The impact of age and fatigue upon acquisition of surgical skills

Dr Ruth Elisa Blackham
MBBS FRACS FACLM MLegM DCH DRANZCOG VGDWH

School of Surgery
The University of Western Australia

This thesis is presented for the degree of Master of Surgery of
The University of Western Australia

The work presented in this thesis was performed in the
School of Surgery
[Surgical Education]

Year of submission 2015
Declaration

This is to certify that this thesis does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma from any university and that, to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Signed

Name: Ruth Elisa Blackham
Date: Thursday, 9 February 17
Abstract

The objective of this study is to quantify the effects of age and fatigue, both independently and as combined factors, upon the performance of newly-acquired surgical skills. The model required “novice” surgeons, in most cases medical students or recently qualified medical graduates, to perform three basic laparoscopic surgical skills at the start and end of the day with the imposition of regular working day fatigue. The assessors timed each session and compared the candidate’s morning and afternoon timings against self-rated fatigue scores and age.

There was no difference found within age and skill timings for the start or end of day tasks. Most of the fatigue scores did not correlate with timing of the day. Sleep did not seemingly affect times. Of note, there was a statistically significant decrease in the time taken to complete each laparoscopic skill at the end of the day compared with the start, regardless of age. There was no correlation of task timings between hours slept the night before or self-reported fatigue scores despite fatigue being higher during the evening session. Based upon this initial study there is no clear evidence to suggest a longer surgical training program would be required to compensate for older surgical trainees, nor does it suggest end-of-day fatigue has a negative impact upon time taken to perform a newly-acquired surgical skill.

Whilst further research is required to better elucidate the relationship of age upon acquisition of surgical skill, the imposition of fatigue upon task performance has previously been shown to increase error rate and impede accuracy. A larger cohort with greater fidelity simulation would be necessary to assess this in future studies.
Acknowledgements

I would like to express my sincere gratitude to Professor Jeffrey Hamdorf in providing supervision and guidance in preparation of this thesis and for sharing his valuable expertise in the area.

I am deeply grateful to my parents Bob and Sandra for their immeasurable encouragement and support in undertaking this project from the beginning. I am greatly indebted to my beloved husband Rob for his learned insights and kind patience through many long nights in bringing this project to completion. A special acknowledgment of my brother Bernard for providing practical advice in undertaking one of the most significant academic challenges I’ve faced thus far.

I dedicate this thesis to my Grandmother, Daw Sein Nyunt. Without her support and strength as inspiration this thesis would not have come to fruition.
Statement of candidate contribution

This thesis is my own composition, all sources have been acknowledged and my contribution is clearly identified in the thesis. No work in this thesis has been co-authored with others.

I have completed this thesis during the course of enrolment in this degree at UWA and it has not previously been accepted for a degree at this or another institution.
## Contents

Declaration................................................................................................................................................... i

Abstract....................................................................................................................................................... ii

Acknowledgements.................................................................................................................................... iii

Statement of candidate contribution ....................................................................................................... iv

List of figures............................................................................................................................................ vii

List of tables............................................................................................................................................. viii

Abbreviations ............................................................................................................................................ ix

1. Introduction ........................................................................................................................................ 1

1.1 Background......................................................................................................................................... 1

1.2 Age and learning ................................................................................................................................ 1

1.2.1 Context .................................................................................................................................. 1

1.2.2 Neural physiology – does age matter? ................................................................................... 3

1.2.3 Older students ........................................................................................................................ 4

1.2.4 Pilots, drivers and age............................................................................................................ 5

1.2.5 The relevance of surgeon age ............................................................................................. 6

1.2.6 Summary ................................................................................................................................ 7

1.3 Human factors ..................................................................................................................................... 7

1.3.1 Definitions ............................................................................................................................. 7

1.3.2 Consequences of fatigue ........................................................................................................ 8

1.3.3 Fatigue and the medical profession ...................................................................................... 10

1.3.4 Fatigue and the surgeon ....................................................................................................... 12

1.3.5 Fatigue and surgical outcomes ............................................................................................. 14

1.3.6 International change ........................................................................................................... 15

1.3.7 The Australian experience ................................................................................................... 19

1.4 Acquisition of surgical skills ............................................................................................................. 21

1.4.1 Acquisition of skills ............................................................................................................. 21

1.4.2 Assessment of surgical competency – skills and judgment ................................................. 21

1.4.3 Role of simulation .............................................................................................................. 22

1.4.4 Comparing simulated environments .................................................................................... 22

1.4.5 Method of teaching ............................................................................................................. 23

1.4.6 Transferability of skills ........................................................................................................ 24

1.4.7 Types of simulators .......................................................................................................... 25

1.4.8 Formalised education programs ........................................................................................... 26

1.5 Training context – becoming a surgeon ............................................................................................ 27

1.5.1 Surgical training .................................................................................................................. 27
## List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>The ACGME common standards for resident duty hours</td>
<td>16</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Basic Laparoscopic Skills Set-up</td>
<td>38</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Task 1 – Cube stack</td>
<td>38</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Task 2 – Peg transfer</td>
<td>39</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Task 3 – String pull-through</td>
<td>40</td>
</tr>
</tbody>
</table>
# List of tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 4.1</td>
<td>Demographic experience</td>
<td>43</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Skill timings (AM vs PM)</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Age vs skill timings</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Age vs fatigue self-rated scores</td>
<td>45</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Fatigue and skill timing</td>
<td>45</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Sleep and skill timing</td>
<td>46</td>
</tr>
</tbody>
</table>
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGME</td>
<td>Accreditation Council for Graduate Medical Education</td>
</tr>
<tr>
<td>AMA</td>
<td>Australian Medical Association</td>
</tr>
<tr>
<td>ASiT</td>
<td>Association of Surgeons in Training</td>
</tr>
<tr>
<td>BST</td>
<td>Basic surgical training</td>
</tr>
<tr>
<td>EWTD</td>
<td>European working time directive</td>
</tr>
<tr>
<td>FLS</td>
<td>Fundamentals of laparoscopic surgery program</td>
</tr>
<tr>
<td>MISTELS</td>
<td>McGill inanimate system for training and evaluation of laparoscopic surgery</td>
</tr>
<tr>
<td>MIST-VR</td>
<td>Minimally invasive surgical trainer – virtual reality</td>
</tr>
<tr>
<td>MMC</td>
<td>Modernising medical careers</td>
</tr>
<tr>
<td>ORIFs</td>
<td>Open reduction internal fixation</td>
</tr>
<tr>
<td>RACS</td>
<td>Royal Australasian College of Surgeons</td>
</tr>
<tr>
<td>SAGES</td>
<td>Society of American Gastrointestinal and Endoscopic Surgeons</td>
</tr>
<tr>
<td>SET</td>
<td>Surgical education training</td>
</tr>
<tr>
<td>SST</td>
<td>Specialist surgical training</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Background

There is increasing worldwide recognition of the competing forces of restriction of trainee working hours against the rising tide of technological advances in surgery. Whilst this has been recognised, the imposition of age due to the trend towards postgraduate medical education in Australia has not been quantified in current research. Simulation will provide an ever-increasing complement to the repertoire of surgical training tools; however, its role has yet to be properly elucidated. This study begins to clarify some of the questions raised within the conversation of the future of surgical training and in particular whether the effect of age or fatigue plays an important role upon the performance of newly-acquired surgical skills such as those performed by a junior trainee surgeon.

1.2 Age and learning

1.2.1 Context

The age of medical graduates in Australia has increased appreciably over the past decade. This began with the establishment of 10 new medical schools in 2000 allowing 19 universities accredited to produce medical graduates (Scott, 2012). The type of medical degree has also changed over this period with a move towards shorter courses (3 or 4 years instead of 6) with postgraduate entry. In 2010, 52.1% of the medical students commencing began their studies after finishing another degree. The majority of these (93%) entered a graduate program (MTRP, 2012). The average age of the medical student
cohort is also increasing. Within the same report, 17.5% of students commencing medical studies were over the age of 25 years.

The older entry cohort has coincided with a decrease in overall clinical contact time in the hospital system of postgraduates. The average weekly hours worked by employed medical practitioners only decreased from 43.6 hours in 2006 to 43.3 hours in 2010, despite an increase in the overall supply of doctors by 13.3% (AIHW, 2012). Feminisation of the workforce has continued, constituting 37% of employed practitioners compared with 33.7% 5 years earlier. The largest distinction lies within the trainee group as 49% of hospital non-specialists are female compared with 25% of specialists. Females currently represent 53% of all medical graduates (MDANZ, 2012).

The principal trainer of surgeons in Australia is the Royal Australasian College of Surgeons (RACS). The RACS training body is responsible for the education of clinical skills and surgical competency in nine sub-specialties. The College underwent a major reform of its training arrangements in 2007, replacing its Basic Surgical Training (initial 2-4 years) and Specialist Surgical Training (subsequent 4 years) with a condensed Surgical Education Training program (SET) over a 5-6 year period, combining the previous BST/SST program. This change ensured a shift from a prolonged, generic model of training followed by specialisation to a program where the undifferentiated pre-vocational doctor is streamed early into a subspecialty, putatively reducing the risk of trainees not proceeding beyond basic surgical training or spending years awaiting a place in specialty training after these basic surgery years as was commonplace prior. However, the College has warned against the assumption of the new competency-based training
program shortening the duration of training; instead confirming trainees may take longer, rather than less time to meet the requirements of the new SET program (RACS, 2011). Whilst acknowledging the role the introduction of the SET program has played in streamlining the training process, RACS has also asserted that any further reduction in the number of training years will produce under-qualified surgeons. Thus it would seem despite the increasing age of medical graduates entering surgical subspecialty training, there is little prospect of the duration of training decreasing from the current 5-6 years.

1.2.2 Neural physiology – does age matter?

Age-related cognitive decline was historically thought to be as a result of quantitative loss of neurons. More recent data indicates age-associated behavioural impairments result from changes to dendritic morphology, cell connectivity, calcium dysregulation and gene expression affective cognitive plasticity (Burke, 2006). The notion that the systemic milieu of the body inhibits or promotes adult neurogenesis in an age-dependent manner is supported by studies in mice, where exposing a young mouse to an old systemic environment decreases synaptic plasticity and impaired spatial learning and memory (Villeda, 2011). The area of the brain most affected by age-related decline is the hippocampus; however, this has not been directly linked to functional deficits associated with normal ageing. There is evidence to suggest age-related changes in synaptic plasticity within the hippocampus and neocortex have similar effects upon synaptic transmission to diabetes and stress (Artola, 2008). This infers the changes produced by diabetes and stress, of persistent inhibition of long-term potentiation and facilitation of long-term depression, might similarly lead to synapse weakening and subsequent cognitive impairment in ageing. Of greater certainty is the relative vulnerability of the
medial temporal lobe and prefrontal cortex to age-related decline, areas that are relevant
to learning, memory and higher executive function. Decreased functional plasticity in
these areas can contribute to behavioural compromise in the absence of significant
pathology (Burke, 2006).

In addition to the anatomical changes, there are age-related adaptations of the
neurotransmitters in the prefrontal cortex. Dopamine concentration, transporter
availability and dopamine D2 receptor density all decline with age (Hedden, 2004).
Serotonin receptor availability in the frontal cortex also declines with age and correlates
with striatal declines of dopamine receptors with largest declines seen in mid-life. Both
volumetric and neurotransmitter changes within the prefrontal cortex in addition to the
striatum has been associated with cognitive performance impairment associated with age.

1.2.3 Older students

There are increasing numbers of older students commencing postgraduate studies. As a
result, the learning style preferences of this group has become of greater interest to
educators. The traditional thinking suggests older learners are active, hands-on learners.
Older adults prefer structure, learning in an auditory manner and learn independently
compared with younger learners who prefer kinaesthetic modes of learning with authority
figures present (Price, 1977). This notion has since been disproven; instead there is a
tendency to become more reflective and observational in the learning environment. A
study of the preferred learning styles of 172 older adults showed the older age group (75+)
preferred to learn by the “assimilator” style of thinking and watching, rather than feeling
or doing (Truluck, 1999).
1.2.4 Pilots, drivers and age

A significant degree of research has been undertaken in the aviation industry comparing the outcomes of older pilots involved in air transport related accidents. A retrospective analysis of 10 years of US National Transport Safety Bureau data indicated pilots older than 60 years of age are more likely to be involved in an accident caused by pilot error (Barazrnga, 2010). In addition, pilots over the age of 60 years are more likely to be involved in a fatal accident. Experience has been independently analysed; studies associating pilot experience and pilot error found the more experienced pilots were less likely to be involved in an accident caused by pilot error, but these pilots are more likely to be involved in fatal accidents (Barazrnga, 2010; Li, 2003). It is postulated this association is due to the higher risk environment of the flights they conduct, in more challenging conditions.

A negative correlation between age and driving performance is often presumed but often is without significant evidence on the basis of seniority rather than merit. Several changes may contribute to a physiological decline, including increased reaction time, vision changes, cognitive changes and decreased neck-turning ability (Wood, 2002). In particular, visual changes such as decreased visual acuity, reduced light sensitivity and increased glare sensitivity may make the ability for visual compensation more difficult for older drivers (Arbesman, 2008). However, current evidence suggests increased age does not cause higher crash rates per exposure, which goes against the traditional notion that older drivers suffer age-related deterioration of driving skills (Hakamie-Blomqvist, 2002; Langford, 2006). Older drivers are indeed over-represented in terms of risk of fatal injury (greater than 70-year-olds) than for younger age groups; this risk is also high for
older adult passengers and pedestrians who represented the majority of older adult fatalities (Rolison, 2012).

In Australia restrictions upon older drivers is state-dependent. For example in Victoria there are no compulsory restrictions for elderly drivers, but drivers over 75 need to renew their licence every 3 years. In NSW this is every 2 years, and Queensland they must carry a medical certificate, which can last for up to 5 years.

1.2.5 The relevance of surgeon age

Whether the age of a surgeon is relevant is a subject that has been considered taboo. Whilst the older surgeon is objectively seen as having greater experience and thus able to make better-informed decisions, little research has looked at whether the decision to operate and the type of operation is influenced by age as an isolated factor. A notable survey by the American Society of Colorectal Surgeons reviewed the surveillance strategies of rectal cancer patients, finding no difference of follow-up testing trained at various time and the motivations for this being nearly uniform among age strata (Patel (2008). However, an orthopaedic study reviewed the extent to which surgeon age influenced treatment patterns for distal radius fractures, hypothesising younger surgeons undertook Open Reduction Internal Fixation (ORIFs) among elderly individuals more frequently than older surgeons. They found a significant difference between surgeons aged 40 and younger being more likely to perform ORIFs. There was no difference in overall outcome (Waljee (2014).
1.2.6 Summary

The increasing age of postgraduate students undertaking medicine and subsequently surgery makes age a relevant factor in surgical training. There has been relatively little research into the impact this has had from an educational perspective, instead the focus has traditionally been upon the effect of the professional doctor, surgeon or pilot with advancing years towards retirement. The question of whether older age affects learning ability and subsequent effect upon length of training has yet to be considered.

1.3 Human factors

1.3.1 Definitions

Fatigue is known to contribute to surgical error; however, its impact is not routinely quantified. The study of human factors seeks to identify the cause of surgical errors within a vulnerable system with the intent of optimising performance (Reason, 2001). It can be used to identify the systems factors affecting surgical performance, patient outcomes and maintenance of surgical skill.

The literature surrounding human factors is relatively recent. Historically, surgical outcomes had been primarily assigned to the technical competency of the surgeon in conjunction with the comorbidities of the patient (Elbardissi, 2012). The unspoken basis of surgeon rankings assumes outcomes to be directly attributable to surgical skill after being adjusted for patient risk factors. The more modern approach considers system factors in its modelling in conjunction with technical considerations and the patient’s premorbid state.
Fatigue is defined as a temporary loss of energy and strength resulting from physical or mental work, usually negatively associated with performance (Nbaa, 2010a). Three types of fatigue can be described (Nbaa, 2010b);

- transient (brought on by extreme sleep restriction or extended waking hours of 1-2 days)
- cumulative (repeated mild sleep restriction) or
- circadian (reduced night-time performance, typically 2:00am – 6:00am).

The individual effect of fatigue is dependent upon a number of factors, including personal sleep requirements, physical condition, diet, age, alcohol and smoking, outside stressors and comorbidities such as pulmonary and cardiovascular pathology (Sinha, 2013).

A systematic review of articles pertaining to human factors within the context of surgery by Shouhed & Catchpole found the main themes relate to development and application of human factor theories, system flow concepts and safety issues (Shouhed, 2012). The authors note there is a dearth of research into adverse events being cause by disruption to the flow of an operation such as teamwork and communication failures. Fatigue is an element of workload factors that can reduce cognitive function contributing to an overall poor outcome. Recognition of this has led to the creation of checklists and specifically timed rest breaks during surgical procedures as noted by El Bardissi (2012).

1.3.2 Consequences of fatigue

The physiological effect of sleeplessness is attributed to hypometabolism of cerebral tissue, in particular the thalamic, pre-frontal, frontal and occipital cortices (Drummond, 2001). A fine tremor becomes noticeable, speech is slower and monotonous with
increased pain sensitivity. It has been noted that the risk of metabolic syndrome with central obesity, diabetes mellitus and cardiovascular disease also increases. The effect of fatigue is frequently compared to blood alcohol levels. The effect of a high blood alcohol level of 0.10% upon attention span, reflexes, task accuracy and perceptiveness is comparable to the effects of a continuous sleep-deprived night or an accumulated sleep deficit of 4-5 hours per night over 7 days (Williamson, 2000).

Much of the research to date into the consequences of sleep deprivation in occupational medicine pertains to machine operators, vehicle drivers and commercial airline pilots. Sleeplessness accounts for impaired perception, difficulty concentrating, slower reaction times and micro-episodes of sleep during waking hours (Orzel-Gryglewska, 2010). These contribute toward lower efficiency of task completion and increased error rate. The effect of sleep deprivation depends upon the task being measured. Mood is more affected than cognitive or motor performance, and partial sleep deprivation has a more profound effect on functioning than either acute or chronic sleep deprivation (Pilcher, 1996).

A significant effect of fatigue upon vehicle drivers has been noted, in particular the duration of sleep and the number of road accidents. After lack of sleep a subject taking a driving simulator test created errors such as driving over the median strip or too close to the kerb in addition to increasing their average driving speed (Peters Rd, 1999). Notably the driver’s age is a statistically significant factor upon the fatigue-related accidents and is a causal determinant of the time of day when the accident occurs; of note, drivers below the age of 25 cause more than twice as many road accidents as other age groups despite making up a minority of drivers (Pack Ai, 1995). Younger drivers are also more likely to
cause accidents at night. It should be noted that these studies pertained to one or more nights of sleeplessness.

As a consequence of the hazards of long shifts there are multiple regulations in place which limit occupational work hours. In the United States, truck drivers are prescribed a maximum of 11 hour shifts. Similarly airline pilots are allowed to fly on domestic routes for a maximum of 8 hours in any given 24-hour period. Train engineers can work for no more than 12 hours at a stretch. Each of these limits have independently been formulated by their regulatory organisations in recognition of the effects of fatigue and enforced by employers, unions and legislature.

1.3.3 Fatigue and the medical profession

A consistent definition of fatigue in the scientific literature is elusive. Chronic partial sleep deprivation can be defined as sleep duration of less than 5-6 hours for consecutive nights. Such chronic sleeplessness is commonplace in medicine with up to 20% of trainees reporting less than 5 hours of sleep each night and 66% sleeping 6 hours or fewer (Baldwin, 2004). Most importantly, the trainees who slept less than 5 hours were more likely to make a medical error and to report working in an “impaired state”.

One of the first sentinel papers describing the effect of sleep loss upon physicians by Friedman defined the negative effect of sleep deprivation upon cognitive function, working memory, fine motor skills, vigilance and mood (Friedman, 1971). The most significant deviations occurred with acute (single night) sleep deprivation in excess of 46 hours dependent on the task being performed (Koslowsky, 1992).
Several disciplines have independently correlated sleep deprivation and specialty tasks with varying results. Sleep-deprived paediatric residents awake for 24 continuous hours required more time to perform umbilical vein catheterisation, without significant effect upon cognitive skills. (Storer, 1989). In epidural anaesthesia the relative risk of unintentional dural puncture increased for night-time than day-time cases (Aya, 1999). The senior anaesthetists were less likely to cause a dural puncture; however, the study was insufficiently powered to elucidate this. In a simulated anaesthetic monitoring situation after 24 hours of in-house call the residents scored significantly worse on the vigilance test than rested (Denisco, 1987). Similarly, acutely sleep-deprived anaesthetic residents in a similar 2003 study demonstrated progressive psychomotor deficits in memory and vigilance and one-third of residents fell asleep during cases. Of note there were no differences in clinical skills or error rates. Finally, interns in the intensive care unit on two different rosters, one on the usual roster of 30 consecutive hours compared to a new roster restricting successive hours to 17 observed the shortened shifts manifested in 36% fewer medical errors and inattention failures (Landrigan, 2004).

Measuring the impact of shift work is a relatively modern concept. Previous research provided insight purely into the effects of acute or chronic sleep deprivation. In terms of patient outcomes, night-time admission to an intensive care unit was not associated with increased mortality or length of stay (Morales, 2003). In terms of healthcare provider outcomes, interns surveyed in the United States report increased needlestick injuries between 11:30pm to 7:30am, with percutaneous injuries more frequent during extended shift hours (“extended work” on average being 29.1 consecutive hours) (Ayas, 2006).
1.3.4 **Fatigue and the surgeon**

The requirement of a doctor to work long hours with overnight call duty and long or rotating shifts have long been considered a standard part of medical practice. Being able to overcome the biological imperative of sleep has historically been seen as a trait of a good surgical trainee. There are myriad factors contributing to this including lack of doctors to adequately staff a roster, plus competing requirements such as study and family responsibilities among others. However, the neuro-behavioural impairments upon cognitive and motor skills that can affect the performance of a surgeon are becoming less excusable in a modern environment with patient safety considered paramount, greater awareness by health practitioners of the effects of fatigue, workplace regulations and an increasingly medico-legal hospital system.

The amount of sleep deprivation required is unclear; a meta-analysis of the relationship between total sleep deprivation and performance found correlations to be highest for duration of 45 hours or greater (Koslowsky, 1992). These affected speed rather than accuracy measures of performance. Other studies specifically measuring laparoscopic performance after 17 hours on-call with disturbed sleep showed a significant slowing of task time with increased number of unnecessary movements across all tasks completed (Grantcharov, 2001).

Several reviews have surveyed the effects of fatigue upon trainees. A physician review indicated 41% of residents cite fatigue as a relevant cause of their most significant medical mistake, with a third resulting in patient fatality (Wu, 1991). A laparoscopic trainer, Minimally Invasive Surgical Trainer – Virtual Reality (MIST-VR)) was used to assess skill after acute sleep deprivation with residents serving as their own control.
comparing pre-on-call skill with post-24 hours of in-house on-call. The authors compared technical skill by way of speed, errors and economy of movement. The number of errors and time to task completion increased at the fatigued assessment (Eastridge, 2003).

The effect of “impaired” sleep is more difficult to quantify in definition but is relevant to trainees on an on-call roster. There is little differentiation in the literature between sleep deprivation, both acute and chronic, versus disturbances in circadian rhythm relevant to more recent trainees on a shift roster. An increase in complications among post-night-time surgical procedures with sleep opportunity of 6 hours or less has been demonstrated in a matched retrospective cohort study by surgeons and gynaecologists; however, the confounding factor of the “sicker” patient whose operation could not be deferred to the next morning is noted (Rothschild, 2009).

Despite the above evidence, surgeons have traditionally been reluctant to discuss their sleep-deprived state with patients in terms of informed consent. The President of the American College of Surgeons Dr Carlos Pellegrini (2013-14) has appealed to the professionalism of surgeons and armed them with information about the effects of sleep deprivation, when questioned whether surgeons should be prohibited from operating without informed consent in the fatigued setting (Czeisler, 2013). The contributing factors of overall fitness, hours worked and unpredictable nature of surgery are not readily quantifiable; in addition the argument that a night off call does not equate to a full night’s rest holds true. A web-based American survey of trainees showed an increase in mean sleep duration of 22 minutes despite falls in overall work hours and shift duration (from 5.91 hours to 6.27) (Landrigan, 2006). Further research into the recovery time for
laparoscopic skills after extended periods on duty should be conducted, in addition to evaluating means of increasing alertness and fatigue reduction (Grantcharov, 2001).

There is a dearth of Australasian data on the effects of fatigue (Watters, 2014). A historical-based cohort compared with a 10-hour fatigued group showed significant deterioration in laparoscopic skills; surgical experience did not influence the primary outcome (Daruwalla, 2014). The lack of such an effect is surprising given the often presumed resilience of the experienced surgeon to fatigue as they would rely upon frequently used skillsets with less cognitive imposition. This has been proven in the significant difference between a novice and an expert calling upon subconscious pattern recognition or “recognition primed” approaches for straightforward clinical decisions (Crebbin, 2013). Thus they avoid overloading conscious working memory.

1.3.5 Fatigue and surgical outcomes

The presumed effects of fatigue upon surgical outcomes are an area of significant interest from both academic and medico–legal perspectives. Whilst electromyography measurements taken through a full day of operating and fatigue indices compared to desk work controls showed fatigue in all subjects operating, with no change compared with those performing desk work (Slack, 2008). No difference was found between consultants and trainee registrars. The most obvious sequela of this would be upon intraoperative complications; one small study examining start-time and duration of orthopaedic surgery with incidence of intraoperative femoral fracture in hip arthroplasty did not show a significant difference (Peskun, 2012).
A systematic review of the effect upon performance and patient safety relating to sleep-deprived surgeons found two of five studies reported increased complications or errors (Sturm, 2011). A further 11 studies assessing psychomotor skills using simulation were evaluated with two reporting no significant differences. As noted by the authors, the heterogeneity of the results did not allow for statistical pooling. In addition, the lack of randomised controlled trials in this area is reasonable given the unethical proposition of deliberate sleep deprivation prior to a surgeon operating on a patient. However, the lack of content validity in simulated environments must be taken into account.

There may be a larger effect of sleep deprivation in non-physicians than physicians, particularly in the areas of cognitive function, memory and vigilance (Philibert, 2005); however, this difference may be due to the chronic reduction of sleep in the physician cohort. The difference could be attributable to “training” in sleep deprivation and acclimatisation to working in a sleep deprived state; or selection bias on the part of the physician choosing to continue to work in their specialty field because of their resilience to the aforementioned effects.

### 1.3.6 International change

In the USA, the ACGME (Accreditation Council for Graduate Medical Education) publicised the Common Program Requirements which restricted working hours for first year residents to a maximum continuous duty period of 16 hours, with postgraduate year two and above able to work 24 hour continuous shifts with 4 hours permitted for handover; however, “strategic napping” is recommended (Philibert, 2002). Duty hours
were limited to a maximum of 80 hours per week in 2003 (including all clinical and administrative duties and conferences) (Figure 1.1).

![The ACGME Common Standards for Resident Duty Hours](image)

**The ACGME Common Standards for Resident Duty Hours**

- An 80-hour limit per week, averaged over 4 weeks
- 1 day in 7 free from all program responsibilities
- Adequate rest must be provided between duty periods (this should generally be 10 hours)
- In-house call no more often than every third night
- A limit on continuous duty of 24 hours up to an added 6 hours for transfer of care and didactics; no new patients are allowed after 24 hours
- In-hospital hours during call from home is counted
- Moonlighting must be approved by program director
- In-house moonlighting counts towards the weekly 80-hour limit

*Accreditation Council on Graduate Medical Education*

**Figure 1.1** The ACGME common standards for resident duty hours

In 2004, Great Britain with the European Union began to adhere to the European Working Time Directive (EWTD) prescribing limits of 58 hours per week for doctors in training. In August 2009 the EWTD came into full effect, further restricting hours (Pickersgill, 2001). This safety directive limited doctors to a maximum of 48-56 hours per week and only 13 consecutive working hours per shift. For doctors in training this meant a maximum 48-hour week averaged over a consecutive 6-month period. The EWTD also limited rest periods and annual leave. This move has proven unpopular with surgeons and trainees due to the perceived effect of limiting theatre exposure for junior doctors during critical training years and decreased case numbers. This effectively increases the requirement of a surgical consultant to be in theatre, at a time of decreased supervisor capacity. As such the United Kingdom’s Association of Surgeons in Training (ASiT) has
petitioned for a change in regulation towards 65 hours per week similar to Australian trainees (Cresswell B, 2009). The New Zealand Employer – Resident Contract similarly restricts working hours to a maximum of 72 hours per week and 16 consecutive hours in a shift.

Whether such measures are effective in guarding against the negative effects of sleep loss upon clinical duties is controversial. A finding that sleep loss of 24-30 hours produces a significant reduction in physicians’ aggregate clinical performance indicates the deficiency in performance may be experienced despite current USA minimum work guidelines (Philibert, 2005). Whether there is a similar effect of sleep deprivation in physicians compared with non-physicians is difficult; in many studies of doctors in training the “rested” controls actually meet the clinical definition of chronic sleep deprivation, below 5-6 hours for several consecutive nights (Carskadon, 1981).

Due to concerns over continuity of care and adequate training hours there have been calls to decrease the strict limitations upon the USA junior doctors, in addition to a perceived erosion of professionalism amongst trainees. Some surgeons have gone so far as to announce the “total destruction of medicine, more specifically general surgery” because of the loss of continuity of care of a surgeon until the completion of the task at hand or a patient’s care by a single surgeon is no longer a requirement (Fischer, 2005). The author states an employee is employed by another, working for salary in the service of an employer is which is unique and independent of a professional ethos, guided by the benefit and welfare of the individual (the patient) whom the professional must. In this general surgeon’s view, the doctor “should be available 24 hours a day” for continuity of
patient care but due to the societal forces surgeons and physicians have been “reduced to the level of employees”. His viewpoint and sentiments underpin a significant element of distrust towards occupational safety regulations that historically foreign to the surgical mainstream.

Since the implementation of the ACGME work-hour limits there have been several studies observing whether the changes have made a difference in patient outcome measures, with varying results. All-cause mortality of all Medicare patients admitted with medical or surgical between 2003 and 2005 showed no worsening or improvement in mortality for the first 2 years after its implementation (Volpp, 2007). A survey by the same author showed similar results among Veterans Affairs patients. Similar studies specific to internal medicine inpatients concurred with the patient mortality outcome; however, intensive care unit utilisation, discharge to rehabilitation or home and pharmacist interventions were significantly reduced. Unfortunately simply reducing the number of work hours does not necessarily increase sleep duration; a survey of trainees showed an increase of only 22 minutes per night, 1 year after ACGME guidelines were implemented (Landrigan, 2006). It should also be noted that shift work alone does not necessarily improve sleep longevity or quality. It is difficult to estimate the impact the ACGME changes have had upon resident education due to lack of evaluation of the formal educational aspects of surgical training.

Despite the above evidence occupational safety researchers have argued for more stringent guidelines for the United States similar to the EWTD of a 48-hour week. Similar shift lengths are not tolerated in other industries, as noted above the Federal Aviation
Administration for a single pilot caps the length of flying time to 8 hours even with autopilot function on the aircraft.

1.3.7 The Australian experience

The Australian experience in safe working hours has been of much slower incremental change towards a guideline-based approach rather than strict legislated rules governing junior doctors. The Australian Medical Association developed a model of safe working hours for doctors in training, which was later adopted by RACS as a report on *Appropriate Working Hours for Surgical Training in Australia and New Zealand* (Surgeons, 2007). This guide provides for a 56-65 hour working week dependent upon the surgical specialty. A 2011 audit by the Australian Medical Association found the longest shift increased from 39 hours in 2006 to 43 hours in 2011 and the maximum hours worked during the audit week increased from 113 to 120 in 2011. The average total hours worked for all hospital doctors was 55.1 hours. To mitigate the rise in working hours a National Code of Practice for rostering and shift hours provided guidelines for managing fatigue (Ama, 2005). The Code is not prescriptive; instead it contains a set of design principles for work roster scheduling and on-call rosters. The aim is to minimise the risk to health and safety of the medical staff and those affected by their actions. The guidelines state the occasions whereby doctors are required to work more than 10 hours should be minimised, a minimum of eight-hour breaks between shifts is preferable to allow for sleep and a forward shift rotation should be used to decrease individual adaptation problems. The guide notes that the principles may not be in accord with current practice, for example permanent night shift staff.
There are significant deviations from the current RACS Safe Working Hours Standard by current Australasian trainees (O'grady, 2010). Around 13% of trainees work in excess of the recommended 70 hours per week in the Standard; more significantly 5% exceeded 80 hours per week. This proportion would be increased if the on-call time spent in work-related activities (such as on the phone) were also included. The majority of trainees often sleep less than the 8 hours of uninterrupted sleep recommended due to 24-hour on-call commitments, particularly in neurosurgery and paediatric surgery. A significant number reported routinely having interrupted sleep of less than 5 hours. This is concerning given the established issues resultant from chronic partial sleep deprivation mentioned previously.

The optimal number of working hours in surgical training and service delivery has not been defined; a survey of Australian and New Zealand trainees responded approximately 60 hours per week would be an appropriate balance for technical vs non-technical training needs although 55 hours were preferable (O'grady, 2012).

The Australian Medical Association (AMA) guide mentions the use of caffeine; however, it has been suggested that although caffeine may restore laparoscopic performance to rested levels they do not reduce errors (Aggarwal, 2011). The use of stimulating agents such as caffeine, taurine and modafinil has been resisted in Australia due to the potential damage to the reputation of the medical profession compared particularly with the experience of footballers, cyclists and athletes (Watters, 2014). A self-regulating profession where fatigue is recognised and managed well is preferable.
1.4 Acquisition of surgical skills

1.4.1 Acquisition of skills

There are several stages in the acquisition of a motor skill – cognition, integration and automation (Kopta, 1971). Cognition describes a state where the learner understands the task to be performed. Integration requires the individual to rehearse the fine motor skills associated with the task. Automation involves the skill being performed without a significant cognitive burden or requirement of external guidance.

The variability of technical ability in the performance of a particular task is well known but not well understood. “Neuroplasticity” is a term that has been used to describe the differential changes that take place in the brain dependent on the experience and training of the individual. Functional neuroimaging studies of test subjects demonstrating a skill being performed has shown a difference in experts of smaller and more refined neuronal networks compared with novices (Leff, 2008). There is a significant burden placed upon the cortical areas of attention and control such as the anterior cingulate gyrus and the prefrontal cortex in learning new skills; this burden should decrease with increased automaticity of the task.

1.4.2 Assessment of surgical competency – skills and judgment

There are two major components required for a surgeon to be “competent” – operative skill as well as judgment.

Operative decision-making is a non-technical skill that is difficult to teach and for which it is difficult to identify standardised methods of assessment. There are potentially great advantages in using simulation applications for assessment of these decision-making
skills. Such an application should include an outline of the operative steps and hazard components of each surgical procedure; a situational awareness component, a ranking system of the options considered within this assessment combined with immediate feedback to the candidate (Andersen, 2012). Such an algorithm combines deliberate practice with incremental challenges to the surgical trainee allowing for progression of simulated experience.

1.4.3 Role of simulation

Simulated settings can be used to emulate a variety of procedures in gynaecology, urology, general surgery and endoscopy. High-fidelity equipment is best suited for minimally invasive technologies, particularly laparoscopy and robotic surgery.

The role of simulation in surgical training is as yet poorly understood and has been implemented in Australia in a haphazard manner. Multiple reasons for this exist, including the myriad of simulation methods with variable validity, lack of clear evidence for transferability of simulated skills to the operative setting, variable access to formalised education programs and the cost of simulation being borne by both the trainee and the institution.

1.4.4 Comparing simulated environments

Before a comparison of simulated environments can be undertaken, a common language is required for assessment of such instruments. There is a distinction between the reliability of a simulator and validity.

- **Reliability** refers to the precision with which consistent results can be achieved with the trainer.
- **Validity** refers to several measures of correlation between the test being applied and whether this accurately corresponds to a given reference.

Within this heading several definitions can be derived.

- **Construct validity** refers to how accurately the physical aspects of a simulator measures the real-life version of the given operative skill.

- **Content validity** concerns whether the representative skill (or skills) being assessed accurately reflects the entirety of an operative procedure. This requires knowledge of the specifications of the task and an authoritative body identifying which important aspects make up a representative sample of the procedure.

- **Face validity** is a less-standardised version of content validity. It refers to how closely a constructed simulator appears to reflect a given task “on face value”.

Whilst it may help engage the trainee in utilising the simulator given its obvious correlation to the real scenario it may be less objective in measuring a task given it is not compared to a given set of criterion within a domain, as in construct validity. More accurate measures of specific tasks are often possible with simulators of low face validity without environmental distractors.

### 1.4.5 Method of teaching

A high degree of heterogeneity exists in the way simulators are used. The simulator itself can be provided for the candidate to use at leisure, or as part of a self-directed program of skill reproduction. Whether a mentor is available to correct errors and provide feedback contributes to the benefit derived from a simulated environment.
Until recently little emphasis had been placed upon the cost of simulation, both in equipment and educator costs. A recent review by the American College of Surgeons analysed the cost and logistics of their tissue-based technical skills curriculum (Henry, 2014) This curriculum covered 35 modules to 35 general surgery trainees across 5 clinical years at an annual cost of $110,300 (or $3,150 per trainee). The modules include cadaveric dry simulation, animal-based plus the paper-based educational assistance. Whilst the cost seemed great the high satisfaction rate of trainees with tissue-based models should be noted.

1.4.6 Transferability of skills

It is assumed that a particular set of skills to complete an operative task should be reproducible to the same level of competency in a surgical environment compared with an artificial scenario. However, there is relatively little evidence for this and where validation has been proven it is not consistent for all parameters. Some tasks, such as those involving minimally invasive equipment, are particularly suited to simulation. Laparoscopic surgery can be simulated inside a box trainer demonstrating use of equipment, simulated depth perception and fine motor control.

In considering the validity of surgical simulators the following the general parameters that need to be considered include (Fairhurst, 2011):

- Resemble the operative task and environment as closely as possible
- Measure surgical performance accurately
- Educate surgical trainees reliable to defined levels of proficiency
Assessing the validity of whether simulation affects real-life performance is highly variable (Sturm, 2008). For example, in simulation of laparoscopic cholecystectomy, to assess the trainee performing the procedure in a live patient the methods of assessment varied from the performance of the entire laparoscopic procedure, only the clip and cut portion or excision of the gallbladder from the hepatic bed.

A more generalised systematic review of transferability assessed studies involving laparoscopic and endoscopic procedures. The consensus showed participants who reached proficiency in simulation-based training performed better than those that did not in the patient-based setting (Dawe, 2014). However, it also showed simulation-based training was equally effective as patient-based for colonoscopy, laparoscopic camera navigation and endoscopic sinus surgery.

1.4.7 Types of simulators

A wide variety of simulators exist in common practice. Equally there exists an array of methods to categorise such models. A task on a human patient can be simulated using a technological model of varying fidelity, animal models (such as porcine) and cadaveric teaching.

The most useful determinant of a simulator in surgical practice would be the sophistication of the technology used in replicating a task. Low-fidelity indicates a simple method of emulating a procedure, such as a box trainer. Box trainers, also known as video trainers in the literature describes a simple box emulating an abdomen or thoracic cavity, with a video camera projected onto a connected monitor. The top of the box is fenestrated, simulating laparoscopic ports to allow placement of intra-cavitary laparoscopic
instruments. One of the most readily used simulators is the McGill Inanimate System for Training and Evaluation of Laparoscopic Surgery box (MISTELS). This has been incorporated into the Fundamentals of Laparoscopic Surgery (FLS) program developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) described below.

High-fidelity simulators, also described as computer simulators, utilise more sophisticated technology to simulate a set of tasks. This group includes virtual reality simulators, endoscopy and endovascular trainers. Virtual reality simulators make up a significant portion of the simulator market. They can emulate single tasks or entire procedures. Often the technology includes methods of measuring performance parameters and allowing this to be compared to other trainees or a pre-approved standard. The most common high-fidelity simulator in practice is the Minimally Invasive Surgical Trainer – Virtual Reality (MIST-VR) created in Gothenburg, Sweden. Others include the LapMentor (Symbionix USA Corporation) and the LapSim (Surgical Science, Sweden).

1.4.8 Formalised education programs

Prior to 1997 there was a relative paucity of formal education in the underlying principles and skillset of laparoscopic surgery. The Fundamentals of Laparoscopic Skills course was launched by the Society of American Gastrointestinal and Endoscopic Surgeons in 2004 in response to the ad-hoc nature of learning laparoscopy without mentors already skilled in the craft. Surgeons were instead being funded through industry courses that were variable in format and utility. Credentialling was equally non-existent.
SAGES created a training package involving didactic learning, clinical judgment and manual skills in conjunction with the MISTELS system and Karl Storz Endoscopy America (Vassiliou, 2010) This tool encompasses tasks of peg transfer, pattern cutting, placement of ligating loop and knot tying, both intracorporeal and extracorporeal which forms the basis of the technical component of the FLS program. FLS certification requires passing a written examination and a proctored manual skills examination in the FLS simulator. Since its inception the FLS course is now used in 30 countries and being followed with a second program based upon flexible endoscopy.

There is an ongoing use of un-coordinated teaching programs in many countries; however, the recognised SAGES program has been adopted by most Western nations, with the American board requiring FLS competency before candidates are able to complete their fellowship exam.

1.5 Training context – becoming a surgeon

1.5.1 Surgical training

Surgical training has undergone multiple iterations given the competing forces of supervision, working hours, trainee expectations and feminisation of the medical workforce. The more modern structured training scheme is based upon the William Halsted model of John Hopkins Hospital from the 1890s (Ong, 2009). The model is predicated upon training the surgeon from a basic skillset to a pre-determined exit point deemed to meet an agreed competency standard of the discipline.
1.5.1.1 United Kingdom

The British surgical model of training arose from The Royal Colleges of Surgeons of Edinburgh and England, which were formed in 1778 and 1800 respectively. Both were formed after separation from the Incorporation of Barber Surgeons. The early part of the 20th century saw the common use of curriculum-based learning and examination-based education in addition to the traditional apprenticeship-type model.

One of the most significant changes to modern surgical training in the United Kingdom came about from the “Modernising Medical Careers” (MMC) program for postgraduate medical training introduced in 2005. The re-structuring of training involved a new Foundation Program for the first 2 years followed by early run-through specialty training, administered by a new board, the Postgraduate Medical Education and Training Board. The system was heavily criticised by the Tooke Inquiry (Tooke, 2008) on multiple levels due to loss of employment for 14,000 doctors, having to choose a specialty too early (mid-way through their second postgraduate year) and loss of local appointment of registrars. The unfortunate timing of the introduction of the European Working Time Directive coinciding with the MMC program led to grave concerns over the competence of a surgeon trained over less years with far lower contact time in the hospital (Delamothe, 2008).

1.5.1.2 United States of America

The American surgeon William Halsted introduced a model not dissimilar to the Germanic surgical model of its day requiring a structured model of learning with graduated responsibility during their years of service under close and direct supervision (Harvey, 1981). This new model of “residency” became the model for all American
hospitals. Later, advanced specialty training was developed as subspecialties utilising the same model of learning with an exit examination. The approach popularised by Halsted was later summarised as the “see one, do one, teach one” method which greatly simplified an entirely new educational model of ensuring the most junior surgeon is supervised through each procedure until the end of training, but also requires a significant amount of preparation, forethought and aptitude on the part of the surgical candidate. Each resident is expected to be familiar with the operation to be performed, the patient, the postoperative care and particular complications of the procedure.

Whilst the American model began to be followed in other parts of North America and Asia, within 50 years the residency structure began to change from the so-called “pyramidal” Halstedian model of autocracy to a “rectangular” model developed by surgeon Edward Churchill at Massachusetts General Hospital (Pellegrini, 2006). The previous system had unintentionally produced a number of surgeons who had not completed their full training requirement and relied upon a “dominant master – docile apprentice” relationship, which was denounced by surgeons and residents alike. Instead, Churchill introduced a system where all residents starting the surgical residency program were expected to complete full training requirements after 4 years. The latter model is more recognisable today as the foundation of a program that must be completed before being awarded a Fellowship in North America, the United Kingdom and Australasia.

1.5.1.3 The Australian model

RACS was founded in 1927 in Melbourne, Victoria. Unlike its English counterpart upon which it was predicated, the College engaged in both training and accreditation of
surgeons. The initial foundation of training was based upon the British model of teaching, like many antipodean processes of its time. From the 1970s the residency model of training began to take hold paving the way for consultant surgeons to engage in teaching of selected junior doctors. This continued from the beginning of surgical training through manual skills and clinical judgment through to the Fellowship examination in a system recognisably similar to the present day. Of note, the Australian model restricted control of the training program to the College of Surgeons alone, whereas in contrast the USA and UK models included influences from the government of the day, hospitals, the universities and surgeons (Wilde, 2003). The apprenticeship model that this sustained involved practising surgery on public hospital patients almost exclusively, a model not without its detractors due to the inherent inequality of “using” those who cannot afford private health insurance as a training mechanism for less experienced doctors.

The most modern iteration of the Australasian surgical training program for all disciplines is known as SET (Surgical Education and Training). It incorporates 5-6 years of streamlined competency-based training aligned to the College’s nine core competencies. The selection of junior doctors as residents into SET carries an expectation of completion of the competencies by the end of the training program, an embodiment of the “rectangular” selection and training model popularised by Churchill. This is based upon a similar program in the United Kingdom introduced as part of their MMC program a decade prior.

Both the Australian/NZ and UK programs discarded the previous basic surgical training (BST) scheme where applicants spent 2-4 years as “BSTs” in either resident or service
registrar positions before vying for a limited number of advanced training posts. The
difficulty with this model, as found previously under Halsted, is the pyramid-shaped
character of candidate attrition. Many half-trained surgeons either wasted years
attemting to become accredited surgical trainees, or settled for the lower rank of “MRCS
surgeons”, those who legitimately undertook minor surgical procedures in an
unsupervised environment without accreditation as a full Fellow of the Royal College of
Surgeons, a distinction most members of the public would not appreciate.

1.5.2 Metrics of surgical training

There have been considerable efforts towards creating a systematic objective structure of
surgical skill assessment in the simulated setting. Ideally this would have commensurate
validity in the operating room. The testing structure would encompass a wide range of
potential tasks simulating skills needed for basic open surgery, laparoscopic surgery, or
newer methods such as those employed in robotic surgery. The ideal device for objective
assessment of real surgical procedures would “… automatically, instantly and objectively
provide valid and reliable data regarding performance in the operating theatre” (Darzi,
1999). One of the founding workshops towards this endeavour was conducted in 2002
resulting in a foundation for communication and a standardisation of definitions,
measurements and criteria (Satava, 2003). It has been proven that current methods of
assessment such as examination, log books and direct observation of procedural skills
lack validity and reliability (Moorthy, 2003).

A method of assessment of laparoscopic skills is the system McGill Inanimate System for
Training and Evaluation of Laparoscopic Skills (MISTELS). It is a system devised with
structured tasks and objective grading systems specifically to assess trainees early in the
acquisition of laparoscopic skills (Vassiliou, 2006). To determine inter-rater reliability
two observers were used to assess 12 subjects performing each task twice. They estimated
internal consistency between tasks scores. Reliability scores were noted to be good for all
tasks for the first assessment of the test-retest.

The FLS program was developed by the Society of American Gastrointestinal and
Endoscopic Surgeons (SAGES). It is a tool designed to assess the criteria of knowledge,
judgment and manual skills (Swanstrom, 2006). The program requires an educational
component of didactic online course materials as well as a box simulator with low-fidelity
tasks for completion. The candidate is evaluated at the end of the course via a multiple-
choice questionnaire and a proctored manual skills test. The tool has been validated for
reliability for evaluating competence (ibid.) as well as predicting intraoperative
laparoscopic skill (Mccluney, 2007). McCluney correctly predicted the simulated scores
of the assessed candidates on the FLS simulator were independently predictive of external
validity of surgeons in the operative setting in terms of laparoscopic skills. Similarly, the
validity of the FLS program was investigated when used in the gynaecological setting in
terms of training and laparoscopic experience. Whilst the testing phase of the program
appropriately measured psychomotor skills, it was found not to discriminate between
novice and advanced operators (Zheng, 2010). Given the validation of the FLS has been
completed it is now a certification requirement for the American Board of Surgery as well
as RACS for trainees in General Surgery.
1.5.3 Lessons learned from modern changes – laparoscopic surgery

The birth of laparoscopic surgery in the early 1980s heralded the largest paradigm shift in the technical aspect of surgery since the dawn of general anaesthesia. Never before had there been a single modular shift in the underlying method of several subspecialties that forced many within the surgical workforce to learn and retrain in a relatively rapid manner. The disciplines affected the most were general and paediatric surgery, urology and gynaecology.

Whilst the principle of laparoscopy was documented in 1910 in Sweden by Hans Christian Jacobaeus, it was not until the first solid state camera was introduced in 1982 that the first stage of human laparoscopy commenced. With a rod-lens optical system and cold light fibre-glass illumination the ability to perform an abdominal laparoscopic procedure became possible, initially for the diagnosis of liver disorders and abdominal trauma (Vecchio, 2000). The first laparoscopic cholecystectomy using a four-trocar technique was performed in 1987 by Philippe Mouret in Lyon, France.

The introduction of laparoscopic surgery was associated with a significantly increased number of complications. This was most acutely seen in the uptake of laparoscopic cholecystectomy. Reports of the early use of laparoscopy in cholecystectomy indicated a higher bile duct injury rate when compared with open cholecystectomy (Deziel, 1993). The assumption of a ‘learning curve’ which would result in a lowering of the rate of bile duct injuries as the procedure became common practice was proven correct in limited studies (Olsen, 2000). However, bile duct injury is still at a higher rate in laparoscopic (0.3-0.7%) than in open surgery (0.125%). Other research from the West of Scotland Cholecystectomy Audit Group found a late downward trend in bile duct injury,
suggesting a prolonged learning curve for the procedure (Richardson, 1996). An
Australian study confirmed the two-fold relative increase in risk of bile duct injury, major
bile leak and other bowel or vascular injury in laparoscopic compared to open surgery
(Fletcher, 1999). Multiple reasons are given for the aberration in risk; some attribute the
increased injury rate to specific technical factors such as tenting, confluence and
diathermy. Others attribute the increased error rate to more general factors based upon
operating on two-dimensional images of the operative field instead of its three-
dimensional reality (Slater, 2002). Relative inexperience of using a laparoscopic approach
accounts for approximately two-thirds of injuries (inexperience being defined as less than
200 cases) with the remaining third attributed to fundamental errors in technique (Archer,
2001). Of note, intraoperative cholangiography was found to be helpful in diagnosis of
bile duct injury (80.9%) in contrast to 45.1% if it was omitted, particularly if the surgeon
defined themselves as “selective” rather than “routine” users of cholangiograms. The
protective effect of routine cholangiograms has been confirmed in another study
(Fletcher, 1999). Overall the inappropriate manner in which minimal access surgery was
adopted has moved the paradigm of surgical training away from an apprenticeship model
towards formal skills programs (Hamdorf, 2000).
2. Aims and hypothesis

2.1 Age of participants

Over the past decade there has been a significant increase in the number of postgraduate medical programs in Australia. Older graduates potentially acquire skills in a different manner; previously learned processes are likely to be acquired more rapidly than those for whom it is a new skill. For older graduates, skills that are unfamiliar to them or without overlap with known cognitive processes may be more difficult to acquire. The study hypothesis proposes a slower acquisition of newly-acquired skills related to increasing age of graduate. In addition it would follow that the more complex the task, represented by the duration required to complete the skill, the more likely older graduates would take longer to complete it. The study aims to assess the effect age has upon the acquisition of surgical skills that are previously unfamiliar to the candidate.

2.2 Fatigue level

Fatigue is more likely to affect the performance of newly-acquired surgical skills. The time taken to complete tasks will be longer and this will be more evident as the complexity of the task increases. This study aims to quantify the effect of fatigue on the timing of laparoscopic skills by comparing the same candidate at the start and end of the day. It is hypothesised that older medical graduates are more likely to be affected by fatigue.
2.3 Related factors

Data were collected on typical sleep hours prior to candidates entering the study. The study protocol did not permit sleep between the test and re-test timings. Longer hours of sleep the night before the study is proposed to be associated with faster completion times in skills testing.

Data were recorded on the activity conducted by the subject between the two tests. We anticipated those who worked during the day would be more fatigued than those undertaking leisure activities.

The amount of caffeine ingested by the research subjects was also recorded. A small number of caffeine-based drinks would negate the effects of fatigue. However, as the number of drinks increases, the effect of the caffeine can be self-defeating. At some point there will be a negative impact on motor skills, similar to drug-induced tremor.
3. Methods

3.1 Study population

Our study population included medical students from two medical schools, interns and residents who volunteered to participate in the trial. Data were collected on the educational level/year and university attended by the subject. Each participant attended a 30-minute briefing session then completed three attempts at each task under timed conditions. Each morning session was completed between 7:00 and 11:00am.

The participants were instructed to undertake their usual tasks during the day but not to sleep or drink caffeine. Hours of sleep prior to the trial date were recorded, in addition to the day’s activity (work, sleep or “other”). During the afternoon sessions participants similarly completed the three tasks, with three timed trials of each and the best of three recorded.

Participants self-rated their fatigue score prior to the morning and afternoon sessions on a scale of 1-5. In addition the regular hours per week each used a computer game was recorded, differentiating between role-player games and console computer games.

3.2 Deconstruction of tasks

3.2.1 Task protocols

The tasks utilised were based upon the FLS course created by SAGES. These are in turn based upon the gold standard of MISTELS program (Sages, 2014). The current FLS course includes the peg transfer task but also requires several more difficult tasks to be
practised including precision cutting and extra/intra-corporeal knot-tying. Due to the
difficulty of these tasks we have modified the protocol to include cube stack and string pull-through. The included tasks – cube stack, peg transfer and string pull-through – can be seen in the following figures (Figure 3.1 – Figure 3.4).

Figure 3.1  Basic Laparoscopic Skills Set-up

Figure 3.2  Task 1 – Cube stack
The first task was used to ensure familiarity with the laparoscopic box and “Maryland” dissector/grasping forceps. The camera is centrally placed within the laparoscopic simulator box. The participant is asked to transfer one cube to stack it above the second cube. The transferred cube must remain atop the first cube; if it falls the candidate must replace the cube on top of the other.

Figure 3.3  Task 2 – Peg transfer

Each pegboard is loaded with six coloured rubber rings. The participant is given two Maryland dissectors to grasp the pegs. The camera is centrally placed within the laparoscopic simulator box. To perform the task the object is grasped with the participant’s non-dominant hand, retrieved from the peg then passed to their dominant hand mid-air, then placed onto a peg of the opposite side. This process is reversed before completion of the task. There is no order placed upon which peg is grasped or placed on either side, as long as it is passed to the opposing side. The activity ceases if a peg is passed or dropped outside of the field of view and the candidate starts again.
Figure 3.4  Task 3 – String pull-through

A board with 12 circular eyelet screws are placed centrally in the simulator. A 30cm piece of thin rope is passed to just within the first peg. The candidate is shown how to pass the rope through each of the pegs towards the opposite side of the board. An order for passing the rope through the board is indicated by a series of arrows marked upon the board itself.

3.2.1.1  Timers

For consistency only one of two researchers timed each candidate completing the three tasks. Each task was first demonstrated to the candidate before being timed three times. The first and second tasks commenced upon grasping the cube with the Maryland dissector, ending with letting go of the cube as long as the cube was stable. Timing for the third task commenced once the Maryland was placed upon the rope that is within the first peg, then finished once the candidate let go of the rope after being passed through the last peg.
3.2.1.2 Effect of fatigue

One of the aims of the study was to simulate the fatigue of the operator at the end of a working day, compared to the beginning. To this end the timings of the skill being assessed was recorded and the difference between start and end time was correlated against their skill timing. In addition the self-rated fatigue score and hours of sleep the night before were taken into consideration.

3.3 Statistical analysis

Statistical analysis was performed using both SPSS and Microsoft Excel. Demographic data were compared using $t$ tests. Skill timings were expressed as the mean score for three timings of each tasks per session. Pearson’s correlation coefficient ($r$) was calculated for assessment of skill timings against factors of age, fatigue and sleep. Significance is defined as a $p$ value of less than 0.05.
4. Results

4.1 Subject demographics

A total of 75 medical students and junior doctors participated in the study (over a 2-week period). All candidates who completed the morning training program completed the evening session. Each of the participants within the study cohort rated their own laparoscopic skill set as “new” to laparoscopic skills or “limited”. Thus none of the participants were experienced in laparoscopic skills, and accordingly are termed “novices” in this study.

The demographic characteristics are shown in Table 4.1. The ages of candidates ranged from 18-43 years. The median age was 23 years. 34 (45.3%) were female and 41 (54.6%) male. 67 participants were medical students (89.3%) with 56 from The University of Western Australia and 11 from the University of Notre Dame Australia. Of the eight participants who were not medical students, three were interns, two were residents, two were registrars and one was a postgraduate PhD student. None of the participants had additional simulator exposure during the day.
Table 4.1  Demographic experience

<table>
<thead>
<tr>
<th>Participants</th>
<th>Overall</th>
<th>n = 75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Students</td>
<td>67 (89.3%)</td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td></td>
<td>23 ± 4.7 years</td>
</tr>
<tr>
<td>Gender</td>
<td>F</td>
<td>34 (45.3%)</td>
</tr>
<tr>
<td>University</td>
<td>UWA</td>
<td>56 (83.5%)</td>
</tr>
<tr>
<td></td>
<td>UNDA</td>
<td>11 (16.5%)</td>
</tr>
<tr>
<td>AM times median (range)</td>
<td></td>
<td>9:00AM (7:30 – 10:00)</td>
</tr>
<tr>
<td>PM times median (range)</td>
<td></td>
<td>4:00PM (15:30 – 17:30)</td>
</tr>
<tr>
<td>Sleep hours (mean ± SD)</td>
<td></td>
<td>6.6 hours (1.12)</td>
</tr>
<tr>
<td>Day-time activity</td>
<td>Work</td>
<td>32 (43%)</td>
</tr>
<tr>
<td></td>
<td>Study</td>
<td>14 (19%)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>29 (38%)</td>
</tr>
<tr>
<td>Computer game playing (hours/week)</td>
<td>Role-player games</td>
<td>n = 62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Range 1 – 7 hours</td>
</tr>
<tr>
<td></td>
<td>Console</td>
<td>n = 64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Range 1 – 7 hours</td>
</tr>
</tbody>
</table>

### 4.2  Skill timing

Three tasks were completed during the morning by all participants (Table 4.2). The median start time was 09:00 hours. The average self-reported fatigue level on a scale of 1-5 was 1.9 (range 1 not fatigued to 4, most fatigued). The average time to complete a specific task was: cube stack 7.7 seconds, peg transfer 13.4 seconds and string pull-through 29.4 seconds.

The median end of day time was 16:00 hours (range 15:30-17:30). The mean self-rated fatigue score was significantly higher at 2.8 (range 1 not fatigued to 5, most fatigued). The average time to complete each specific task in the evening session was markedly lower than in the morning session: cube stack 5.0 seconds, peg transfer 10.7 seconds,
string pull-through 21.2 seconds. Each of these comparisons was statistically significant (p < 0.05).

### Table 4.2 Skill timings (AM vs PM)

<table>
<thead>
<tr>
<th></th>
<th>AM time</th>
<th>PM time</th>
<th>t-test *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatigue score (mean ± SD)</td>
<td>1.9 ± 0.8</td>
<td>2.8 ± 0.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cube stack (mean ± SD)</td>
<td>7.7 ± 7.5</td>
<td>5.0 ± 5.9</td>
<td>0.002</td>
</tr>
<tr>
<td>Peg transfer (mean ± SD)</td>
<td>13.4 ± 6.1</td>
<td>10.7 ± 3.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>String pull-through (mean ± SD)</td>
<td>29.4 ± 13.3</td>
<td>21.2 ± 7.7</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* Paired t-test comparison of AM vs PM values

### 4.3 Age and skill timing

It was hypothesised that the older participants would perform significantly slower on each of the three tasks. The three measures of cube stack, peg transfer and string pull-through were correlated according to age using Pearson’s correlation coefficient for both the morning and evening pairs. In terms of age correlating with time to complete tasks, the results were not statistically significant (Table 4.3).

### Table 4.3 Age vs skill timings

<table>
<thead>
<tr>
<th></th>
<th>AM vs age</th>
<th>PM vs age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r*</td>
<td>p value</td>
</tr>
<tr>
<td>Cube stack</td>
<td>-0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>Peg transfer</td>
<td>-0.08</td>
<td>0.50</td>
</tr>
<tr>
<td>String pull-through</td>
<td>-0.11</td>
<td>0.35</td>
</tr>
</tbody>
</table>

* r is Pearson’s correlation coefficient

### 4.4 Age and fatigue

A statistically significant association was noted when the age of the participant was correlated with the self-rated fatigue score at the end of the day (p=0.017). However, it is
to be noted, that this finding was inconsistent with the results collected in the morning (Table 4.4).

Table 4.4 Age vs fatigue self-rated scores

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Range</th>
<th>r</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM times</td>
<td>1.9</td>
<td>(1-4)</td>
<td>0.167</td>
<td>0.074</td>
</tr>
<tr>
<td>PM times</td>
<td>2.8</td>
<td>(1-5)</td>
<td>0.244</td>
<td>0.017</td>
</tr>
</tbody>
</table>

4.5 Fatigue and skill timings

The study investigated whether there was an association between the time taken on each of the three tasks and the self-rated fatigue score for each participant, divided into morning and afternoon sessions. Of the correlation coefficients, only one was deemed statistically significant (peg task timing in the morning, $p = 0.043$). None of the afternoon times nor the other morning timings were related to the participant’s subjective fatigue level (Table 4.5).

Table 4.5 Fatigue and skill timing

<table>
<thead>
<tr>
<th></th>
<th>AM times</th>
<th>PM times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r$</td>
<td>$p$</td>
</tr>
<tr>
<td>Cube stack</td>
<td>-0.032</td>
<td>0.393</td>
</tr>
<tr>
<td>Peg transfer</td>
<td>-0.199</td>
<td>0.043</td>
</tr>
<tr>
<td>String pull-through</td>
<td>0.174</td>
<td>0.067</td>
</tr>
</tbody>
</table>

4.6 Sleep and skill timing

Each candidate reported the number of hours sleep they had on the night prior to the task trials. The mean was 6.6 hours (standard deviation 1.1 hours). There was no correlation between the time taken for each candidate to complete the task and the reported hours of sleep the previous night (Table 4.6).
Table 4.6  Sleep and skill timing

<table>
<thead>
<tr>
<th></th>
<th>AM times</th>
<th>PM times</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p</td>
</tr>
<tr>
<td>Cube stack</td>
<td>-0.033</td>
<td>0.390</td>
</tr>
<tr>
<td>Peg transfer</td>
<td>0.180</td>
<td>0.060</td>
</tr>
<tr>
<td>String pull-through</td>
<td>0.047</td>
<td>0.344</td>
</tr>
</tbody>
</table>
5. Discussion

5.1 Major findings

This research demonstrates the novice candidates performed better in regards to time taken during the end-of-day session for each of the three skill sets learned the same morning. This remained despite a statistically significant increase in the self-rated fatigue score of the participants.

The result is initially surprising in the context of a hypothesised decrease in time to complete a task; as a fatigued practitioner would be assumed to have a longer time in demonstrating a repeated skill. Similar previous studies utilising laparoscopic training modules and knot-tying techniques did not find significant differences across three sessions for errors of time over 2 days (Figert, 2001). The conclusion is not necessarily transferable as significant decrements in proficiency upon the psychomotor and cognitive skills of surgical residents have been demonstrated dependent upon the time of day during which their on-call period occurred, not solely upon the duration of the call (Brandenberger, 2010).

Whilst the above results are seemingly incongruous they do fit with the demonstrated neurocognitive ability to learn a skill and process the steps at a subconscious level, even if they are not actively repeating or thinking about the steps involved during the day. Described as “neuroplasticity” (see “Acquisition of surgical skills”) there is a differential change dependent on the individual’s experience and training to date prior to the skill being learned. This ability to process new skillsets in the novice practitioner leads to a
potential area for further research wherein the “experienced” practitioner, such as an expert surgeon with known experience in laparoscopic surgery may exhibit an even greater capacity for transferability of skill set to a new technique.

The study findings are in part reassuring as the positive correlation of age and self-rated fatigue at the end-of-day session did not translate into a slowing of skill time. Only one of six skill sessions showed a correlation between self-rated fatigue score and timing (morning peg transfer task) which of itself is unlikely to be a true reflection of correlation between the two. A great deal of research has been undertaken into the negative effects of the advancing age of the patient upon surgical outcome. In contrast there has been a lack of rigorous endeavour seeking to ascertain the impact of surgeon frailty upon outcomes, other than in an anecdotal fashion (Blasier, 2009).

Our results demonstrate that despite the fact that entirely new skills were being tested, the practitioner was not only able to compensate for this but was also able to demonstrate the skill with greater proficiency later the same day.

This was unaffected by age, either de novo or in combination with fatigue scores. Perhaps of greater interest was the lack of significant correlation between age and skill timing, or fatigue vs. skill in the majority of tests (five of six did not show a statistically significant difference).

5.2 Limitations

Combining the factors of age and fatigue upon the more quantifiable aspects of surgical skill would require more resources and study of further influences in addition to those
measured in this study. These limitations include a larger cohort of novice participants, more complex laparoscopic skills and the ability to measure accuracy and error in the performance of each surgical skill, with acceptable inter-rater reliability. A greater panel of surgical biometric measures would ideally replace a simple timing measure of each task. The inconsistency of older participants rating their own fatigue level as being higher in the evenings should be tempered against the fact that as self-rated fatigue scores they are not objective measures of fatigue.

5.2.1 Cohort size

Whilst 75 participants is a reasonably large number, for further subset analysis to be accurate the cohorts would need to be of a larger size. Of greatest utility would be a more sizable cohort of candidates in each age group with further diversity in the amount of hours spent on computer skills (console and role-player games) to allow this to be included in future studies with meaningful results. Despite the above recommendation of increasing the number of participants, it is to be noted that other investigations of surgical skill learners have reported results on far fewer participants (Macdonald, 2003, Seymour, 2002).

5.2.2 Complexity

The lack of complexity in the given tasks is relevant in terms of transferability to real-world measurability of skill. The tasks themselves are low-fidelity representations of several key elements of routine laparoscopic surgery. These skills are taught as part of the FLS course described earlier. Utilising more high-fidelity representations could potentially benefit the external validity of the study and allow further interpretation of
results such as accuracy, error measures and more complex procedures such as clipping vessels.

There are several issues that prevented use of such equipment in this study. Firstly the learning curve of complex, non-validated tasks makes them ill-suited to novice laparoscopic users such as the medical students and resident medical officers who participated in this study. To enable usable results would require many more hours of training in each skill with either a trainer or laparoscopic simulator program prior to measuring the effects of age and fatigue. Secondly the availability of such machines is limited for large numbers of research subjects.

5.2.3 Error measures and fidelity

Indeed whether a higher fidelity version of the surgical skill being imitated would improve transferability from the simulated setting to the operating room is highly controversial. Concerns around the cost of high-fidelity equipment, without demonstrated improvement in teaching outcomes, has led to a number of “bench models” being tested. A study involving “hand-on” training using models of differing fidelity prior to the performance of microsurgical anastomoses on rats, showed no difference in scores, irrespective of model fidelity (Grober, 2004).

A literature review of studies contrasting high- and low-fidelity training in anaesthetic auscultation skills, surgical techniques and complex management skills sought to clarify the relationship between simulator fidelity and learning. Both high- and low-fidelity learning resulted in improvements in performance compared against control groups without interventions. However, nearly all the studies demonstrated a lack of significant
advantage of one over the other with average differences ranging from 1-2% only (Grober, 2004). The most important conclusion seems to be that virtual reality is likely to be successful only if systematically integrated into an education and training program utilising objective assessments of technical skill improvement close to the learning experience (Gallagher, 2005). Such assessment requires well-defined surgical metrics to measure consistent performance to an agreed standard.

Whilst the problems associated with sleep-deprived practitioners are well known, the ability to quantify these risks within surgery have proven more difficult to accurately ascertain. This study did not demonstrate an association between hours of sleep the night prior and deterioration of performance. Prior research has demonstrated neurocognitive performance decrements as well as a loss of insight into the effect of fatigue in subjects affected by sleep deprivation (Durmer, 2005). According to Pilcher and Huffcutt, sleep loss in residents promotes neuro-behavioural impairment. Further, this finding applies irrespective of whether the sleep loss is short-term sleep loss (within the past 24 hours) or chronic partial sleep restriction (less than 6 hours per night for at least 1 week) (Pilcher, 1996). Of note, the effects of chronic sleep loss are cumulative.

The imposition of fatigue via sleep loss is markedly different to tiredness due to work and might be considered a limitation of this study. A better method of ensuring the practitioners were truly fatigued could be to impose a sleep restriction then retest the candidates within a 24-hour period. Such studies upon work-imposed sleep restriction have concluded acute sleep deprivation of less than 4 hours does not affect performance (Grantcharov, 2001). To replicate this outside a work-imposed setting would require
supervision on the part of the researchers as well as methods of ensuring the safety of the practitioners before and after the test. The current study did not replicate this scenario. Given the trend towards shift work medicine, with various working time restrictions in place in most first-world countries, it is becoming less of a pertinent issue for the major subspecialties of surgery.
6. Conclusion

The study sought to provide an insight into the effect of age and fatigue upon newly-acquired surgical skills. End of day fatigue was found not to impair performance of basic laparoscopic tasks. Indeed, it seems there is a demonstrable improvement in skill timing at the end of the day regardless of intervening activity. The hours of sleep the night before did not correlate with skill timing, nor did the age of the participant.

As more medical schools move towards postgraduate education concerns have been raised around whether the advancing age of interns and “junior” doctors will have a negative impact on performance. This is especially relevant to the consideration of whether, in addition, longer surgical training schemes are required. It is reassuring that older candidates did not perform differently in the completion of the three surgical tasks.

Future research into the technical aspects of training are required. Ideally skill accuracy and measures of error would be taken into account utilising rigorous application of surgical biometrics. A graduated series of tasks of increasing complexity would be more externally valid using a combination of low and high-fidelity surgical simulators. Finally, the imposition of sleep restriction could be created artificially; however, as medicine internationally moves towards regulated hours such studies are likely to be less necessary. It is imperative future iterations of surgical technology utilises surgical simulation to a significant extent, as a complement to hands-on training, prior to its application upon surgical patients.
7. References


