013 and revealed that the technical skills of bariatric surgeons had wide variations and greater skills were associated with fewer postoperative complications, reoperations, readmissions and returns to the emergency departments (Birkmeyer et al., 2013). Although this was a landmark study, the measurement of surgical skills had some limitations, which were addressed by the second study in this thesis: (1) instead of self-selected operations, all available cases were included and (2) the rater was extensively trained and calibrated prior the review of (3) full length, unedited operations. Study No. 2, with its strengths and limitations, added to and enhanced the evidence that technical skills do contribute to postoperative complications, therefore their ongoing assessment and improvement could potentially improve outcomes as well. Although the true identity of the level of the operating physicians remained unknown, from the survey study (Study No. 3) it can be inferred that mainly practicing surgeons performed the critical steps of these operations.

Regardless of the level of operating surgeon (trainee vs. staff), the appropriate way of assessing intraoperative technical performance, as described in Study No. 1, as well as in other studies is the direct observation or review of the video recorded procedures by trained, independent observer(s), with content expertise along with the use of appropriate assessment instruments to
provide structured and individualized feedback, based on the principles of deliberate practice (Fecso, Szasz, et al., 2017; Hu et al., 2012; Moorthy et al., 2003; Szasz et al., 2015).

Video-based coaching is an educational modality to target intraoperative judgment, technique and teaching and in the study by Hu et al. (Hu et al., 2012) surgeons of all levels (chief resident, junior-, senior-, and very senior surgeon) found it highly instructive. This methodology may provide a practical, much needed approach for continuous professional development (Hu et al., 2012).

Of course, one of the main limitations of video-recordings or direct observations is the Hawthorne effect. The Hawthorne effect concerns research participation and describes the phenomenon of unintentional changes in behavior in response to the presence of the observer or a recording device (McCambridge, Witton, & Elbourne, 2014; Wickstrom & Bendix, 2000). It can affect any direct observational study and therefore cannot be completely eliminated (Kohli et al., 2009), however the Hawthorne effect typically fades with time, as the subjects are getting used to observation, especially if the presence of an observer is not directly visible. Therefore a long duration or continuous observation of intraoperative performances using small, unobtrusive video-audio recording devices can help to accelerate the habituation and to reduce the Hawthorne effect.

One such system recently developed and being studied is the Operating Room Black Box™ (ORBB) which is a multifaceted system that consists of a Surgical Data Encoder, Operating Room Control Console, Cameras, Microphones, and interfaces to the Laparoscopic Camera and Patient Monitor, developed by SST (Surgical Safety Technologies) Inc. The ORBB HD audio-video recording technology integrates the panoramic view of the operating room, the output from the laparoscopic view used by the surgical team as well as the patient’s physiological signs from the anesthesia machine. Audio information is captured using four strategically positioned microphones, designed to capture all relevant audio information. Video and audio feeds are synchronized (surgical timeline) for the rater to assess both technical and non-technical aspects of the operation, to assess both individuals and teams. Currently this system is available only for laparoscopic (endoscopic) procedures only, but technology is being developed to capture the technical aspects of open surgeries as well. The amount and granularity of intraoperative data that being captured now with this technology has never been captured before. This gives
opportunity to analyze, besides technical skills, human factors, non-technical skills and disruptions on a synchronized surgical timeline and to study the relationship of these variables to each other and in time. Figure 6 shows an example of a synchronized surgical timeline.
Figure 6 – Synchronized surgical timeline produced by Operating Room Black Box™

Fig 6: NOTSS – Non-Technical Skills for Surgeons; SPLINTS – Scrub Practitioners’ List of Intraoperative Non-Technical Skills; DISI – Disruption in Surgery Index; DO – Door opening

©SST Inc.

Reproduced with kind permission from Surgical Safety Technologies (SST) Inc. (Toronto, ON)
6.2.2 Technical and Non-Technical Skills

Human factors and non-technical or intrinsic performances, such as poor communication and teamwork, distractions, excessive workload and fatigue have been identified as potential contributors to surgical errors, complications and perioperative patient injury (A. A. Gawande et al., 2003; Greenberg et al., 2007; Lingard et al., 2004; Mazzocco et al., 2009; Sevdalis et al., 2009).

Although, both technical and non-technical skills are important contributors to a successful patient outcome, the interaction between the two remained under-investigated. The ORBB technology offered the opportunity to study these interactions in great details and Study No. 4 provided this initial evidence on how these two might be related.

Both technical and non-technical skills are essential part of the surgical competency as well ("ACGME Program Requirements for Graduate Medical Education in General Surgery," 2016; Frank JR et al., 2015). Competency in one area (e.g. good technical skills) doesn’t necessarily mean competency in other (e.g. good non-technical skills) (Szasz et al., 2016), therefore equal emphasis has to be placed on the continuous assessment and improvement of both skills. Guidelines for teaching non-technical skills for trainees have been introduced across North America, United Kingdom, and Australasia, however less so were the recommendations on how to deliver and incorporate this knowledge into the residency program (Dedy, Fecso, Szasz, Bonrath, & Grantcharov, 2016). More and more research and work is being done to incorporate non-technical skills into the curriculum for trainees (Dedy, Bonrath, et al., 2016; Dedy, Fecso, et al., 2016), however when it comes to practicing surgeons, coaching of both technical and non-technical performances is still in the initial phase of development and acceptance. Just as coaching and deliberate practice are mainstays in athletics and other high-performance industries, and are considered to be effective in improving performance at any skill level, they could be mainstays in surgery as well. In fact, the feasibility and benefits of one-on-one coaching amongst surgeons have been described (A. Gawande, 2011; Hu et al., 2012) and initiated already (Birkmeyer et al., 2013). However, the effectiveness of such strategies to enhance skills of practicing surgeons and to improve postoperative patient outcomes remains to be established (Birkmeyer et al., 2013).
6.2.3 Added Validity

As discussed in the introduction, assessment instruments are designed to measure a particular “construct” (e.g. performance, attitude, behavior or skill) and it is important to ensure that these measurements are valid (they measure what they were designed to measure) and reliable (the measurements are accurate) (Bonrath, 2015). In section 1.3 of the introduction the validity evidence was also presented, that supported the use of OSATS, GERT, NOTSS and SPLINTS (Table 5). The projects in this thesis not only used these instruments, but also, by doing that, added to their validity evidence.

Similar to Table 5, Table 21 shows the different sources of validity and how did the studies in this thesis contribute to each source.

Since the scope of this thesis was to assess the performance itself and not to determine pass or fail scores, or the impact of test scores on the persons assessed, or to develop and test the efficacy of educational interventions, the “Consequences” source of validity was not applicable. However, as discussed previously and will be discussed in the Future Directions chapter, studies should focus on developing and testing such interventions.
### Table 21 – Added Validity Evidence by this Thesis

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Sources of validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Content</td>
</tr>
<tr>
<td>OSATS</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Study No. 2 &amp; 4</td>
</tr>
<tr>
<td>GERT</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Study No. 2 &amp; 4</td>
</tr>
<tr>
<td>NOTSS</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Study No. 2 &amp; 4</td>
</tr>
<tr>
<td>SPLINTS</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Study No. 2 &amp; 4</td>
</tr>
</tbody>
</table>

OSATS = Objective Structured Assessment of Technical Skills; GERT = Generic Error Rating Tool; NOTSS = Non-Technical Skills for Surgeons; SPLINTS = Scrub Practitioners' List of Intraoperative Non-Technical Skills
Study No. 2 (Chapter 3) – Technical Performance as a Predictor of Clinical Outcomes in Laparoscopic Gastric Cancer Surgery
Study No. 4 (Chapter 5) – The Relationship Between Intraoperative Non-Technical Performances and Technical Events in Bariatric Surgery
6.2.4 Outcome measures

As touched upon, patient outcomes are important measures for both the patient and the healthcare system. The outcome metrics are one of the components that healthcare providers use to improve care as part of the Institute for Healthcare Improvement’s “Triple Aim” (IHI, 2008). Patients use these metrics in the process of their decision making and informed consent, to get a clear picture of the potential risks and benefits of any particular procedure (Birkmeyer et al., 2004).

Besides the hard endpoints of morbidity and mortality more emphasis is being placed on patient reported outcome measures (PROM) or patient centered outcomes (PCO) (e.g. functional status; health related quality of life, etc.) and the combination of these could be a more comprehensive measure of success of any particular medical or surgical treatment (Soreide & Soreide, 2013). However, as demonstrated in Study No.1 all of the included studies (up to mid 2014) collected and reported short – term morbidity and mortality, but only two of them (< 10 %) reported PROMs, therefore opportunities exist in this field as well and going forward, for comprehensive and accurate outcome measurements, both, traditional and patient reported outcomes, should be collected and analyzed together in a standardized and proper manner.

Several arguments support the need for PROM/PCO and the research on this topic (Gabriel & Normand, 2012). First of all, the aging of the population and the presence of the increasing number of comorbidities, along with the primary disease needs, require integrating knowledge about PCO in the discussion with patients. Second, in deciding on a particular treatment option, to accurately weigh the potential risks and benefits, the expected patient oriented outcomes need to be factored in as well. Lastly, in the era of individualized medicine, especially in the area of oncology, patient characteristics need to be accounted for on an individual rather than a group level. Patient reported outcomes should be included in decision-makings, however it must be based on valid, reliable and clinically useful measures (Lipscomb, Gotay, & Snyder, 2007).

Valid and reliable instruments to measure PROM/PCO are available (McDowell, 2006), however collecting such data efficiently and cost effectively is more difficult (Birkmeyer et al., 2004) and standards for translational processes from one language to another are somewhat still lacking (Soreide & Soreide, 2013).
It is important, that PROM/PCO comes directly from the patient, without the prior interpretation of it by anyone else (Fiscella et al., 2011). PROM can be investigated either as generic (i.e., overall health) or disease-specific (i.e., cancer specific) measures (Soreide & Soreide, 2013). Patient satisfaction, as part of PROM, is independent to morbidity, however emphasizes the fact that PROM may capture information not easily obtained through other metrics (Avery et al., 2006; Soreide & Soreide, 2013). Surgery in itself could potentially have immediate and long-lasting effects of patient related outcome-, and correspondingly health related quality of life measures (Soreide & Soreide, 2013). Surgeons’ proper knowledge and accurate application of PROM/PCO has great potential in contributing to a better understanding of the patients’ well being (Soreide & Soreide, 2013).

Finally, as indicated by Birkmeyer et al. (Birkmeyer et al., 2004), the quality in health care means “doing the right things right” and care must be taken to avoid missing the big picture, by focusing only on the latter component of “doing things right”. Incorporating patient preferences into clinical decision-makings will lead to focus on the initial component, and that is “doing the right thing” as well.
6.3 Thesis Conclusion

The goal of the present dissertation was to study the intraoperative environment and to highlight the importance of technical skills on patient outcomes, as well as to generate the momentum for future research and educational opportunities in the world of surgery. The studies contained within this thesis collectively described the existing-, and provided new evidence to support the importance of these skills as well as opened up exciting, new opportunities to study this unique environment, the interactions between technical and non-technical performances, between teams and individuals. The ultimate goal is to improve patient outcomes.

Patient outcomes stem from more than just a perfectly executed operation, and occasionally, despite of the most flawless technique, the patient will develop a complication during the recovery. We might not be able to improve all potential contributing factors to outcomes, e.g. the patient’s age, comorbidities, extent of the primary disease or the type of procedure needed, to name only a few, however technical skills of the surgeon is one of the contributing factors that has the greatest potential for optimization.

We as surgeons and physicians, have the obligation towards our patients to provide the best possible care for them and to continuously strive to improve our performance, outcomes and to decrease our footprint. Therefore, the continuing development and improvement of these skills, both technical and non-technical, remains important at all stages of one’s career, from the very first try on a dry lab training model all the way up to peer-to-peer coaching of senior surgeons.

The research area of skills and outcomes are still in its infancy, however studying this piece of the puzzle is exciting and promising with great potential for future opportunities in the field of surgical education and patient safety.
CHAPTER 7

The purpose of this chapter is to:

I. Review the general limitations
II. Suggest future opportunities to address the limitations
7 Limitations

7.1 Generalizability of the results

The studies reported in this thesis (Study No. 2 - 4) were all conducted within one training environment, therefore the generalizability and applicability of these results to other educational environments needs to be further determined.

Chapter 3 and 4 studied the technical aspects and level of the operating surgeon in the area of laparoscopic gastric cancer within a large university center with multiple participating surgeons and hospitals. The approach to this particular operation and the involvement of trainees and operative task assignments might vary between university centers. Aim #2 of this thesis was to directly assess technical performance itself and not the level of the operating surgeon and this approach might have addressed, at least partially, the generalizability limitation. Along the same line, the educational needs and possibilities of trainees, might vary from center to center, depending on their level of operative involvement, however what is unanimous across all centers is the work-hour restrictions, therefore the potential educational opportunities discussed in Chapters 3 and 6 might be applicable to all.

It seems intuitive that technical skills of the surgeon will effect the success of the operation, however this intuition has been proven only in a few settings, such as laparoscopic Roux-en-Y gastric bypass (Birkmeyer et al., 2013) and laparoscopic gastric cancer operations (Fecso, Bhatti, et al., 2017). Whether these results are generalizable to other operations requires further investigations, especially that Varban et al. (Varban et al., 2016) suggested that every procedure, might require its own independent assessment of technical proficiency.

Lastly, Chapter 5 (Study No. 4) revealed exciting pilot findings and relationships within an elective, highly standardized operation performed by a well functioning team. Whether these findings can be generalized and extrapolated to an emergency setting (e.g.: trauma laparotomy), in the middle of the night, working with a team not familiar to each other, remains to be determined.
7.2 Comprehensive assessment of the intraoperative environment

Each of the studies in Chapters 3, 4 and 5 assessed the intraoperative environment, however none of them could include all aspects (technical and non-technical) and participants (surgical-, nursing and anesthesia teams) of this environment.

To have a comprehensive intraoperative assessment, all team members should be included (surgical, nursing and anesthesia team) and besides technical aspects of the operating surgeon, non-technical (aka. Intrinsic) interactions on both individual and team level need to be assessed, human behaviors and technological aspects should be accounted for as well. Furthermore, objective recording and assessment of all aspects of the operating room environment, instead of relying on ones’ memory to recall information (survey study) would certainly improve the reliability of the findings.

Future studies should focus on designing methodologies to address these limitations.
CHAPTER 8

The purpose of this chapter is to:

I. Present possible future directions
8 Future directions

8.1 Comprehensive assessment of the intraoperative environment

The Operating Room Black Box™ platform (ORBB) provides “endless” opportunities to prospectively and comprehensively assess the intraoperative environment, to analyze technical and non-technical performances of each team member and to measure the environmental and technical factors and how they interact with each other. Furthermore, it allows to evaluate and study team members’ awareness of adverse events as well as the overall patient safety.

Figure 7 depicts the structure and potential opportunities using the ORBB.

The synchronized surgical timeline (Figure 6) produced by the ORBB allows for in detail analysis of the entire intraoperative environment at any given time and to determine the relationships and causalities between the variables of interest (i.e.: team members, technology, environmental factors, disruptions). This is a novel approach and gives details that were never been available before. These analyses will be used to identify error patterns leading to adverse outcomes, define area of improvement and set foundations for future educational interventions aimed to improve the performance of the surgical teams and ultimately to improve patient safety in the operating room.
Figure 7 – Opportunities with the ORBB™

©SST Inc.

Reproduced with kind permission from Surgical Safety Technologies (SST) Inc. (Toronto, ON)
8.2 Increase generalizability

One of the main limitations discussed in the previous chapter was the limited generalizability of the data. To address and improve these limitations, we propose that future studies should include diverse operating room settings, surgical specialties, teams and cultures.

The awareness of the ORBB platform is increasing and more and more hospitals on national and international level enroll in this novel research. This expansion will address the generalizability limitation and provide an enormous amount of exciting, never seen before intraoperative data to research.

8.3 Design of educational interventions

This thesis provided evidence, that technical performance does affect postoperative patient outcomes. The above-mentioned future goals will add further evidence on how non-technical performances, environmental factors, technology interact to each other and affect these outcomes. The main purpose is to ultimately improve patient safety and decrease adverse events and complications.

A great deal of educational interventions can be developed and evaluated both for individual as well as teams. These can include, didactic lectures with video presentations for visual representation, simulated operating rooms for both technical and non-technical teaching, animal and cadaveric models mainly for the technical skill improvements, as well as interventions to enhance the intraoperative learning at all training levels.

Improvement of performances as a result of an objective assessment, targeted feedback and individualized training interventions, using modern techniques of coaching and deliberate practice is of critical importance for both trainees, practicing surgeons and the entire operating room team. The superiority of these individualized, modern interventions, compared to conventional training has been demonstrated in multiple setting, using randomized control trials, including both technical and non-technical performances (Bonrath, Dedy, Gordon, & Grantcharov, 2015; Dedy, Bonrath, et al., 2016).
The intersection of surgical education, quality and safety and patient outcomes is a promising field of research with significant opportunities for state-of-the-art, cutting-edge research and knowledge translation.
References


Palter, V. N., & Grantcharov, T. P. (2012b). A prospective study demonstrating the reliability and validity of two procedure-specific evaluation tools to assess operative competence in


Copyright Acknowledgements

Permission to reproduce Table 1 in Chapter 1:

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
May 08, 2017

This Agreement between Andras B Fecso ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number: 4078310971108
License date: 
Licensed Content Publisher: John Wiley and Sons
Licensed Content Publication: British Journal of Surgery
Licensed Content Title: Objective structured assessment of technical skill (OSATS) for surgical residents
Licensed Content Author: J. A. Martin, G. Regehr, R. Reznick, H. Macrae, J. Murnaghan, C. Hutchinson, M. Brown
Licensed Content Date: Dec 8, 2005
Licensed Content Pages: 6
Type of use: Dissertation/Thesis
Requestor type: University/Academic
Format: Print and electronic
Portion: Figure/table
Number of figures/tables: 1
Original Wiley figure/table number(s): Figure 2
Will you be translating?: No
Title of your thesis / dissertation: Technical Performance and Patient Outcomes in Surgery: Another Piece of the Puzzle
Expected completion date: Jun 2017
Expected size (number of pages): 180
Permission to reproduce Table 2 in Chapter 1:

License Details

This Agreement between Andras B. Feczo ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

<table>
<thead>
<tr>
<th>Details</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>License Number</td>
<td>4070030230706</td>
</tr>
<tr>
<td>License Date</td>
<td>Mar 30, 2017</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>British Journal of Surgery</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Error rating tool to identify and analyse technical errors and events in laparoscopic surgery</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>E. M. Bonnath, B. Zevin, N. J. Dedy, T. P. Grantcharov</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Jun 10, 2013</td>
</tr>
<tr>
<td>Licensed Content Pages</td>
<td>9</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/Table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>1</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>1</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Technical Performance and Patient Outcomes in Surgery: Another Piece of the Puzzle</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jun 2017</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>108</td>
</tr>
</tbody>
</table>
Permission to reproduce Table 6 in Chapter 1:

<table>
<thead>
<tr>
<th>License Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Agreement between Andras B Fecso (“You”) and Wolters Kluwer Health, Inc. (&quot;Wolters Kluwer Health, Inc.&quot;) consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.</td>
</tr>
<tr>
<td>License Number</td>
</tr>
<tr>
<td>License date</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
</tr>
<tr>
<td>Licensed Content Title</td>
</tr>
<tr>
<td>Licensed Content Author</td>
</tr>
<tr>
<td>Licensed Content Date</td>
</tr>
<tr>
<td>Licensed Content Volume</td>
</tr>
<tr>
<td>Licensed Content Issue</td>
</tr>
<tr>
<td>Type of Use</td>
</tr>
<tr>
<td>Requestor type</td>
</tr>
<tr>
<td>Portion</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
</tr>
<tr>
<td>Figures/Tables/illustrations used</td>
</tr>
<tr>
<td>Author of this Wolters Kluwer article</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
</tr>
<tr>
<td>Expected completion date</td>
</tr>
<tr>
<td>Estimated size(pages)</td>
</tr>
</tbody>
</table>
Permission to reproduce Table 7 in Chapter 1:

<table>
<thead>
<tr>
<th>License Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Agreement between Andras B Fecso (&quot;You&quot;) and Wolters Kluwer Health, Inc. (&quot;Wolters Kluwer Health, Inc.&quot;) consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Field</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>License Number</td>
<td>408766000005206</td>
</tr>
<tr>
<td>License Date</td>
<td>Apr 14, 2017</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>Wolters Kluwer Health, Inc.</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Anesthesiology</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Adding Examples to the ASA-Physical Status Classification Improves Correct Assignment to Patients</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Hurwitz, Erin E.; Simon, Michelle</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Apr 1, 2017</td>
</tr>
<tr>
<td>Licensed Content Volume</td>
<td>126</td>
</tr>
<tr>
<td>Licensed Content Issue</td>
<td>4</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor Type</td>
<td>Individual Account</td>
</tr>
<tr>
<td>Portion</td>
<td>Figures/Table/Illustration</td>
</tr>
<tr>
<td>Number of figures/table/illustrations</td>
<td>Table 1</td>
</tr>
<tr>
<td>Figures/Table/Illustrations used</td>
<td>No</td>
</tr>
<tr>
<td>Author of this Wolters Kluwer article</td>
<td>Technical Performance and Patient Outcomes in Surgery: Another Piece of the Puzzle</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Jun 2017</td>
</tr>
<tr>
<td>Expected Completion Date</td>
<td>180</td>
</tr>
<tr>
<td>Estimated size (pages)</td>
<td></td>
</tr>
</tbody>
</table>
Permission to reproduce Table 8 in Chapter 1:
Permission to use our previously published manuscript (Section 1.3 Chapter 1):

Title: The Effect of Technical Performance on Patient Outcomes in Surgery: A Systematic Review
Author: Andras Fecso, Peter Szasz, Georgi Kerezov, et al
Publication: Annals of Surgery
Publisher: Wolters Kluwer Health, Inc.
Date: Aug 17, 0302
Copyright © 2017, Copyright (C) 2017 Wolters Kluwer Health, Inc. All rights reserved.

Request Denied
This request is granted gratis and no formal license is required from Wolters Kluwer. Please note that modifications are not permitted. Please use the following citation format: author(s), title of article, title of journal, volume number, issue number, inclusive pages and website URL to the journal page.
Permission to use our previously published manuscript (Chapter 4):

Accept your approved request

Dear Dr. Andras Fecso,

Elsevier has approved your recent request described below. Before you can use this content, you must accept the license fee and terms set by the publisher.

Use this link to accept (or decline) the publisher's fee and terms for this order.

Order Summary

Order Date:    Mar 10, 2017  
Order Number: 501245339  
Publication:  Journal of Surgical Education  
Title:  Training in Laparoscopic Gastric Cancer Surgery in the Western World: Current Educational Practices, Challenges, and Potential Opportunities at a Large University Centre  
Type of Use:  reuse in a thesis/dissertation

View or print complete details of your request.
Permission to reproduce Figure 6 in Chapter 6 and Figure 7 in Chapter 8:

Dear Dr. Fecso,

SST Inc. grants permission to reuse, for the purpose of your thesis/dissertation, the Synchronized Surgical Timeline sample and the Opportunities with the ORB5 figures (Figure 6 and 7 in the thesis).

Sincerely,

Surgical Safety Technologies Inc.