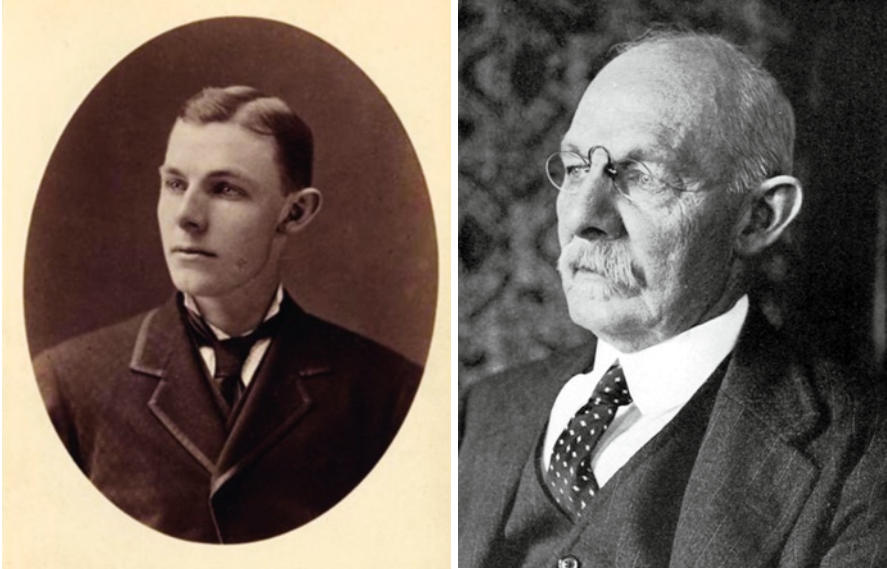


# Chapter 1

## Agents of Change

### Halsted: The Beginning of the Modern Surgical Training Program

In August 1922, Dr. William Stewart Halsted returned to Baltimore from his summer retreat (High Hampton, North Carolina) with symptoms of choledocholithiasis. He had had his gallbladder successfully removed at Johns Hopkins Hospital in August 1919 and had remained symptom free until this occasion. However, despite a successful reoperation and attentive care by his colleagues he developed pneumonia and pleurisy of which he died on Thursday, 7 September 1922. Even at the start of the twenty-first century the stature of Halsted's contribution to medicine remains undiminished despite revelations about his private life. He was educated at Yale University (where there is no record of him ever borrowing a book from the library) and the University College of Physicians and Surgeons in New York, after which he took up a position as a house physician at the New York Hospital. One of his earliest contributions to patient care that still exists to this day is the introduction of the temperature, pulse, and respiration recordings to the patients chart. In 1884, Halsted commenced a series of experiments on himself and his colleagues investigating the anesthetic powers of cocaine. Unfortunately, during the process of these experiments Halsted and several of his colleagues became addicted to cocaine. Although hospitalized and treated for his addiction on at least two occasions it emerged after his death that his addiction had been treated by switching from cocaine to morphine, to which he remained addicted throughout his life. Most of his peers and colleagues assumed that his addiction to cocaine had been cured during his hospitalization in Rhode Island. However, private diary notes by Sir William Osler (the first chief of medicine at the Johns Hopkins Hospital), who was also Halsted's physician, clearly indicated that Halsted was never able to eliminate his daily use of morphine. Osler noted that Halsted could work comfortably and maintain his "excellent physical vigor" on three grains of morphine per day (about 180 mg). In later years (i.e., 1912), Osler noted that Halsted had reduced his consumption to about 1½ grains/day.



**Fig. 1.1** Dr. William Stewart Halsted (1852–1922). Portrait of William Stewart Halsted, Yale College class of 1874 (Photograph courtesy of the Yale University Manuscripts & Archives Digital Images Database, Yale University, New Haven, CT)

During his time living and working in New York, Halsted was outgoing, gregarious, sociable, energetic, and vigorous. However, when he moved to Baltimore he led a quiet and scholarly life which bordered on reclusive. He appeared to be a solitary figure with few friends or close acquaintances at Hopkins throughout his career. Dr. John Cameron (1997) a subsequent Chairman of Surgery at Johns Hopkins speculated that this marked change in Halsted's demeanor probably resulted from his humiliation by his addiction. Despite this burden Halsted's contributions to surgery included recognizing the importance of submucosal suturing for intestinal anastomosis, development of radical mastectomy for cancer of the breast, and development of a surgical procedure for inguinal hernia repair. He was also the first surgeon to promulgate the philosophy of safe surgery by introducing rubber gloves into the operating room and advocating that the gentle handling of tissues, careful hemostasis, and the use of meticulous surgical technique. Even though general anesthesia had been introduced in the early nineteenth century, during Halsted's time most surgeons still operated rapidly with little concern for hemostasis (as though the patient was still awake during the procedure). By the time of his death the American surgical community had accepted his philosophy of safe surgery and took full advantage of the operative benefits anesthesia afforded for technical skills application during surgery. However, Halsted's (Fig. 1.1) single greatest contribution to modern healthcare was the development and implementation of the first system to train young surgeons.

## Surgical Training

In the latter part of the nineteenth century there were no formal training programs in surgery. Individuals who were qualified or experienced in the practice of surgery were not particularly interested in training other surgeons who might then become competitors in private practice. Halsted devised a surgical training program at Johns Hopkins Hospital based on what he had learned from a number of well-known European surgeons. He established a surgical training program that was based on strict dedication to the bedside study of disease and graded responsibility with a clinical teacher. He also established that surgery was best learned by hands on experience and education within a hierarchical training program. His training program consisted of a 1-year internship, followed by 6 years as an assistant resident. If successfully navigated, this period culminated in 2 years as a house surgeon. The term (surgical) “resident” comes from Halsted’s training program. His trainee surgeons were discouraged from marrying, lived in the hospital where room, board, and training were provided in exchange for service to the hospital 24 h a day, 7 days a week. This pattern of long work hours and service commitment was wedded (and indeed probably still is in some quarters) to the persona of becoming a surgeon.

The training system developed by Halsted at John Hopkins Hospital was based on the German system, and as such, it was autocratic and pyramidal in structure. Although eight residents entered training in first year, four of these positions were for only 1 year and of the remaining four, only one became a surgeon and the other three spent long periods of time with no guarantee of becoming staff surgeons. The system aimed at producing one outstanding surgeon that then went on to become a Professor (Grillo 2004). In this sense, the Johns Hopkins training model worked very well as graduating surgeons went on to establish training programs at other distinguished institutions such as Yale, Duke, and Brigham Hospital based on the Halsted training model.

One of the first major changes to this training system was introduced by Dr. Edward Delos Churchill (1895–1972) at Massachusetts General Hospital (MGH). Churchill was critical of the Halstedian training model for two reasons. The first was that the training model developed at Johns Hopkins unintendedly produced a number of poorly trained surgeons because they left training after completion of 1 year or shortly afterward. The second reason was that the training system was somewhat authoritarian in that it depended on the formation of a relationship between the dominant master surgeon and the docile trainee. Churchill believed that this was anti-intellectual (Pellegrini 2006). Churchill proposed a new training structure at MGH which intellectually and philosophically departed considerably from the traditional Halstedian approach to training. In the traditional MGH training structure there were eight residents, six of which were trained for 2 years with two being advanced to the 4th year level. The first change that Churchill advocated was that the total number of residents entering the training system in any given year should be decreased to six, with four of them obtaining a 4-year training (which

meant they were fully trained) and two would remain in the hospital and might be destined to become master surgeons at MGH or go on to take up leading academic positions at other institutions. However, he also proposed that the residents should be trained by a group of master surgeons rather than a single dominant personality. One of his intentions in implementing this training structure appears to have been to minimize or obviate the self-aggrandizing and authoritarian relationship which was such an integral part of the apprenticeship model of training (Grillo 2004). The rectangular system proposed by Churchill would remain, with minor modifications, the core structure of the residency training systems in the USA until the end of the twentieth century. As Pellegrini (2006) points out, Churchill believed that the residency training structure should be implemented in such a way that it allowed for flexibility which enabled individual residents to follow up any specific interests they had and it also allowed the acquisition and development of proficiency. This idea of proficiency and flexibility in progression will be discussed further in Chap. 8.

The enactment of the servicemen's readjustment act of 1944 (or the GI BILL) was a defining moment for surgical training. It was created to train medical officers returning from World War II and marked the first time that surgical trainees in the USA received payment (Sheldon 2007). Although surgical trainees received some payment, the life of surgical trainees remained austere up until the 1970s. Just as in Halsted's era, they rarely left the hospital which provided them with meals, white uniforms, laundry, and somewhere to sleep. The next major change in surgical trainees' lifestyle was initiated by the Medicare and Medicaid Act of 1965 which provided a mechanism for surgical trainees to receive financial compensation for care that they had previously given for free (Sheldon 2007). Surgical trainees observed huge increases in their salaries as a result of this landmark health care legislation. Possibly as a result of these changes and changes in attitudes to work during the 1960s and 1970s, the restrictive lifestyle of the Halstedian training paradigm began to lessen. Trainees began to marry and move out of the hospital and were no longer available for service delivery 24 h a day (Wallack and Chao 2001). Despite these changes, surgical training remained arduous with the trainees working long hours, frequently on call every other night and going home only after the work was completed. Indeed, this work ethic and culture persisted in surgical training until the late twentieth century when the death of a young woman in a New York hospital brought into question the safety of having trainee doctors who had been on duty for long hours take care of sick patients.

## Agents of Change

The Halstedian approach to training in surgery existed for the best part of a century, and despite its critics was effective. Indeed, it was so effective that the rest of medicine, more or less imitated the training program that had been pioneered at the Johns Hopkins and refined at MGH in Boston. However, all that was to change in the latter part of the twentieth century. Surgical training was about to undergo a

paradigm shift in the way surgeons were trained and this revolution would impact on how *all* doctors were trained. Thomas Kuhn (1962) argued that science does not progress via a linear accumulation of new knowledge but undergoes periodic revolutions or so-called *paradigm shifts* in which the nature of scientific enquiry within a particular field is abruptly transformed. He also argued that paradigm shifts do not occur by accident, but instead are driven by agents of change. An agent of change can be something as simple as a growing body of evidence that demonstrates significant anomalies against an accepted paradigm or approach (such as the Halstedian approach to training). At some point in the accrual of this evidence the discipline is thrown into a state of crisis. During this crisis, new ideas, perhaps ones previously discarded are tried. Eventually, a new paradigm is formed which gains its own new followers and an intellectual battle takes place between the followers of the new paradigm and those who held on to the old paradigm. However, Kuhn (1962) argues that this is not simply an evolution of ideas, but a revolution. Furthermore, the new paradigm is *always* better and not just different. Paradigm shifts have occurred most frequently in the natural sciences and have always been dramatic, particularly in what appeared to be a stable and mature area of research and study. For example, Lord Kelvin in an address to an assemblage of physicists at the British Association for the Advancement of Science in 1900 famously stated that “there is nothing new to be discovered in physics now. All that remains is more and more precise measurement” (Smith and Wise 1989). Five years later, Albert Einstein published his paper on special theory of relativity which fundamentally challenged the bases of Newtonian mechanics (Pais 2005). In this chapter we will argue that the agents of change impinging on the discipline of surgery were worldwide, varied, pervasive and persuasive and cried out for a different and better way to prepare surgeons for operating on patients. The outcome of this revolution has been precisely that. In the coming pages, we will describe what we believe have been the agents of change.

## **The Libby Zion Case, USA**

Libby Zion was an 18-year-old woman admitted to the New York Hospital, Cornell Medical Center, with fever, agitation, delirium, and strange jerking movements of her body on March 4, 1984 (Asch and Parker 1988). Within 8 h of admission, she was dead. The exact cause of her death was never conclusively demonstrated although it is widely suspected that she died because of serotonin syndrome. Her father, a lawyer and New York Times columnist, believed that she had died as a result of inadequate care from overworked and inadequately supervised medical residents. Her father conducted a very public and emotional campaign against the hospital and doctors and claimed that the death of his daughter was tantamount to murder. In 1987, the intern and resident who cared for Libby Zion were charged with 38 counts of gross negligence and/or gross incompetence. The grand jury considered evidence that a series of mistakes contributed to Libby Zion’s death

including the improper prescription drugs and the failure to perform adequate diagnostic tests. Under New York law, the investigative body for these charges was the Hearing Committee of the State Board for Professional Medical Conduct. The hearing committee unanimously decided that none of the 38 charges against the two residents were supported by evidence (Spritz 1991). However, the final deliberations on this case rested with another body, the Board of Regents. In a surprise decision the Board of Regents voted to censure and reprimand the resident physicians for acts of gross negligence. Although the decision did not affect their right to practice as doctors and was overturned in the appeal Court in 1991, the decision of the Board of Regents caused considerable concern among practicing physicians in New York City and nationally.

As a result of a grand jury indictment of the two residents, the New York State Health Commissioner (David Axelrod) established a blue ribbon panel of experts headed by Dr. Bertrand M. Bell from Albert Einstein College of Medicine to address the problems in residency training. The Bell Commission put forward a series of recommendations that addressed several patient care issues one of which was resident work hours (Asch and Parker 1988). In particular, they recommended that residents could not work more than 80 h a week or more than 24 consecutive hours. In 2003 the Accreditation Council for Graduate Medical Education (ACGME) adopted similar regulations for all accredited medical training institutions in the USA (Philibert et al. 2002). These changes in training practices shook the medical establishment to its very roots and continue to reverberate. In general, both residents in training and attending surgeons thought that the quality of care given to patients had been negatively affected by the introduction of an 80 h work week (Whang et al. 2003) despite objective evidence that found no differences in the quality of care received by patients or quality of education experience received by trainees pre-and post the introduction of the ACGME work hour limit (Hutter et al. 2006).

## European Working Time Directive

In the USA, pressures to reduce the number of hours worked by doctors in training emanated from an incident that occurred in medicine. However, pressures to reduce the number of hours worked by junior doctors in training in the UK and Europe derived from an entirely different source. The European Union Working Time Directive (EWTD) was first drafted in 1993 and was introduced to improve the living and employment conditions of workers within the European Economic Community. The most commonly known clause within the directive is that which is associated with a 48-h working week and the opt-out associated with it (Adnett and Hardy 2001). The directive, adopted in 1993 and amended in 2003 has been incrementally introduced in European nations with the final stage introduced on August 1, 2009. When first adopted in November 1993 the working time directive excluded the air, rail, road, sea, inland waterway and lake transport, sea fishing, offshore work, and the activities of doctors in training, as it was decided that these sectors

required individual specific legislation to accommodate working time measures. A further directive covering these sectors, known as Horizontal Amending Directive was adopted on August 1, 2000. The entitlements in this legislation include:

- A limit of an average of 48 h work a week, up to maximum of 60 in any one week
- A limit of an average of 8 h work in 24, but no more than 10
- A right for night workers to receive free health assessments
- A right to 11 h rest a day
- A right to a day off each week
- A right to an in-work rest break if the working day is longer than 6 h
- A right to 4 weeks paid leave per year

It is fair to say that few issues have generated as much controversy or legal challenges as this directive, particularly within the medical profession. Doctors' leaders argued that if their American colleagues found it challenging to train doctors in the ACGME mandated 80 h/week, they would find it impossible within a 48-h time frame. When the legislation was first introduced there was some compromise with its implementation. However, in 2008 the European Parliament voted to end the right of individual doctors in member states to opt out of the directive. There is little doubt that the EWTD posed considerable organizational difficulties for its implementation in medicine. It was also widely believed that the directive compromised the training of future surgeons (Lowry and Cripps 2005) and as such was unpopular with UK trainee and trainer surgeons. In the UK, the implementation of the EWTD meant that doctors had to move to a shift pattern of working. This type of work practice allows important information loss about clinical care during the increased number of handovers. However, it should be remembered why this legislation was introduced in the first place.

The practice of working at night was made possible by Edison's commercialization of electric light in 1882. This extended the working day to 24 h a day, 7 days a week; fatigue caused by working longer hours and round-the-clock became a major social issue. The emerging labor movement in the early 1900s eventually influenced work hour regulations and laws and the concept of hours of service regulation emerged. As a result, the issue of workplace fatigue became intertwined with labor pay and rights issues and led to regulatory limits on work duration and minimums of off-duty time duration in all transportation modes by the middle of the twentieth century (Moore-Ede 1993). Research conducted in the late 1970s demonstrated that the brain's circadian clock exerted strong control over time, duration, and stages of sleep. Because of this circadian regulation of sleep, there was an important difference between sleep opportunity and the amount of sleep it was possible to obtain during that opportunity. For example, even under ideal sleeping conditions, individuals who slept 8 h when they went to bed at 11 p.m. would only sleep 6 h if they went to bed at 3 a.m., and only 4 h if they went to bed at 11 a.m. even though they had been kept awake all night (Åkerstedt and Gillberg 1986; Daan et al. 1984).

Around about the same time studies reporting on the link between sleep pattern, fatigue, and accidents started to appear in the scientific literature (Dembe et al. 2005; Samkoff and Jacques 1991; Schuster and Rhodes 1985; Wojtczak-Jaroszowa and Jarosz 1987). Furthermore, a series of major industrial accidents occurred between 1970 and 1990 where human operating errors related to fatigue were linked. These included:

- The Chernobyl nuclear reactor explosion in the Ukraine, where 237 people suffered from acute radiation sickness of whom 31 died within the first 3 months, 135,000 people were evacuated from the area (Hallenbeck 1994).
- Flixborough, where a chemical plant explosion destroyed an English Village on 1 June 1974, killing 28 people and seriously injuring 36.
- Piper Alpha North Sea oil platform which exploded and killed 167 people in 1988.
- In the city of Bhopal, India, December 3, 1984 a poisonous gas cloud escaped from the Union Carbide India Limited (UCIL) pesticide factory. The cloud contained 15 metric tons of methyl isocyanate (MIC) covering an area of more than 30 square miles. The gas leak killed at least 4,000 local residents instantly and caused health problems for at least 50,000 others.

These types of incidents led to in-depth analyses of how they occurred and precipitated the evolution of a systematic understanding of the relationship between human operative error and fatigue. These efforts have been greatly informed by the work of Prof. James Reason (1990) who had been an advisor to the Royal Air Force and NASA on human error. Reason pointed out that most major accidents are the result of multiple latent system errors and not just by the immediately obvious act of error by the human controller (Reason 1990). He suggested that many accidents were in fact not accidents but a series of events that set the occasion for an adverse event to happen. All that it took for these “accidents” to occur was the right set of environmental circumstances which invariably revolved around a person or persons. Avoidable human factors such as fatigue due to sleep deprivation which are known to be associated with increased probability of errors should not be allowed to happen, should be specifically anticipated and dealt with at a senior organizational level.

The relationship between errors in medicine and sleep deprivation was established in the 1970s (Friedman et al. 1971). Friedman et al., reported that interns made almost twice as many errors reading electrocardiograms after an extended workshift (i.e., 24 h or more) than after a night’s sleep. More recent studies have shown that surgical residents make up to twice as many errors in the performance of a simulated laparoscopic surgical task after working overnight than after a night of sleep (Grantcharov et al. 2001). Although the literature as a whole suggests that sleep deprivation causes substantial decrements in physicians’ performance (Gaba and Howard 2002; Weinger and Ancoli-Israel 2002) this is not accepted by some in the medical community. For example, Dr. Malcolm Lewis, Director of Postgraduate Education for General Practice at the School of Postgraduate Medical and Dental Education in Cardiff University (Wales) and chairman of the Committee of General Practice Education Directors, (a UK-based forum) has questioned the relationship between fatigue, work hours, and medical errors. In an interview for a Canadian medical Journal, he stated that “the



perceived advantages [of the EWTD] are of less tired workforce and of improved patient safety as a result. This is of course theoretical and I am not aware of a body of evidence to support the perception” (Villaneuva 2010). It is of course possible that Dr. Lewis is unaware of the large volume of well-controlled, quantitative research that directly links decrements in performance to fatigue and sleep deprivation. However, what is less believable is that he is unaware of the results from studies in medicine, published in leading medical journals that have directly established a relationship between medical error, sleep deprivation, and fatigue. For example, Landrigan et al. (2004) investigated the effects of reducing intern work hours on serious medical errors in intensive care units, using a prospective, randomized study design. They compared performance of interns working according to a traditional schedule with extended (i.e., 24 h or more) work shifts every other shift (i.e., and every third night call schedule) and a schedule that eliminated extended work shifts and reduced the number of hours worked per week to 63 h. They found that interns made significantly more serious medical errors when they worked frequent shifts of 24 h than when they worked shorter shifts. Interns made approximately 21% more serious medication errors during the traditional schedule and they were also five times more likely to make a serious diagnostic error. Furthermore, the data for this study was from direct observation of the intern’s performance rather than self-reported.

From the wealth of published data on the effects of fatigue on performance in a variety of industrial and occupational settings, the results are unambiguous, i.e., it significantly degrades human performance and considerably increases the probability that an error will be enacted. However, fatigue poses a particular and very real problem on a daily basis for particular types of surgical specialties such as neurosurgery, ophthalmic surgery, otolaryngology surgery, plastic surgery, or any type of surgery requiring a microsurgical techniques (e.g., tendon repair, vascular anastomosis, etc.). Physiological tremor arises from mechanical and neuromuscular sources and is made worse by a number of factors such as dehydration, caffeine, cigarettes, anger, fear, stress, and fatigue (Patkin 1977). Unfortunately for surgeons using this particular technique, increased hand tremor is a natural result of normal operating procedures and is a simple fact of the job resulting from muscle fatigue (Slack and Ma 2007). Surgeons who employ microsurgical techniques on a regular basis go to great lengths in an effort to control their hand tremor. These include biofeedback training, maintenance of a healthy lifestyle, ensuring they are well hydrated before operating, abstaining from coffee and nicotine, and sometimes resorting to taking beta-blockers (Elman et al. 1998; Ferguson and Jobe 2004; Harwell and Ferguson 1983). However, within these operators, fatigue is recognized as the most tremor producing factor and situations which induce fatigue prior to operating should be, where possible, avoided. Unfortunately, injuries which require the application of these types of surgical skills occur irregularly but commonly at inconvenient times such as during the night, in a patient admitted to accident and emergency as a result of a road traffic accident. The only safe approach to this type of scenario is for the surgeons to maintain a state of readiness, and that means minimizing surgical interventions by fatigued surgeons.

Other factors that need to be kept in mind are the findings from the 1960s, relating performance to levels of arousal and the presence of others, who would appear to

have important implications for the practice of surgery. Scientific investigation of the effects of an audience dates back over a century. In 1904, a German researcher conducted experiments concerned with muscular effort and fatigue. He noted that the subjects were able to exert far more muscle effort on days that he watched as compared to the days on which they were not watched (Meumann 1904). However, Zajonc (1965) suggested that the situation was not that simple, and that the presence of others energized individuals and increased their drive level. An increase in drive strengthens the dominant response of the organism, i.e., the response most likely to occur. At the same time, an increase in drive weakens responses that already are weak. What this means is that under stressful conditions individuals will respond in a way that is very familiar or is easier for them. For example, in a simple or well-learned task, familiarity with what is required exists or the task has been practiced several times, thus the strongest and most likely response is the one that is appropriate and correct. In a complex and difficult task on the other hand, the strongest response is likely to be the wrong one. Complicating matters further is the Yerkes–Dodson law (Yerkes and Dodson 1908) which establishes an empirical relationship between arousal and performance. The law dictates that performance increases with physiological or mental arousal, but only up to a point. When levels of arousal become too high, performance decreases. The process is often illustrated graphically as a curvilinear, inverted U-shaped curve which increases and then decreases with higher levels of arousal. What this means for the practicing surgeon is that the skills which are very familiar and or well trained are more likely to be performed well in situations of stress whereas surgical skills which are unfamiliar and or novel to them will not be performed well. These predictions have profound implications for trainee surgeons, particularly in stress provoking situations such as in accident and emergency or in the operating room when unanticipated complications occur. This type of response is most likely to occur for surgical trainees (of *any* level of seniority) if the skills they are required to practice are novel, unpredictable, not under the control of the individual, and required to be practiced in the presence of an experienced evaluator (e.g., a more senior surgeon part of whose job is to appraise their performance). Lupien et al. (2007) have reviewed the evidence of the psychophysiological effects of stress hormones (glucocorticoids) on the process of forming long-term memory. They concluded that mildly elevated levels of glucocorticoids enhanced long-term memory formation. In contrast, long-term memory formation is impaired after adrenalectomy (which causes chronic low glucocorticoid levels) or after exogenous glucocorticoids administration (e.g., subcutaneous injection) thus demonstrating an inverted U-shaped performance reminiscent of the Yerkes–Dodson effect.

## **The Bristol Case, UK**

In 1989 Dr. Stephen Bolsin moved from the Brompton Hospital in London to take up position as a consultant cardiac anesthetist at the Bristol Royal Infirmary. He very quickly formed the opinion that the Bristol Royal infirmary had significantly

higher complication and mortality rates than what he was accustomed to, and probably higher than the national average complication rate. He identified that too many babies were dying during heart surgery and although he raised his concerns with senior hospital administrators, they refused to investigate. He eventually took his concerns to the media and the ensuing investigation became known as *The Bristol Case* (Smith 1998). The Bristol case centered around three doctors: Mr. James Wisheart, a former medical director of the United Bristol Healthcare trust; Mr. Janardan Dhasmana, a pediatric and adult cardiac surgeon; and Dr. John Roylance, a former radiologist and Chief Executive of the Trust. The central allegations against these individuals were that they knowingly allowed to be carried out or carried out operations on children, knowing that the mortality rates for these operations in the hands of the surgeons were higher than the national average. Furthermore, the operating surgeons were accused of not communicating to the parents the correct risk of death for these operations in *their* hands.

One of the earliest concerns raised by Dr. Bolsin was that Mr. Wisheart's operations took up to three times as long as those at the Brompton Hospital and were associated with more complications. By 1993, he had concluded a formal audit that showed that while national average mortality rate for repair of tetralogy of Fallot was 7%, Mr. Wisheart's was 33% and Mr. Dhasmana's was 25%. The audit also showed that while national average mortality rate for atrioventricular canal surgery was 10%, Mr. Wisheart's was 60% and Mr. Dhasmana's was 17%. By the time Mr. Wisheart had retired in 1995, seven of the last eight children that he operated on died. At about the same time Mr. Dhasmana began performing arterial switch procedures on neonates. Although he stopped after performing the procedure on 13 patients, 9 of them died and 1 of them had sustained serious brain damage. A team in Birmingham (87 miles north-east from Bristol) who were performing the same procedure had only 1 death in 200 patients. Mr. Dhasmana's results in older children were also cause for concern with a mortality of 30% compared to about 1% in centers of excellence.

Although Dr. Bolsin contacted the Department of Health in 1993, it was not until 1995 that a new consultant cardiac surgeon was appointed. The Bristol Royal Infirmary Inquiry was chaired by Professor Sir Ian Kennedy and was a landmark case in that it changed how medicine was learned and practiced in the UK (Bristol Royal Infirmary Inquiry 2001). Mr. Wisheart and Dr. Roylance were struck off the medical register and Mr. Dhasmana was disqualified from practicing pediatric cardiac surgery for 3 years. The enquiry concluded that a substantially and statistically significant number of excess deaths (between 30 and 35) occurred in children between 1991 and 1995. The mortality rate over the period was probably double the rate in England at the time for children under one and was even higher in children under 30 days (Bristol Royal Infirmary Inquiry 2001).

Dr. Richard Smith (1998) in his editorial in the British Medical Journal (BMJ) seemed to summarize very well the impact that the Bristol Case would have on medicine in the UK and the international reverberations from it when he said that medicine would be transformed by the case. It had thrown up a long list of important issues that British medical practitioners would take years to address which has proved correct. These included:

- The need for clearly understood clinical standards
- How clinical competence and technical expertise are assessed and evaluated
- The training of doctors in advanced procedures
- How to approach the so-called learning curve of doctors undertaking established procedures
- The reliability and validity of data used to monitor doctors' personal performance

There were many other issues raised, which included an appreciation of factors other than purely clinical ones that affect clinical judgment performance and outcome, team leadership, and responsibility and communicating with patients and families. One of the more uncomfortable issues that The Bristol Case raised was the need for doctors to take prompt action at an early stage when a colleague was in difficulty in order to offer the best chance of avoiding damage to patients and a colleague to put things right.

Just like the Libby Zion case in New York, the problems that were encountered in Bristol met with intense and sustained political, media, and public interest both in the UK and internationally (Walshe and Offen 2001). It also brought into sharp focus issues relating to professional regulation, clinical competence, and health care quality improvement in medicine. Furthermore, much of this debate was conducted on the front page of national newspapers and television chat shows. One of the aspects of this case that was very striking to the UK general public was the fact that senior hospital managers (some of whom were doctors themselves) knew that some of their surgeons were underperforming and despite frequent, often public, protestations from clinical colleagues they did not act. The trust between doctors and patients had been compromised by this case and the general public was unambiguously aware of this fact. It was a very public failure of doctors and the health care system.

## **The Neary Case, Ireland**

Our Lady of Lourdes Hospital is a 340-bed public hospital located in Drogheda, County Louth, Ireland. It provides acute-care hospital services, including a 24-hour emergency department for the population of County Louth and the North East of the Irish Republic. It serves a population of about 110,000 out-patients, and more than 20,000 in-patients. It is also a very busy maternity hospital with more than 4,000 births a year. It had previously been owned by the Medical Missionaries of Mary who were founded in 1939 by Mother Mary Martin. It was the first hospital founded by the order. The order set up the hospital, then called the "International Missionary Training Hospital," in Drogheda. It served the people of Drogheda and the surrounding regions, and it also served to train personnel for hospitals in Africa. Nurses and patients for the most part refer to the hospital as "The Lourdes," a shortened version of its full title Our Lady of Lourdes Hospital. Many of the older consultants

referred and still refer to the hospital as the IMTH (International Missionary Training Hospital). The hospital provided services that accorded with the ethos of the Roman Catholic Church – including its teachings on human reproduction (Harding Clark 2006). In sum, the hospital had a strong Irish Catholic history.

In 1998 Dr. Michael Neary was asked by his employer to take administrative leave for 2 weeks from his post as a consultant gynecologist at the hospital after concerns about his clinical practice were expressed by two experienced midwives. In 1998 three senior consultant obstetricians from major teaching hospitals in Dublin were asked to review the practices of Dr. Neary between the years 1996 and 1998. Seventeen caesarean hysterectomies identified from the maternity theater register were to be reviewed. The three obstetricians met with Dr. Neary and considered each case in turn. However, of the 17 cases they were asked to review, 8 were excluded on the bases that Dr. Neary had informed them that these were consented hysterectomies necessitated because of the prohibition in the hospital of tubal ligation. Their reports exonerated Dr. Neary's clinical practice. The health board was uncomfortable with this report and asked for a fourth opinion. They requested a review from a very senior practicing obstetrician consultant at St Mary's Hospital in Manchester where he was lead clinician in the labor ward which had more than 6,000 births each year. The Manchester obstetrician reviewed the same nine cases previously reviewed by the three obstetricians acting for Dr. Neary. His report stated that he had major concerns about Dr. Neary continuing to practice as a consultant obstetrician. Unfortunately this report was leaked to the press and it made national headlines throughout the subsequent investigation into the case.

The Medical Council of Ireland received complaints from 15 patients who had procedures carried out by Dr. Neary during the years 1986–1990, including ten complaints alleging unwarranted peripartum hysterectomies. The Medical Council commenced its enquiry on 6 June 2000 and continued taking evidence over the next 2 years. These ten complaints included the nine cases reviewed by Dr. Neary's review group and the English obstetrician from Manchester. On 29 June 2003 the Medical Council's Fitness to Practice Committee found that the facts in relation to the ten complaints alleging unwarranted peripartum hysterectomies were proved and that Dr. Neary was found guilty of professional misconduct. The Medical Council determined that his name should be erased from the General Register of Registered Medical Practitioners.

The Inquiry into peripartum hysterectomy at Our Lady of Lourdes Hospital, Drogheda, chaired by Judge Maureen Harding Clarke S.C. was established by the Government in 2004 following the decision of the Medical Council to remove Dr. Neary from the Register of Medical Practitioners. They found that a total of 188 peripartum hysterectomies were carried out in the 25-year period between 1974 and 1998. Of the 188 cases, 129 cases were attributed to Dr. Neary. An average consultant obstetrician would perform about five or six operations in an entire career. The rate of Caesarean hysterectomies at the hospital for the period was 1 in every 37 Caesarean sections. In contrast, the rate at other hospitals of similar ethos ranged from 1/300 to 1/254 Caesarean sections. Although concerns were raised in 1978/1979 by the then matron her concerns were not heeded. Indeed, no issues were raised

about Dr. Neary's practice until 1998 when the two midwives raised the issue with the Health Board solicitor. Furthermore, the unit was passed for training by the Royal College of Obstetricians and Gynaecologists in 1987 and again in 1992. The unit was also passed by the Royal College of Surgeons in Ireland (RCSI) for undergraduate training and by Bord Altranais for midwifery training. The inquiry also found that 23.4% of obstetric hysterectomy records (44 cases) for the period 1974–1998 were missing and were intentionally and unlawfully removed from the hospital with the object of protecting those involved in hysterectomies or protecting the reputation of the hospital. In 40 of the 44 cases the birth registers were also missing (Harding Clark 2006).

This case is important because, just like the Bristol Case in the UK, it changed the public's perception of doctors in an otherwise very conservative country where doctors were held in very high esteem. Indeed, when Dr. Neary was suspended there was enormous support for him and outrage by many of his patients and colleagues at his treatment. However, as the facts of the case emerged sympathy turned to anger. In particular, there was considerable anger at what the public perceived as the medical profession's attempts to cover up its own mistakes. The three consultant obstetricians who conducted the original review were perceived as trying to protect their own and this was specifically commented on in the inquiry (Harding Clark 2006).

This was the worst case of medical misconduct ever to have occurred in Ireland. It resulted in significant modifications to the Medical Practitioners Bill in the country which made continuing professional development and education compulsory for all medical practitioners. It also established in law for the first time a statutory obligation for competence assurance for medical practitioners. More than a decade after questions first started to be asked about this case, medical practitioners are still dealing with the impact of the changes initiated by it.

## **The Bundaberg Hospital Scandal, Australia**

In 2003 Dr. Jayant Patel, who trained in India and the USA, was appointed surgical medical officer and later promoted to the post of Director of Surgery at Bundaberg Base Hospital, Bundaberg, in central Queensland. Over the following 2 years he operated on about 1,000 patients of whom 88 died and 14 suffered from serious complications (Burton 2005). However, all this may not have happened had the 2003 registration of Dr. Patel by the Queensland Medical Board been more rigorously scrutinized (Van Der Weyden 2005).

Although Dr. Patel obtained his preliminary medical education in India, and was awarded a Masters Degree in Surgery, he completed his intern year and residency training at the University of Rochester School of Medicine in Upstate New York. While working at a hospital in the city of Buffalo (New York) in 1984, Dr Patel was cited by New York health officials for failing to examine patients before surgery and placed on 3 years clinical probation. In 1989, he moved to Portland, Oregon, to

work at the Kaiser Permanente Hospital system. Staff reported that his practices (including hygiene) were unusual and bizarrely he would frequently turn up to perform surgery on patients, some of whom were not even his responsibility. In some cases, the surgery was not required and in other instances he caused serious injuries and death to patients. After a review in 1998 the Kaiser Permanente Hospital system in Portland restricted his practice and banned him from doing liver and pancreatic surgeries and required him to seek second opinions before performing surgeries. After a further review, the Oregon Board of Medical Examiners made the practice restrictions statewide in September 2000 and in relation to a separate (previous) case, New York State health officials required him to surrender his license to practice in April 2001.

After this Dr. Patel moved to work for the Queensland Health Department in Australia. Unfortunately, they employed him without conducting due diligence regarding his qualifications and experience. Had this review been conducted by the Queensland Medical Board they would have discovered his placement on probation in 1983 by Rochester Hospital, New York, for “gross negligence”; they would have discovered the Oregon Board of Medical Examiners placing restrictions on his surgical practice; and they would also have discovered the threat by New York state to have his license to practice revoked before he voluntarily surrendered it. None of this information was disclosed by Dr. Patel at the time of his appointment. Also, just like the Neary case in Ireland and the Bristol Case in the UK, the concerns about Dr. Patel’s performance at Bundaberg Hospital did not emerge from clinical governance systems but from concerns expressed by individual doctors and nurses about his surgical performance and prowess. Once again it was a communication from a member of the nursing staff about this matter which led to a question being tabled at the Queensland Parliament, which eventually resulted in the establishment of a Commission of Inquiry (Van Der Weyden 2005).

After the issues pertaining to Dr. Patel were raised in the Queensland Parliament, an award-winning Australian journalist succeeded in uncovering Patel’s past which resulted in a media frenzy surrounding the case. Dr. Patel left Australia shortly after this and returned to his home in Portland, Oregon. A warrant was issued for his extradition from the USA on three charges of manslaughter, five charges of causing grievous bodily harm, four of negligent acts causing harm, and eight charges of fraud. He was extradited to Australia on 21 July 2008. He was tried in the Queensland Supreme Court for the unlawful killing of three patients and grievous bodily harm to a fourth. On 29 June 2010, Dr Patel was found guilty of four charges and on 1 July 2010 he was sentenced to 7 years in prison. Even after his sentencing there was considerable public anger as many believed his sentence was too lenient considering the gravity of the charges and the lack of remorse that Dr. Patel showed during the trial.

Just like the UK and Irish cases outlined here, a similar pattern appears to be emerging. At the center of this pattern is a doctor who is underperforming but fails to recognize that he is or fails to do anything about. In fact, Dr. Patel went to great lengths to cover up and deny his failures. Deficits in his clinical performance were not brought to light by clinical governance systems but by concerned members of

staff who had to go outside the health care system to raise their concerns. Once the case reached the public scrutiny brought about by the media, the facts of the situation exploded on to the front pages of the Australian and world press. Similar to the Bristol and Nery cases, the Queensland Health Care system came under considerable criticism. It was depicted as a gigantic dysfunctional conglomerate with a corporate center that was more concerned with performance indicators, revenue generation, and cost control than people. Furthermore, of the 64,000 employees of Queensland Health, fewer than one in five were clinicians (Forster 2005).

The Medical Board of Queensland has since introduced extensive measures for the registration of overseas doctors, including receiving a certificate of good standing on each and every jurisdiction in which a doctor has practiced and getting the primary degree, registration, and transcripts of applicants verified by the Educational Commission for Foreign Medical Graduates International Credentialing Service. Just like the UK and Ireland, corporate and professional medicine moved to put structures in place that would ensure that this type of incident did not occur again. However, by the stage that this had happened the good standing of medicine and doctors had once again been significantly undermined by a doctor who had behaved less than honorably but also by a medical system that was patently seen to fail to regulate itself.

## **The Institute of Medicine Report, USA**

The Institute of Medicine (IOM) is an independent, nonprofit organization that works outside of government in the USA to provide unbiased and authoritative advice to decision makers and the public. Established in 1970, the IOM is the health arm of the National Academy of Sciences, which was chartered under President Abraham Lincoln in 1863. Nearly 150 years later, the National Academy of Sciences has expanded into what is collectively known as the National Academies, which comprises the National Academy of Sciences, the National Academy of Engineering, the National Research Council and the IOM.

In 1999, the IOM published the report, “To Err is Human; Building a Safer Health System,” (Kohn et al. 2000) which made the astonishing claim that between 44,000 and 98,000 people die in USA hospitals each year as a result of medical errors that could have been prevented. This report very quickly became a citation classic and was the focus of discussion in almost every major health care journal across the world. The content of the report shocked USA citizens and health care workers by the starkness of the message. In a single publication they had brought the issue of medical errors and patient safety to the forefront of discussions about health care. Ironically, the data that the IOM used to make these claims had been published in two papers in the *New England Journal of Medicine* almost a decade earlier (Brennan et al. 1991; Leape et al. 1991). In these two reports, the researchers reviewed 30,121 randomly selected records from 51 randomly selected acute care, nonpsychiatric hospitals in New York State in 1984. From these records, the researchers developed population estimates of injuries and computed rates according to age



**Table 1.1** Rates of adverse events and negligence among clinical specialty group

Specialty	Rate of adverse events		Rate of negligence	
	Percent	Population estimate	Percent	Population estimate
Orthopedics	4.1±0.6	6,746	22.4±4.7	1,514
Urology	4.9±0.8	4,819	19.4±6.5	933
Neurosurgery	9.9±2.1	2,987	35.6±8.6	1,063
Thoracic and cardiac surgery	10.8±2.4	3,588	23.0±9.3	826
Vascular surgery	16.1±3.0	3,187	18.0±8.1	575
Obstetrics	1.5±0.2	5,5013	38.3±7.0	1,920
Neonatology	0.6±0.1	1,713	25.8±6.9	442
General surgery	7.0±0.5	22,324	28.0±3.4	6,247
General medicine	3.6±0.3	37,135	30.9±4.4	11,475
Other	3.0±0.4	11,097	19.7±4.9	2,183

Plus-minus values are means ± SE. Values differ from the sums of those reported above because of rounding

and gender of the patients as well as error rates for the specialties of the physicians. The study was the largest and most comprehensive ever to investigate the incidents of adverse events that occurred to patients, while they were being cared for in hospital. In general, the medical profession and the general public had some awareness that hospitals were associated with an increased risk of bad things happening to patients while they were hospitalized. However, their estimates were on nothing like the same scale of adverse events reported by these two studies and discussed in detail by the IOM report. It is fair to say that the data shocked citizens and health-care workers in the USA and around the world.

Adverse events occurred in 3.7% (95% confidence interval; 3.2–4.2%) of hospital admissions, and of these 27.6% (95% confidence interval; 22.5–32.6%) were due to negligence (i.e., 1%).

It should be noted that error and negligence may be correlated but they are not the same. Medical negligence is defined as failure to meet the standard of practice of an average physician practicing in the specialty in question (Oxford English Dictionary 2004). Negligence occurs, not merely when there is error, but when the degree of error exceeds an accepted norm. The presence of error is a necessary but not sufficient condition for the determination of negligence. Sometimes the evidence of negligence appears clear cut as when a physician fails to evaluate a patient with rectal bleeding. Other cases are less obvious.

Using weighted averages they estimated that in the 2,671,863 patients discharged from New York hospitals in 1984 there were 98,609 adverse events of which 27,179 were due to negligence. Rates of adverse events rose with age with more adverse events due to negligence occurring in the elderly group. There were also marked differences between the rates of adverse events among the different physician groups and these are shown in Table 1.1.

Table 1.1 shows that the highest percentage of adverse events observed in the study was for vascular surgery (16.1%), followed by the thoracic and cardiac surgery (10.8%), neurosurgery (9.9%) and then general surgery (7%). The actual

**Table 1.2** Incidence of specific types of performance errors ( $n=697$ )

Type of error	No.	%
Inadequate preparation of patient before procedure	59	9
Technical error	559	76
Inadequate monitoring of patient after procedure	61	10
Use of inappropriate or outmoded form of therapy	24	3
Avoidable delay in treatment	41	7
Physician or other professional practicing outside area of expertise	13	2
Other	75	14

population estimates that these percentages represent was 22,324 adverse events in general surgery and an even higher incidence of 37,135 in general medicine. Despite the difference in observed incidence of adverse events between the different medical specialties the percentage of adverse events judged to have occurred as a result of negligence was fairly similar across the different specialties (range 35.6–18%). Obstetrics had the highest incidence of negligence (38.3%) followed by neurosurgery (35.6%). The incidence of adverse events as a result of negligence was 28% in general surgery (which represents 6,247 incidents) and 30.9% in general medicine (which represented 11,475 incidents). The data from this study are probably more accurate than the estimates from the only other large-scale study to have been conducted. The California Medical Association's Medical Insurance Feasibility Study (Mills 1978) was carried out in the 1970s to estimate the incidence of iatrogenic injury and substandard care. In this study, adverse events were estimated as occurring in 4.6% of the cases examined, with a negligence rate of 0.8% which was 20% lower than the Brennan et al. (1991) study. Of the 98,609 adverse events studied by Leape et al. (1991) 56,042 (56.8%) of them led to minimal disability with complete recovery in 1-month. In 13,521 (13.7%) incidents, the adverse events led to minimal disability with complete recovery in 6 months. However, 2,550 (2.6%) of them produced permanent total disability and in 13,451 (13.6%) led to death.

The researchers expressed surprise at the number of adverse events caused by negligence. In the New York study in 1984 they estimated that 27,179 injuries, including 6,895 deaths and 877 cases of permanent and total disability resulted from negligent care. Furthermore, the researchers (Brennan et al. 1991; Leape et al. 1991) point out that they did not measure all negligent acts, but only those that led to injury. Thus, their figures only reflected a consequence of negligence and not the actual true rate and as such probably represented a significant underestimation of the true rate of negligence in clinical care in the 30,121 randomly selected records that they studied.

The researchers also categorized the different types of errors as to their perceived cause. There were 397 events that were attributable to prevention errors, 265 events that were attributable to diagnostic errors; 153 that were due to drug treatment errors; and 68 that were due to system errors. However, the greatest single category was performance errors (697) and these are summarized in Table 1.2. More than three quarters of this type of error were due to technical performance. Nearly half of all adverse events (48%) resulted from operations and the location of the largest

**Table 1.3** Number of adverse events by medical specialty

Specialty	No. (%) of records reviewed	No. of patients with adverse events DETECTED	
		All (%) of records	Preventable (% of events)
General medicine	273 (27)	24 (8.8)	18 (75)
General surgery	290 (29)	41 (14.1)	17 (41)
Obstetrics	174 (17)	7 (4)	5 (71)
Orthopedics	277 (27)	38 (13.7)	12 (32)
Total	1014	110 (10.8)	52 (47)

percentage of adverse events was the operating room (41%) followed by the patient’s own hospital room (25%). The emergency room, intensive care units, labor and delivery rooms sites accounts for approximately 3% of the adverse events.

In a similar but less extensive study Vincent et al. (2001) assessed the incidence of adverse events in 1,014 hospital case notes randomly selected from two acute hospitals in London between July and September 1999 in one hospital and December 1999 and February 2000 in the second hospital. Table 1.3 shows the number and percentage of records reviewed by medical specialty which they were drawn from. The highest number of adverse events occurred in general surgery ( $n=41$ ), followed by orthopedics ( $n=38$ ) and then general medicine ( $n=24$ ). The greatest number of preventable adverse events occurred in medicine ( $n=18$ ), surgery ( $n=17$ ) and orthopedics ( $n=12$ ). They found that 10.8% of patients admitted to hospital experienced adverse events and an overall 11.7% rate of adverse events when multiple adverse events are included. About half of these events were judged preventable. A third of adverse events led to moderate or greater disability or death.

## The Rhetoric and the Reality of Follow-Up to the IOM Report

The Institute of Medicine (IOM) report had a profound effect on the health care community in the USA and across the world. It is not clear why the report made such an impact given it was based on data that was more than 10 years old (Brennan et al. 1991; Leape et al. 1991). Possibly it was the unambiguous and sheer number of adverse events and deaths as a consequence of health care that shocked and emboldened the healthcare community to do something about it. Within days of the Institute of Medicine’s report, the Clinton administration asked a federal task force to examine the recommendations made in it. The task force quickly agreed with the majority of the recommendations that were made in the report (Quality Interagency Coordination Task Force (QuIC) 2000). In spite of the initial flurry of activity that the report stimulated, activity and progress slowed once the media moved on to the next crisis. When the IOM published a follow-up report in March 2001 the release barely registered with the media and the public (Millenson 2002). Indeed, one of the architects of the IOM report and scientific lead of the study on which the report was

based concluded that movement toward systematic change to the healthcare system remains frustratingly slow (Leape and Berwick 2005). More than a decade after the release of the IOM report, efforts to reduce the harm caused by medical care systems have been few and fragmented.

The IOM report included recommendations to prevent medication errors, create accountability within the healthcare system through transparency and to establish a national focus by actually measuring the extent of the problem. The IOM report identified medication errors as a substantial source of preventable error in hospital. They recommended stronger oversight by the Food and Drug Administration to address safety issues connected with drug packaging and labeling, similar name drugs and post marketing surveillance of doctors and pharmacists (Kohn et al. 2000). Many medication errors are caused by the confusion of medicines with similar names and labels. Despite the fact that the FDA has had procedures in place since 1999 for assessing the potential of name confusion and monitoring of the market for instances of medication confusion, few existing names have been changed. Available evidence suggests that prescribing and administration problems associated with look-alike/sound-alike drugs has not been adequately dealt with by the FDA. Furthermore, the use of technology to minimize prescription or administration errors has been inadequately adopted by healthcare institutions and so patients continue to receive the wrong drug or the wrong dosage because of a doctor's poor handwriting. A federal law passed in 2008 offers bonus Medicare payments to physicians who use e-prescribing and physicians not using this facility will face reductions in Medicare payments starting in 2012. These relatively simple changes in physician behavior have failed to happen despite the existence of evidence that shows physician e-prescribing reduces medication errors by 81% (Bates et al. 1999) and the inclusion of the pharmacists with the team when doing rounds results in 66–78% reduction of preventable adverse drug reactions (Kucukarslan et al. 2003; Leape et al. 1999).

One of the primary recommendations made by the IOM report was for better data collection, particularly on adverse events to more reliably quantify the extent of the problem but also so that doctors and other members of the healthcare community could learn from mistakes. However, even this simple goal has met with only very variable success. For example, the National Quality Forum is a private membership group that works to set national priorities and goals for performance improvement. It publishes a list of voluntary consensus standards related to patient safety and includes a list of medical events that should never occur. The list of serious reportable events (sometimes known as the “never event” list) includes:

- Surgery performed on the wrong body part
- Surgery performed on the wrong patient
- Wrong surgical procedure performed on a patient
- Intra-operative or immediately post-operative death in an ASA Class I patient
- Patient death or serious disability associated with the use or function of a device in patient care, in which the device is used for functions other than as intended (Leape 2002; Wachter 2004)

Despite apparent widespread consensus on these types of efforts only 17 states had established a confidential reporting system by the time a federal framework was created in the Patient Safety and Quality Improvement Act of 2005 (which was not implemented until 2008 (Fassett 2006)). Progress has been almost entirely focused on voluntarily, confidential or aggregate reporting systems which although offer some benefits it hinders efforts to identify specific hazards, their antecedents (which are extremely valuable in helping to identify solutions) and the outcome of interventions put in place as a result of this analysis.

Knowing the incidence and potential origins of adverse events is a very valuable starting point in the development of a strategy to reduce these adverse events. One study which clearly showed the effectiveness of this approach targeted catheter-associated infections in the Michigan-affiliated Intensive Care Units. They found that they had an incidence of 7.7 bloodstream infections for every 1,000 days catheter used. In response to this problem they initiated a state-wide safety initiative called “Michigan Health and Hospital Association” Keystone: ICU and set a goal of reducing catheter-associated bloodstream infections. They instituted a short checklist of best practices related to catheter use and empowered nurses to ensure that doctors were following best practices. They tracked catheter associated bloodstream infection rates in 103 participating ICUs. Bloodstream infections across the participating ICUs dropped from 7.7 to 1.4/1,000 days catheter use during the study (Pronovost et al. 2006). The results also showed that 18 months after the study began the Michigan Health and Hospital Association reported that at least 50% of the participating ICUs had completely eradicated catheter-associated bloodstream infections.

These efforts represent a rare success story. The Agency for Health Research Quality (AHRQ) is the closest federal agency to the IOM’s vision of a center for patient safety and coordinated national resources on patient safety. It was established as a direct result of initiatives that stemmed from the IOM report and funds numerous research projects on quality and safety. It also publishes the National Healthcare Quality Report (NHQR), which is the vehicle for discussion and reporting on progress on patient safety which includes collecting evidence on the prevalence adverse events. The agencies patient safety indicators focus mainly on surgical errors and does not use data contained in forms such as patients case-notes. As an indicator of how little progress has been made towards accounting for preventable medical harm the 2009 NHQR (Agency for Healthcare Research and Quality (AHRQ) 2009) used data from the IOM work (Kohn et al. 2000) as the best estimate of the magnitude of medical errors!

One of the reasons for the limited impact of the IOM report may have to do with how well the AHRQ has been funded. The AHRQ was established around 1999 with a funding stream from Congress starting in the same year and it was tasked with dealing with issues pertaining to patient safety. Table 1.4 shows the amount of funding received by AHRQ in the years 1999, 2000, 2005, and 2010. It also shows the amount of funding received by different National Institutes of Health. AHRQ received \$28 million in 1999 which has increased to \$55 million by 2010. In contrast, the National Library of Medicine received \$35 million in 1999 but had

**Table 1.4** The amount of research dollars (in \$000,000) available from the Agency for Healthcare Research Quality for funding research from 1999 through to 2010 in comparison to the amount of funding available to a number of organizations within the National Institutes of Health (NIH)

	1999	2000	2005	2010
AHRQ	28	36	55	55
National Library of Medicine (NLM)	35	43	67	70
National Institute of Mental Health (NIMH)	860	973	1,400	1,500
National Cancer Institute (NCI)	3,000	3,300	4,800	5,100
National Centre for Complimentary and Alternative Medicine (NCCAM)	NA	68	122	129
Office of the Director (OD)	306	282	358	1,200
Total NIH Budget	15,000	17,800	28,500	31,000

increased to \$70 million in 2010. The amount of research monies received by AHRQ pales into insignificance when compared with the amount of research money received by the National Cancer Institute which received \$3 billion in 1999 and had increased to \$5.1 billion by 2010. Even the National Institute of Mental Health (which globally is notoriously underfunded), received significantly more funding than AHRQ. In fact, even the Office of the Director of National Institutes of Health (NIH) was significantly better funded than AHRQ. In 1999, the Office of the Director received 11 times the funding of AHRQ and by 2010 this had increased to 22 times more funding. Even the National Centre for Complimentary and Alternative Medicine (NCCAM) received more funding than AHRQ, despite not being established until the year after (NIH 2010). Lack of funding for this laudable enterprise is surprising, particularly given the furore that the IOM report created when it was published in 1999. However, with hindsight, perhaps the lack of funding is understandable, given that the majority of the first decade of the twenty-first century was under a Bush and Republican administration and all efforts took second place after the horrendous events of the World Trade Center in 2001. However, the problem of medical errors and adverse events is not going to go away on its own, in the USA or in any other country around the world. The issue will require a concerted and systematic approach to understand the problems and then develop evidence based solutions.

## Patients as Consumers

Unfortunately for medicine and surgery, all of these events occurred at a particular time in the historical development of public health care delivery when governments sought to empower patients. The clearest example of this occurred in the UK where the Conservative government led by Margaret Thatcher introduced the internal market to the National Health Service (NHS). This was outlined in the 1989 White Paper, Working for Patients (Health Committee 1989) which passed into law as the NHS Community Care Act 1990. The bill had been designed to increase

the responsiveness of the service to the consumer, to foster innovation, and to challenge the monopolistic influence of the hospitals on health-care in which community-based services were increasingly important. After the establishment of the internal market and the purchaser–provider split (purchasers’ were health authorities and some family doctors and providers were acute hospitals, organizations providing care for the mentally ill, people with learning disabilities, and ambulance services), purchasers were given budgets to buy health care from providers. One of the goals of this major NHS reorganization was to reduce waiting lists and to make health-care more efficient and responsive to patients. However, one of the unexpected consequences was precisely how much the general public would take the concept of “consumer” to heart. Around about this same time the consumer society was taking off and the general public had more and more access to better information and communication technologies such as satellite TV and the Internet. Under Prime Minister John Major, the Patient’s Charter reflected the idea of an “empowered client” as seen in the Citizen’s Charter, which was enacted in 1991. Although this charter did not have the force of law, it encouraged patients to complain and assert their health care rights (Harpwood 2001). It set out details of what patients could expect from the NHS, thereby establishing a standard by which doctors could be judged. In this respect, it significantly raised public awareness of rights and standards and encouraged the health care providers to focus on the gap between perceived and actual levels of care. The result was that the general public expected more from the health services and was better informed by the media about whether they were or were not getting better health care. Scandals such as the Bristol Royal Infirmary case could not have occurred at a worse time. Furthermore, although Bristol and many other scandals originated in the 1980s and early 1990s, responsibility for dealing with many of them came under the watch of a Labour government led by Tony Blair, who had publicly committed to the expansion of the NHS and ensuring better quality of patient care.

During this time also, there was a process of demystification of the medical profession. In the past, the public generally regarded physicians highly for three main reasons: (1) physicians’ control of knowledge, (2) the public’s perception that physicians worked in the patient’s best interest, and (3) physicians control of the decision-making process with regard to health care. The cause and effect of this relationship is not clear, but doctors have gradually lost their status as keepers and infallible source of medical knowledge (Haug 1973). In part this may be due to the fact that the average length of formal education among the general public has increased. This, combined with growing access to information, especially through Internet websites like WebMD ([www.webmd.com](http://www.webmd.com)), have decreased the knowledge gap between the patient and physician. What is clear is a greater level of information has empowered patients to question the decisions about their diagnosis and treatment (McKee and Healy 2002). In general, during the 1980s, people started to question these assumptions. There was a growing awareness standards of care and decisions about the single best treatment based on past effectiveness did not exist for many illnesses. Furthermore, as treatments became more technical it was difficult to know with certainty that one treatment option

was better than another. In addition, the general public became more aware of small variations and different treatments for similar conditions, based not on clinical determinants but on other factors, including physician preferences, in a particular region (Charles et al. 1999). Patients also began to realize since they had to live with the consequences of the doctor's medical decisions they should participate in the evaluation of the trade-offs.

There was also growing concern about whether the doctor really was acting in the patient's best interest. The medical profession to a greater extent throughout the twentieth century was almost entirely self-regulated. The profession chose to establish the General Medical Council (GMC) which is financed by the profession but accountable to Parliament as a form of self-regulation (Salter 1998). Originally, the GMC was charged with establishing a register of medical practitioners who were qualified to treat patients (Davies 2007). As part of discharging this duty, the GMC has the authority to discipline members whose actions are of poor quality and if necessary to revoke medical licenses. In theory, the GMC concerns itself with claims of serious professional misconduct, which according to the GMC means no more than serious misconduct judged according to the rules, written or non-written, governing the profession. One therefore might expect the GMC to review a wide range and large number of claims. However, it interprets this mandate narrowly. For example, from 1970 to 1984, no doctor was struck from the register for failing to attend a patient, but four were struck off for sexual misconduct with patients. Brazier (1992) summarizes the appearance of this situation particularly well when she asks, "has the GMC got its priorities right in punishing the adulterer with greater vigor than the uncaring doctor"? The answer, she explains, is that serious professional misconduct is interpreted to consist not of negligence or failure to attend to patients, but rather of actions that disgraced the profession. The general public and parliamentarians had access to this information and the ensuing discussions. The consequence for medicine was that doctors were not held in the same esteem that they had been when the NHS was first established. Overall, the relationship between a patient and their physician has changed considerably over the past few decades. The power of doctors associated with their professional autonomy and dominance has gradually weakened. The image of an idealized, infallible medical professional has undergone significant changes.

### **“Keyhole Surgery”: The Tipping Point**

The tipping point for the belief that surgery, and perhaps medicine needed to consider a radical change in the way that doctors were trained came with the widespread introduction of minimally invasive surgery (MIS) in the 1990s. “The Tipping Point” (Gladwell 2000) was a very influential book by a staff writer from the *New Yorker* magazine (Malcolm Gladwell). He argued that certain exceptional people can initiate change. These individuals can be characterized (individually or simultaneously) as “Connectors,” “Mavens,” and “Salesmen.” Connectors are individuals



who know lots of people and establish large social networks which means they have the capacity to spread information on ideas or products which they are particularly taken with. Mavericks are individuals who enjoy collecting information and then sharing that information with others. The third characteristic which Gladwell described was Salesmen, who are characterized by charm, enthusiasm, and likeability, i.e., the personality elements required to win others to a particular way of thinking. Gladwell suggested these characteristics Connectors (as the social glue), Mavericks (as databank or informationists) and salesmen (selling the idea, concept or product) brought to a project interacted to create a very powerful endorsement. The surgeons who were learning and practicing minimally invasive surgery at the outset were probably best typified by all of these three characteristics. They were (youngish) enthusiastic adopters of a new advanced technology which allowed for the performance of traditional surgery in a very novel way. Furthermore, the surgical Establishment was not particularly in favor of this new approach to performing surgery and so the proponents seemed like rebels in an otherwise very conservative profession. This small group of surgeons traveled the world giving lectures and seminars at international surgical meetings describing their experience of this new approach to performing surgery. However, Gladwell also suggests that for a message or idea to take hold, it has to be somehow memorable. The media supplied this last ingredient when they referred to minimally invasive surgery as “keyhole surgery”. This term captured the world’s attention and news of it spread like wildfire around the globe.

The surgery proponents of MIS did not actively discourage the use of the term “keyhole surgery” and this new approach to the performance of surgery quickly captured the public imagination. Surgeons were regular guests on news programs and documentaries promoting this approach to the performance of surgery for certain surgical procedures. The approach seemed to resonate with a consumer minded general public because it meant less scarring due to smaller incisions; incisions that were made were much easier to disguise (e.g., around the umbilicus); there was less pain associated with the procedure and patients returned to normal activities faster than they would after recovering from the same procedure performed with a traditional open surgical incision. It was also popular with cost-conscious hospital administrators because patients could have major surgical procedures performed minimally invasively with a much shorter hospital stay than they would with a traditional surgical incision. This approach to surgery was also very popular within industry as new types of surgical instruments, laparoscopes, cameras, monitors, etc. (some of them in the developmental stages) were required for the performance of surgery and the majority of the surgical instruments were disposable (and not inexpensive). This created a new large volume market from an existing customer base (surgeons) who traditionally, rarely replaced operating room instruments. In many respects, the development and evolution of MIS equipment manufacturers morphed into something resembling the pharmaceutical industry. However, nothing in life is that simple!

It soon became clear this new approach to performing surgery was associated with a higher complication rate than the traditional approach to performing the same

procedure by the open technique, particularly for establishing procedures such as laparoscopic cholecystectomy (Davidoff et al. 1992; Peters et al. 1991; The Southern Surgeons Club 1991). The minimally invasive approach to diagnostic procedures was not particularly new and had been used throughout the 1980s. Fiber-optic technology, closed-circuit television, and electrocoagulation equipment led to widespread introduction of laparoscopic techniques by gynecologists throughout the 1970s. General surgeons incorporated diagnostic laparoscopy into their practice during the 1980s for laparoscopic liver biopsy and cancer staging (Litynski 1999). The first laparoscopic cholecystectomies were in fact performed by European gynecologists in the late 1980s. Kurt Semm, a German gynecologist, performed the first laparoscopic appendectomy in 1983; the first documented laparoscopic cholecystectomy was performed by Eric Mühe in Germany in 1985, but Phillipe Mouret has been credited with performing the first laparoscopic cholecystectomy in Lyon, France, using video technique in 1987. This is important because it was the laparoscopic cholecystectomy operation which proved to be the precise tipping point for a revolutionary change in the way some surgical procedures were performed and as a consequence how surgeons were trained to perform them safely. Despite this period of exposure to the technique and the technology surgery was unprepared for the changes in training that were required for the safe adoption of this procedure.

By the time surgery had accepted there were “difficulties” in learning to perform surgery laparoscopically the approach had already achieved widespread acceptance by the general public, hospital administrators, and the health care establishment. What surprised many in the surgical and medical community was the degree of the difficulties associated with acquiring the surgical skills to practice this technique safely. After all, medical courses around the world attracted the brightest and best, and in general, surgery recruited the cream of them. However, surgery had made a fundamental and important miscalculation about the human factor difficulties associated with the practice of minimally invasive surgery. These will be discussed in detail in Chaps. 3 and 4. The trainee has to overcome considerable psychomotor and perceptual problems before even learning to perform MIS surgery safely and these problems are considerable and multiple. Firstly, the surgeon has to learn to coordinate 18 in. long surgical instruments that pass through trocars in the patients abdominal wall. Thus, they had lost important tactile and haptic information that they would normally receive through their fingers and the palms of their hands. They also had to perform surgery while looking at a pixelated image on a monitor. It may be a high quality image but it is still a pixelated image which required the brain to work harder than if it was processing information captured by the eye while viewing under natural seeing conditions. Images displayed on the monitor are captured from a single camera which means that many of the binocular cues that were associated with the judgment of depth of field are also lost. Lastly, perhaps the most significant obstacle to the learning and practice of safe laparoscopic surgery was the apparent counterintuitive movement of surgical instruments. For example, when the surgeon moved his or her hand (holding the handle of a surgical instrument) to the right inside the patient’s abdomen, the working end of the instrument moves to the left on the monitor. This causes a fundamental proprioceptive-visual conflict for the operator.

Their proprioceptive system tells the brain that the instrument is moving to the right while their visual system simultaneously informs the brain that the working end of the instrument is moving to left. Compounding these problems are the reduced degrees of freedom (in comparison to the hand and fingers) afforded by these new surgical instruments. These complexities make learning the psychomotor coordination necessary to perform laparoscopic surgery difficult and protracted. Furthermore, the reduced degrees of freedom afforded by the surgical instruments also meant that new techniques had to be developed for relatively straightforward surgical maneuvers such as suturing. In traditional open surgery, suturing is a precise but a very straightforward technique to learn. The widespread acceptance of MIS changed all that and it quickly became apparent that the traditional apprenticeship model (of learning on-the-job while practicing on patients) which had served surgery well for more than a century was not a viable training model, particularly for the early stages of the learning curve.

## “More” Training

In September 1992 the American NIH convened a consensus development conference on Gallstones and Laparoscopic Cholecystectomy (NIH 1993). They brought together surgeons, endoscopists, hepatologists, gastroenterologists, radiologists, and epidemiologists as well as other health care professionals and the public. They came to a number of conclusions one of which was that most patients who experienced symptoms of gallstones should be treated and that laparoscopic cholecystectomy provided a safe and effective alternative treatment to open cholecystectomy, for most patients. They also concluded that every effort should be made to ensure that surgeons performing laparoscopic cholecystectomy were properly trained and credentialed (and proctored for their first 15 procedures). As a result, there was a rapid expansion of training courses in laparoscopic surgical technique each expounding the ethos of the course organizer. It also led to the establishment of national and regional training centers around the world. However, a fundamental and detailed understanding of the specific human factor aspects of this surgical technique which made it difficult to learn eluded the majority of the surgical community except for leaders such as Prof Sir Alfred Cuschieri at Ninewells Hospital in Dundee and Dr Michael Patkin in Australia. The precise explanation as to why laparoscopic surgery is difficult to learn will be discussed in detail in Chaps. 3 and 4 but it is fair to say that it was not until the late 1990s that the extent and magnitude of these difficulties were documented and quantified (Berguer 1999; Crothers et al. 1999; Cuschieri 1995; Gallagher et al. 1998; Patkin and Isabel 1993). However, in the interim surgeons who wanted to learn to practice MIS needed more training.

These early courses were primarily led by industry. In the early 1990s, device manufacturers such as Ethicon Endo Surgery, Auto Suture (later to morph into US Surgical), Karl Storz, to name but a few, arranged courses for consultant surgeons who wanted to learn to perform surgery using the new laparoscopic technique.

These were very well run courses, staffed with well known national and international surgical faculty and they were also exceptionally well resourced. However, it is generally acknowledged that industry ran these courses in their efforts to increase sales of their product. Nevertheless, this should not detract from the quality of the courses offered by these organizations when at that time, academic surgery departments around the world were completely unprepared and had nothing to offer. These courses taught surgeons what they could and could not do with the devices they were going to use to perform the surgery. In reality, this meant that the surgeon was not going to have to work out what they were going to do with an instrument the first time they opened the packaging in the operating room, just prior to operating on a real patient. In this sense, industry provided the first human-factor safety training for devices in surgery. These same industrial organizations and a great many more continue to organize courses to this day. However, academic departments of surgery have become much more proactive in the establishment and running of a wide variety of courses as have professional organizations such as SAGES and EEES.

In general, these courses (industry and academic) lasted 1 or 2 days, usually over a weekend. Although the didactic and knowledge aspect of the course was well developed and reasonably standardized, there was a complete absence of standardization for the skills training component. Surgeons were familiarized with the imaging equipment, endoscopes, electrocautery, surgical instruments, and had some opportunities to acquire the psychomotor skills necessary for instrument handling. The training models used varied from course to course (and are discussed in more detail in Chap. 2) and included anything from an anesthetized pig in a fully equipped operating room through to inanimate bench top animal parts (e.g., chicken leg) or silicon models. Training simply consisted of exposure and, time permitting, some repeated practice. Performance metrics were subjective appraisal of task performance and possibly task completion time. There were no benchmarks for trainees to reach before applying their “skills” on a real patient and there was an implicit assumption that these types of course would be more than sufficient to familiarize and prepare the surgeon for this new type of surgical practice.

What is probably most surprising about this whole state of affairs is that the problems encountered by the surgeon in their efforts to acquire the skills for the safe practice of laparoscopic surgery were entirely predictable and understandable from human factors perspective (Gallagher and Smith 2003) and had been for at least half a century. What is also hard to believe is the fact that surgeons leading the vast majority of these training courses were blissfully unaware of this fact. However, this ignorance was not malicious and slowly but surely, detailed quantitative analysis of the human factor difficulties associated with the acquisition of the skills necessary for the practice of MIS started to appear in mainstream surgical journals. Furthermore, there was increasing awareness by the leaders in the MIS community that human factors, ergonomics, education, training, and validation were assuming an increasing importance in surgery. The endoscopic surgical movement grasped this reality first and started to populate their mainstream journals such as *Endoscopy* and *Surgical Endoscopy* with studies that validated the basic laparoscopic surgical

approach by comparing minimally invasive surgery to the traditional open approach for the same procedure. They also started to publish studies on the learning curve for a particular surgical procedure and the different approaches to training. Although at this time the understanding of “metrics” was crude e.g., using completion time as surrogate measure of performance, surgeons appeared to grasp the basic premise that subjective appraisal of trainee and intra-operative performance was inadequate for quality assurance purposes.

## Virtual Reality Simulation

For a period the surgical community offered training courses for a diverse and wide variety of laparoscopic procedures (even before their clinical efficacy over open procedures had been demonstrated). However, after the initial novelty of these offerings, which were very popular, widely covered in the media, well attended and well sponsored by industry, the actual costs of running courses became clear, i.e., they were relatively expensive to run in terms of faculty and course materials, such as instrumentation, consumables (e.g., suture material), and surgical training tasks. The most expensive training models were live animals that were fully anesthetized and operated on in a very high spec operating theater. Of course, training courses that offered operating experience on a live animal were the most popular with the surgical community, probably because they had the greatest face validity to the attending surgeons. As well as the expense of these courses, there was also the issue of animal rights which meant that running these types of courses was a sensitive issue. The surgical community argued that to train them to operate safely on patients they needed the highest fidelity training model possible. The irony is that although the porcine model offers some similar features to operating on a patient there is minimal direct anatomical equivalency.

Dr. Richard Martin Satava was a general surgeon in the U.S. Army who had been seconded to work for the Defence Advanced Research Projects Agency (DARPA) at the start of the 1990s. Thus, he was very well-informed about the difficulties which learning MIS posed for surgeons. He was also aware of the risk that taking a traditional Halstedian approach to training MIS skills would pose for the patient. In the military, whenever something is too dangerous, expensive, or distant in time or place or imagination, physically experience, there have been attempts to simulate the experience (Satava 1993). This is the approach the military and NASA had taken over the training of aviation and space flight skills. Some years after that, he wrote that simulation is a fundamental activity of virtually all species; it is the replacement of one dangerous activity by the enactment of a similar activity in a non-dangerous environment. It is the primary way in which children are taught to deal safely with the real world, and frequently includes the setting of play, theater, practice or sports. Surrogates (simulators) are used as replacements for real objects; they include dolls and puppets, props, and games among other substitutes (Satava 2008). During his first secondment to DARPA, Satava began to envision a simulation approach to

solving the problem of training minimally invasive surgical skills. Although anesthetists had been using mannequin simulation for team training for a number of years (Gaba and DeAnda 1988), no virtual reality simulator existed for training surgical skills. Asmund Laerdal, a successful plastic toys manufacturer had produced Resusci-Anne in the 1960s, which made possible the training of ABC (airway, breathing, circulation) for cardiopulmonary resuscitating (Safar et al. 1961). Cooper and Taqueti (2004) have reviewed the development of mannequin simulators and concluded that despite more than two decades of development the acceptance and market penetration of this type of simulation for clinical education and training was small. They also concluded that the acceptance of simulation and training would not occur until there was substantial validation evidence showing efficacy and cost-effectiveness for improving learning and producing better patient outcomes. Indeed, at the time of writing, this type of validation is was still not forthcoming for mannequin type simulations.

## Validation

Dr. Dave Gaba one of the pioneers of mannequin type simulations believes that there are many obstacles to obtaining definitive proof of the impact simulation on clinical care. He also pointed out that “no industry in which human lives depend on skilled performance or responsible operator has unequivocal proof of the benefits of simulation before embracing it” (Gaba 1992). One of the major obstacles that Gaba alluded to was the development of reliable and valid measurement instruments and methodologies necessary for the assessment of performance and behavior change as a result of simulation training. We believe that these were very perceptive insights by Dr. Gaba. Furthermore, we believe that the lack of widespread acceptance and penetration of simulation into education and training in medicine is primarily linked to the dearth of validation evidence. Furthermore, we also believe that there is a lack of validation evidence relating to simulation in medicine because it is fundamentally misunderstood. These issues will be addressed directly in Chaps. 10 and 11.

There is some clinical validation of the utility of simulation training in surgery (Seymour et al. 2002), however this evidence is still scant. There is a growing body of evidence in relation to the psychometric properties of simulation devices but there needs to be an expansion in the volume and quality of studies examining the value of simulation training for clinical performance. Like Gaba, we believe that these studies shouldn't really be necessary to convince the medical community that there is a better way to train clinical skills. However, all the indicators are that they are indeed required. The methodology which led up to and was used by the Seymour et al., study is probably the most robust clinical validation study that has been conducted on surgical simulation to date. The methodology used in the metric validation of the simulator used in this clinical trial was not new, and was derived from extensive knowledge of validation studies in the behavioral sciences. Likewise, the clinical validation methodology (i.e., proficiency-based progression training and

objective assessment of intra-operative performance) used, were also drawn from the behavioral sciences. These methods will be covered in detail in Chap. 5 (where we will detail how to identify, define, and measure performance); Chap. 6 integrates metrics into simulation training; Chap. 7 validates the metrics that have been developed and Chap. 8 harnesses the simulation for metric-based training to proficiency for improved intra-operative performance.

## **Understanding and Implementing an Alternative Training Strategy**

What we have tried to do in this book is draw together the knowledge and quantitative findings that help to explain why certain types of surgical procedures are difficult to learn. It is only from an extensive and thorough knowledge of these factors and how they relate to normal cognitive and information processing that methodology for more efficient and effective skill acquisition in surgery can be developed (Chaps. 3 and 4). This knowledge is necessary to help us understand precisely what we want to simulate and why we want to simulate it. Unfortunately, in the past simulations that have been developed for training medical skills have concentrated on what the simulator looks like. Most physicians mistakenly believe that a simulator that looks like real patient anatomy is a good simulator. As the reader will become aware in the chapters ahead, this is not a belief we hold to. Physicians tend to accept the validity and utility of this type of simulator purely on how it looks, in other words, how *pretty* it is. While this feature of a simulator is nice to have, there are a lot more important functional features which are higher up the priority list when making an effective and efficient simulation training device. One of these is the capacity to emulate the device and procedure to be learned and give detailed, reliable, and valid quantitative measures of performance, i.e., metrics. We will make the point time and time again; a simulator without these metric attributes is nothing more than a fancy video game, no matter how pretty it is. In Chaps. 5 and 6 we shall outline in detail how metrics are developed from first principles in a variety of contexts. We will give a number of examples to demonstrate that the principles are always the same and can be applied to any procedure to be simulated, learned, and assessed. Much of this methodology will be new to readers from a medical (and possibly engineering) background however; they have been well tried and tested for about half century in psychology. These are probably two of most important chapters in the book and this can be applied to any area of medicine and any medical procedure (if the principles are fully understood).

While it is all well and good knowing how to develop simulation and the metrics necessary for making it an efficient and effective training device there is still the “small” matter of convincing the medical community that the simulation and metrics that have been developed actually work. There are two steps to this process. The first is the validation of the psychometric properties of the metrics you have developed. This is an extremely important part of the validation of simulation,

particularly as the metric-based simulation and assessments are likely to be used for high-stakes purposes, such as determining training progression. Shoddy validation studies will not do! In Chap. 7, we will discuss the different types a psychometric validation required in the process of validating a simulator and its metrics. Again, this knowledge and expertise has been drawn from the psychology and educational testing sectors where this issue has been debated at length and international gold standard methodologies agreed standards (American Psychological Association, APA 1999). However, the nuances for their application in procedural medical simulation and clinical validation are somewhat novel. But, the rules of validation for these efforts are clear, if not completely understood as evidenced by some of the validation efforts in objective assessment of procedural skills and the development of metrics (see Chap. 7 for a full discussion of this issue). We will give explicit examples of what is acceptable and what is not acceptable, particularly regarding the assessment of inter-rater reliability.

Armed with validated metrics it is then necessary to demonstrate that metric based simulation training improves clinical performance in comparison to traditional training. In Chaps. 8, 9, and 10 we will describe how a complete education (e.g., e-learning) and training package should be put together, how it should be implemented and evaluated. We will also discuss lessons learned from training programs that have already been developed and implemented. The novelty of this approach to training means that mistakes (e.g., inefficiencies) are inevitable. However, mistakes are very valuable learning opportunities (for those who wish to learn). That is precisely the point that is made in Chap. 11 when we discuss the issue of feedback and deliberate practice in determining how education and training should be optimally configured.

## The Paradigm Shifts

The combined impact of all of these events on surgery was profound and disruptive. Just as Kuhn (1962) had predicted, it created a crisis within surgery in particular and medicine in general. Medicine and surgery have been subjected to high profile medical error and negligence cases in the past. However, the cases that we have outlined here had an enormous impact on the medical community, but they also impacted on the general public's perception of doctors and how they treated their patients. Furthermore, these cases occupied the headlines in the popular press for years, with the graphic lurid details of each case being discussed in detail in front of a shell-shocked general public. There is little doubt that in the aftermath of these cases the general public's perception and possibly confidence in their doctors had been significantly shaken. Furthermore, these cases also brought about a fundamental and radical reconfiguration of how doctors were trained. There was a move away from the perception that doctors were competent once they "knew how" to do something. In the new configuration of training doctors had to "demonstrate" that they knew how (see Chap. 8).



Compounding these considerable problems was a demand by health care providers for greater productivity in medical care which meant targets for operations, targets for waiting lists etc. Although the development of MIS may have helped on this front it was a tipping point for the change in how surgeons were trained. The widespread introduction of MIS into clinical practice meant two things for surgeons in training. The first was that the introduction of MIS eliminated many potential training opportunities for the junior surgeon. For example, hernia repair and open cholecystectomy were relatively straightforward surgical procedures that provided frequent opportunities for the vast majority of surgeons to acquire their basic technical surgical skills in a relatively low-risk environment. The second was that not only were these training opportunities removed from the basic surgical training opportunities, but they had now become more advanced procedures, which required advanced training. Even when they did get an opportunity to perform them, this was probably as a result of a patient being too ill to perform the procedure laparoscopically or an MIS procedure that was converted to an open procedure; neither of these scenarios could be described as straightforward! Making this situation even worse was the reduced work hours that surgeons had available to them during which to learn their craft. Moreover, the rate of change through the introduction of new approaches and new technologies to the performance of surgery had increased exponentially. Although this was not just a problem for surgery, but all of medicine, it impacted worst on surgery and other procedural specialties because the acquisition of their skills could only occur (traditionally that is) in a relatively specialized environment, i.e., the operating room. To make matters worse, surgeons were also being required to achieve certain standards or levels of competence. Furthermore, although the assessment of the skills that had traditionally been left to the prerogative of the supervising consultant surgeon, new standardized assessment methodologies had been introduced and were a mandatory part of training and career progression.

There can be little doubt that all of these factors combined to create a sense of crisis among the surgical establishment. There can also be little doubt that surgery was confronted with an unanticipated training crisis of global proportions. Kuhn (1962) also predicted that during transitions and periods of crisis a wide range of potential solutions are examined and sometimes existing solutions are re-examined. This is precisely what happened with simulation. As we have described earlier, anesthesiologists had been using simulation since the 1960s in their educational and training curriculum, but this had not registered with the surgical community as a technology they were particularly interested in. However, the development of one of the first surgical simulators by Satava (1993) started a process which would move surgical simulation from a proof of concept, through to clinical validation (Seymour et al. 2002) to widespread acceptance as a primary training modality for the new training paradigm in surgery (Pellegrini et al. 2006). We will argue here that simulation based training was accepted and implemented before it was fully understood by the surgical and medical establishment. Although widely believed to offer the opportunity for repeated practice that had been lost in the operating room, simulation in fact provides the opportunity for deliberate practice. Deliberate practice differs from repeated practice in the use of metric-based formative feedback to hone

the skills of the trainee. This means that the optimal development and application of metrics lies at the very core of effective and efficient simulation training. One of the goals of this book is to explain how these factors work together and should be optimally configured for an efficient and effective approach to training.

Lastly, we will also argue that although simulation based training technology affords the opportunity for more efficient and effective training in disciplines such as surgery, it also offers the opportunity to quality assure the skill levels of graduating trainees. Traditionally, surgeons have acquired their skills in an apprenticeship model, where they practiced on patients. This meant that individual surgeons had a considerably varied experience, i.e., it depended on what hospital they were working in during training, on what consultant they worked with, and what patients they got to operate on, which meant they graduated with variable skills levels. In the past, *in all probability* their skills would have been trained to at least a safe level of operating simply by the sheer volume of operating they had performed during their training. In a twenty-first-century health care this guarantee of case volume no longer exists during training. Furthermore, the number of cases performed and the amount of time in training are very poor predictors of the skill level of the surgeon (which we will discuss in detail in Chap. 8). The approach that has been taken in most developed medicine training programs around the world is to require trainees to reach a level of competency. Unfortunately, this is a basic level of competency about which there is widespread unease among very experienced practicing surgeons. To be frank, we share this unease, not least because of the lack of transparency of these levels of competency, their lack of unambiguous operational definitions (see Chap. 5), and the impact that this has on the reliability of the assessment process (see Chap. 7). What we have proposed here is that trainees should train until they reach a performance criterion level, i.e., a level of proficiency. Furthermore, this level of proficiency should be quantitatively defined using validated metrics implemented in simulation technology and based on the *in vivo* performance of experienced and practicing surgeons. This strategy achieves two things: First, it establishes an unambiguous, objective, transparent, and fair training goal for a trainee, which is based on the performance of practicing surgeons in the real world and not some abstract concept of “a just passing performance” (i.e., competence). Second, it ensures a considerably less variable level of skills of graduates. Both of these factors would go some way to reassuring the surgical establishment that this new approach to training surgical skills stands a better than average chance of producing surgeons who can become as good if not better as they are.

## Conclusions

Whether by design or by accident, Halsted developed a training program which has served medicine well for a century. However, considerably more is known today about the cluster of human factors which are essential for the education and training of advanced skills such as surgery. The process of education and skill acquisition is

not some unknown black box. Surgery has a unique opportunity to develop a training program that will serve medicine well for many years. However, this program should be built on an explicit and detailed understanding of human sensation, perception, cognition, kinesthetics, psychomotor learning, and performance. Considerably more is known about the performance characteristics and parameters of these human factors and on how they impinge on human learning and the practice of skilled performance. Equipped with this knowledge, surgery will be better able to build simulations which are optimally configured for the training and assessment of advanced procedural skills in surgery. This approach is important because other procedural disciplines in medicine are confronting the same problems as surgery. However, surgery has reached this point first, and is duty bound to ask and address the important questions that will shape the future of procedural training in medicine. This approach will also inform surgery of the deficits in simulations that currently exist for training surgical skills and ensure that these are not repeated in the next generation of simulations. We also believe that this revolution which started in surgery, probably one of the most conservative disciplines within medicine, will change all of medicine.

## References

- Adnett N, Hardy S. Reviewing the working time directive: rationale, implementation and case law. *Ind Relat J*. 2001;32(2):114-125.
- Agency for Healthcare Research and Quality (AHRQ). National healthcare quality report. *Journal*. From: <http://www.ahrq.gov/qual/qdr09.htm>. 2009 (accessed 19th September 2010).
- Åkerstedt T, Gillberg M. A dose-response study of sleep loss and spontaneous sleep termination. *Psychophysiology*. 1986;23(3):293-297.
- APA. NCME (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education). *Standards for educational and psychological tests*; Washington DC: AERA Publications Sales; 1999.
- Asch DA, Parker RM. The Libby Zion case. One step forward or two steps backward? *N Engl J Med*. 1988;318(12):771.
- Bates DW, Teich JM, Lee J, et al. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc*. 1999;6(4):313-321.
- Berguer R. Surgery and ergonomics. *Arch Surg*. 1999;134(9):1011-1016.
- Brazier M. *Medicine, Patients and the Law*. Harmondsworth: Penguin-Printed Resource; 1992.
- Brennan TA, Hebert LE, Laird NM, et al. Hospital characteristics associated with adverse events and substandard care. *JAMA*. 1991;265(24):3265-3269.
- Bristol Royal Infirmary Inquiry. *Learning from Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary 1984-1995*. London: Stationery Office; 2001.
- Burton B. Court ruling ends Patel inquiry. *BMJ*. 2005;331(7516):536.
- Cameron JL. William Stewart Halsted: our surgical heritage. *Ann Surg*. 1997;225(5):445-458.
- Charles C, Gafni A, Whelan T. Decision-making in the physician-patient encounter: revisiting the shared treatment decision-making model. *Soc Sci Med*. 1999;49(5):651-661.
- Cooper JB, Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *Br Med J*. 2004;13(suppl 1):i11-i18.
- Crothers IR, Gallagher AG, McClure N, James DT, McGuigan J. Experienced laparoscopic surgeons are automated to the "fulcrum effect": an ergonomic demonstration. *Endoscopy*. 1999;31(5):365-369.

- Cuschieri A. Whither minimal access surgery: tribulations and expectations. *Am J Surg*. 1995; 169(1):9-19.
- Daan S, Beersma DG, Borbely AA. Timing of human sleep: recovery process gated by a circadian pacemaker. *Am J Physiol Regul Integr Comp Physiol*. 1984;246(2):R161-R183.
- Davidoff AM, Pappas TN, Murray EA, et al. Mechanisms of major biliary injury during laparoscopic cholecystectomy. *Ann Surg*. 1992;215(3):196-202.
- Davies M. *Medical Self-Regulation: Crisis and Change*. Hampshire: Ashgate Publishing Company; 2007.
- Dembe AE, Erickson JB, Delbos RG, Banks SM. The impact of overtime and long work hours on occupational injuries and illnesses: new evidence from the United States. *Br Med J*. 2005;62(9):588-597.
- Elman MJ, Sugar J, Fiscella R, et al. The effect of propranolol versus placebo on resident surgical performance. *Trans Am Ophthalmol Soc*. 1998;96:283-294.
- Fassett WE. Patient safety and quality improvement act of 2005. *The Annals of pharmacotherapy* 2006;40(5):917-924.
- Ferguson RL, Jobe K. A quiet hand for microneurosurgery: twiddle your thumb. *J Neurosurg*. 2004;101(3):541-544.
- Forster P. *Queensland Health Systems Review Final Report*. Queensland: Queensland Department of Health; 2005 [http://www.health.qld.gov.au/health\\_sys\\_review/final/qhsr\\_final\\_report.pdf](http://www.health.qld.gov.au/health_sys_review/final/qhsr_final_report.pdf) (accessed 5th July 2011).
- Friedman RC, Bigger JT, Kornfeld DS. The intern and sleep loss. *N Engl J Med*. 1971;285(4): 201-203.
- Gaba DM. Dynamic decision-making in anesthesiology: cognitive models and training approaches. In: Evans DA, Patel VL, eds. *Advanced Models of Cognition for Medical Training and Practice*. Berlin: Springer; 1992:123-147.
- Gaba DM, DeAnda A. A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology*. 1988;69(3):387-394.
- Gaba DM, Howard SK. Patient safety: fatigue among clinicians and the safety of patients. *N Engl J Med*. 2002;347(16):1249-1255.
- Gallagher AG, Smith CD. From the operating room of the present to the operating room of the future. Human-factors lessons learned from the minimally invasive surgery revolution. *Semin Laparosc Surg*. 2003;10(3):127-139.
- Gallagher AG, McClure N, McGuigan J, Ritchie K, Sheehy NP. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy*. 1998;30(7):617-620.
- Gladwell M. *The Tipping Point: How Little Things Can Make a Big Difference*. Boston: Little, Brown & Company; 2000.
- Grantcharov TP, Bardram L, Peter FJ, Rosenberg J. Laparoscopic performance after one night on call in a surgical department: prospective study. *BMJ*. 2001;323(7323):1222-1223.
- Grillo HC, Edward D. Churchill and the "rectangular" surgical residency. *Surgery*. 2004;136(5): 947-952.
- Hallenbeck WH. *Radiation Protection*. Boca Raton: CRC Press; 1994.
- Harding Clark M. The Lourdes Hospital Inquiry—An Inquiry into peripartum hysterectomy at Our Lady of Lourdes Hospital, Drogheda. *Report of Judge Maureen Harding Clark SC. The Stationery Office, Dublin*; 2006.
- Harpwood V. *Negligence in Healthcare: Clinical Claims and Risk in Context*. London: Informa; 2001.
- Harwell RC, Ferguson RL. Physiologic tremor and microsurgery. *Microsurgery*. 1983;4(3):187-192.
- Haug M. Deprofessionalization: an alternative hypothesis for the future. *Sociol Rev Monogr*. 1973;20:195-211.
- Health Committee. Working for patients. Vol. 159 cc556-9W. 1989. From: [http://hansard.millbanksystems.com/written\\_answers/1989/nov/07/working-for-patients](http://hansard.millbanksystems.com/written_answers/1989/nov/07/working-for-patients). Retrieved October 21, 2010.
- Hutter MM, Kellogg KC, Ferguson CM, Abbott WM, Warshaw AL. The impact of the 80-hour resident workweek on surgical residents and attending surgeons. *Ann Surg*. 2006;243(6):864-871.

- Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington DC: National Academy Press; 2000:196-197.
- Kucukarslan SN, Peters M, Mlynarek M, Nafziger DA. Pharmacists on rounding teams reduce preventable adverse drug events in hospital general medicine units. *Arch Intern Med*. 2003;163(17):2014-2018.
- Kuhn TS. *The Structure of Scientific Revolutions. International Encyclopaedia of United Science*. Chicago: University of Chicago Press; 1962.
- Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med*. 2004;351(18):1838-1848.
- Leape LL. Reporting of adverse events. *N Engl J Med*. 2002;347(20):1633-1638.
- Leape LL, Berwick DM. Five years after to err is human: what have we learned? *JAMA*. 2005;293(19):2384-2390.
- Leape LL, Brennan TA, Laird N, et al. The nature of adverse events in hospitalized patients. Results of the Harvard medical practice study II. *N Engl J Med*. 1991;324(6):377-384.
- Leape LL, Cullen DJ, Clapp MD, et al. Pharmacist participation on physician rounds and adverse drug events in the intensive care unit. *JAMA*. 1999;282(3):267-270.
- Litynski GS. Endoscopic surgery: the history, the pioneers. *World J Surg*. 1999;23(8):745-753.
- Lowry J, Cripps J. Results of the online EWTD trainee survey. *Bull R Coll Surg Engl*. 2005;87(3):86-87.
- Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: implications for the field of brain and cognition. *Brain Cogn*. 2007;65(3):209-237.
- McKee M, Healy J. *Hospitals in a Changing Europe*. Buckingham: Open University Press; 2002.
- Meumann E. Haus- und Schularbeit: Experimente an Kindern der Volksschule. [Home and school-work: experiments on children in school]. *Die Deutsche Schule*. 1904;8:278-303; 337-359; 416-431.
- Millenson ML. Pushing the profession: how the news media turned patient safety into a priority. *Qual Saf Health Care*. 2002;11(1):57-63.
- Mills DH. Medical insurance feasibility study: a technical summary. *West J Med*. 1978;128(4):360-365.
- Moore-Ede M. *Twenty-Four-Hour Society: Understanding Human Limits in a World That Never Stops*. Reading: Addison Wesley Publishing Company; 1993.
- NIH. Consensus development panel. Gallstones and laparoscopic cholecystectomy. (National Institutes of Health Consensus Development Panel on Gallstone and Laparoscopic Cholecystectomy). *JAMA*. 1993;269(8):1018-1024.
- NIH. The NIH almanac – appropriations. May 14, 2010. From: <http://www.nih.gov/about/almanac/appropriations/index.htm>. Retrieved September 19, 2010.
- Oxford English Dictionary (ed) (Vol. 24) Oxford: Oxford University Press; 2004.
- Pais A. *Subtle Is the Lord: The Science and the Life of Albert Einstein*. Oxford: Oxford University Press; 2005.
- Patkin M. Ergonomics applied to the practice of microsurgery. *Aust N Z J Surg*. 1977;47(3):320-329.
- Patkin M, Isabel L. *Ergonomics and Laparoscopic General Surgery. Laparoscopic Abdominal Surgery*. New York: McGraw-Hill; 1993.
- Pellegrini CA. Surgical education in the United States: navigating the white waters. *Ann Surg*. 2006;244(3):335-342.
- Pellegrini CA, Sachdeva AK, Johnson KA. Accreditation of education institutes by the American College of Surgeons: a new program following an old tradition. *Bull Am Coll Surg*. 2006;91(3):8-12.
- Peters JH, Ellison EC, Innes JT, et al. Safety and efficacy of laparoscopic cholecystectomy. A prospective analysis of 100 initial patients. *Ann Surg*. 1991;213(1):3-12.
- Philibert I, Friedmann P, Williams WT. New requirements for resident duty hours. *JAMA*. 2002;288(9):1112-1114.

- Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med*. 2006;355(26):2725-2732
- Quality Interagency Coordination Task Force (QuIC). Doing what counts for patient safety: federal actions to reduce medical errors and their impact. *Journal*. From: <http://www.quic.gov/report/mederr2.htm>. 2000 (accessed 19th September 2010).
- Reason J. *Human Error*. Cambridge: Cambridge University Press; 1990.
- Safar P, Brown TC, Holtey WJ, Wilder RJ. Ventilation and circulation with closed-chest cardiac massage in man. *JAMA*. 1961;176(7):574-576.
- Salter B. *The Politics of Change in the Health Service*. Basingstoke: Macmillan; 1998.
- Samkoff JS, Jacques CH. A review of studies concerning effects of sleep deprivation and fatigue on residents' performance. *Acad Med*. 1991;66(11):687-693.
- Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc*. 1993;7(3):203-205.
- Satava RM. Historical review of surgical simulation – a personal perspective. *World J Surg*. 2008;32(2):141-148.
- Schuster M, Rhodes S. The impact of overtime work on industrial accident rates. *Ind Relat J Economy Soc*. 1985;24(2):234-246.
- Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236(4):458-463; discussion 463-454.
- Sheldon GF. Surgical workforce since the 1975 study of surgical services in the United States: an update. *Ann Surg*. 2007;246(4):541-545.
- Slack PS, Ma X. Time dependency assessment of muscular fatigue index and hand tremor under operating conditions. *IEEE*. 2007:4822-4825.
- Smith R. All changed, changed utterly. *Br Med J*. 1998;316(7149):1917-1918.
- Smith C, Wise MN. *Energy and Empire: A Biographical Study of Lord Kelvin*. Cambridge: Cambridge University Press; 1989.
- Spritz N. Oversight of physicians' conduct by state licensing agencies. *Ann Intern Med*. 1991;115(3):219-222.
- The Southern Surgeons Club. A prospective analysis of 1518 laparoscopic cholecystectomies. *N Engl J Med*. 1991;324(16):1073-1078.
- Van Der Weyden MB. The Bundaberg Hospital scandal: the need for reform in Queensland and beyond. *Med J Aust*. 2005;183(6):284-285.
- Villaneuva T. European working time directive faces challenges. *Can Med Assoc J*. 2010;182(1):E39-E40.
- Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ*. 2001;322(7285):517-519.
- Wachter RM. The end of the beginning: patient safety five years after 'to err is human'. *Journal*, W4-534-545. From: <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.534v1>. 2004.
- Wallack MK, Chao L. Resident work hours: the evolution of a revolution. *Arch Surg*. 2001;136(12):1426-1431.
- Walshe K, Offen N. A very public failure: lessons for quality improvement in healthcare organisations from the Bristol Royal Infirmary. *Qual Health Care*. 2001;10(4):250-256.
- Weinger MB, Ancoli-Israel S. Sleep deprivation and clinical performance. *JAMA*. 2002;287(8):955-957.
- Whang EE, Mello MM, Ashley SW, Zinner MJ. Implementing resident work hour limitations: lessons from the New York State experience. *Ann Surg*. 2003;237(4):449-455.
- Wojtczak-Jaroszowa J, Jarosz D. Time-related distribution of occupational accidents\* 1. *J Safety Res*. 1987;18(1):33-41.
- Yerkes RM, Dodson JD. The relationship of stimulus to rapidity of habit formation. *J Comp Neurol Psychol*. 1908;18:459-482.
- Zajonc RB. Social facilitation. *Science*. 1965;149(3681):269-274.