

Chapter 1

The History of Clinical Trials



Janice Hu, Justin Barr, and Georgia M. Beasley

1.1 Introduction

A clinical trial is a purposeful comparison of medical interventions, including placebos, against one another to determine the safest, most efficacious means of treating pathology. The history of clinical research in surgery sheds light on both the successes and the challenges that academic surgeons faced when developing therapies for their patients.

1.2 Early History

Clinical trials had little role in the ancient world where accepted disease theories rendered them all but irrelevant. In many older cultures, disease and healing were perceived to stem from supernatural and divine forces. In Greece during the fifth century BCE, patients sought healing through “incubation,” or sleep, in temples of the healing god Asclepius. It was around this time that a new form of medicine arose, marking a major innovation in the treatment of disease. Unlike supernatural theories, Hippocrates’ method involved seeking the cause of illness in natural

J. Hu
Duke University School of Medicine, Durham, NC, USA

J. Barr
Duke Department of Surgery, Durham, NC, USA

G. M. Beasley (✉)
Duke University School of Medicine, Durham, NC, USA

Duke Department of Surgery, Durham, NC, USA
e-mail: georgia.beasley@duke.edu

factors involving the composition of the body's humors. An oeuvre of texts known as the *Hippocratic Corpus*, written by numerous authors over many decades until the first half of the fourth century BCE, established that physicians could learn through observations and actions. Yet the ancient Greeks did not perform clinical trials to test their hypotheses. Moreover, the highly individualized understanding of disease made broadly applicable treatments rare, vitiating the value of clinical trials. The Greek had “freed himself of religion to become the prisoner of philosophy” [1]. This dogma largely continued through the Roman world [2].

In 1025 CE, the Persian physician Avicenna wrote the widely used medical treatise *The Canon of Medicine* in which he laid down a precise guide for empirical investigation of the effectiveness of medical drugs and substances [3]. He recommended studying two cases of contrary types, along with the timing and reproducibility of drug effects so that consequence and accident are not confused. Moreover, he advocated for experimentation on the human body, since testing a drug on a lion or a horse might not prove anything about its effect on man. The pharmacology discussed in Avicenna's treatise was used extensively in medical schools across Europe as late as 1650 [4]. Although Avicenna advocated for the empirical study of drugs, his *Canon* did not lead to the widespread engagement of experiments and empiricism. Instead, the Medieval Era (800–1400 CE) was characterized by textual dependence and interpretation that prized the authority of the ancients over experimental evidence [5]. Moreover, while extant sources such as the *Hippocratic Corpus* and the *Canon* defined elite, academic-based medicine, the vast majority of medical care was delivered by untrained, unlicensed, and irregular practitioners, most of whom were illiterate. This practice went largely unrecorded and likely relied on a combination of superstition, tradition, and empiricism.

1.2.1 Early Modern Era (1500–1800)

With the dawn of the early modern era in the sixteenth century, there was a general intellectual shift away from dogmatic textual dependence and toward empirical investigation. This was evident in multiple arenas including heliocentric theories of astronomy put forth by Nicolaus Copernicus, anatomical observations made by Andreas Vesalius, and navigational feats like those by Christopher Columbus. They also appeared in medicine.

One of the first clinical trials was accidentally conducted in 1537 by the French surgeon Ambroise Paré when he ran out of the boiling oil that was conventionally used to treat bullet wounds and resorted to giving some soldiers a balm made from egg yolks, rose oil, and turpentine [6]. He awoke the following morning to find that patients who received the new treatment were resting well with little discomfort and swelling, whereas those who were cauterized with oil were “feverish with much

Fig. 1.1 Ambroise Paré et l'examen d'un malade [Ambroise Paré examining a patient] by James Bertrand (1823–1887), from the Charles de Bruyères Museum collection in Remiremont. (Source: Ji-Elle, license CC:BY-SA) [8]



pain and swelling about their wounds.” Reflecting on this experience, he noted “I resolved with myself never more to burn thus cruelly poor men wounded with gunshot” [7]. This observation, widely published, changed clinical practice as military surgeons across Europe began to eschew boiling oil in favor of less painful remedies (Fig. 1.1).

Systematic tests of disease management tackled the pre-fifteenth century Galenic tradition of wound management, characterized by gradual “wet healing” that involved forcing wounds open and applying emollients. This conventional method often led to poor outcomes. From 1580 to 1583, Spanish surgeon Bartolomé Hidalgo de Agüero challenged this notion by examining hospital records, finding that his own method of “dry healing”—cleaning the wound with white wine, removing damaged tissue, bringing the edges together, applying drying compounds, and covering the wound with a bandage—led to a far lower mortality rate compared to the Galenic technique [9].

The trend of empiricism continued to grow as physicians set forth hypotheses and began testing them through observation. Paré and Agüero belonged to a group of sixteenth century practitioners who were willing to trust their observations and personal experience over ancient traditions and dogma. Yet two centuries would pass before the launch of the first rigorous prospective trial.

Scottish surgeon James Lind randomized six pairs of sailors to different treatments for scurvy in 1747, finding that citrus fruits were the most effective therapy [10]. Despite the soundness of his methods and the irrefragability of his results, his conclusion had little impact on medical opinion in Britain, exposing an ongoing theme through this history: the challenge of even the best clinical trial actually changing medical practice. It ultimately requires many more decades, with thousands of additional deaths, for professional opinion to adopt lemons as a scurvy prophylactic.

1.3 The Emerging Importance of Statistics

Comparative retrospective analyses played an important role in building toward controlled trials in medicine and surgery. Statistics, or the practice of collecting and analyzing large amount of numerical data, emerged as an important tool in treatment evaluation. By the eighteenth century, several case series propelled arguments about the utility, methods, and timing of limb amputations [11, 12]. Lithotomists published numerical evidence on bladder stone removal, debating the merits of lithotripsy compared to lithotomy and examining mortality among age subgroups [13–15]. In the 1820s, Pierre-Charles-Alexandre Louis used his “numerical method” on aggregated clinical data to cast doubt on the practice of bloodletting [16, 17]. Furthermore, statistics featured prominently in debates surrounding perioperative innovations such as anesthesia and Lister’s “antiseptic method” of carbolic acid for surgical wounds beginning in 1867 [18, 19]. This portended the clear role and need for stronger evidence to evaluate theories of disease management. It also demonstrated the shift from highly individualized disease states as understood in ancient and medieval medicine to a more ontological notion of sickness where a single intervention had the potential to apply to all patients suffering from the same pathology. This critical theoretical transition made clinical trials relevant. Moreover, as anesthesia and antiseptics allowed surgeons to delve further into internal organs and conduct more elective procedures, there arose a clear need to provide proof of safety and benefit.

1.4 Prospective Clinical Trials Begin

In the nineteenth century, surgeons joined in performing prospective trials by first using nonrandom methods of treatment assignment such as alternate allocation. In perhaps the earliest example of this, an 1816 medical dissertation describes how military surgeons performed a controlled trial on 366 soldiers in the Peninsular War to assess the effects of bloodletting for fever. Although there are uncertainties surrounding the authenticity of this report [20], it nonetheless illustrates the emerging desire among surgeons to control for factors other than the treatment of interest:

It had been so arranged, that this number was admitted, alternately, in such a manner that each of us had one third of the whole. The sick were indiscriminately received, and were attended as nearly as possible with the same care and accommodated with the same comforts. One third of the whole were soldiers of the 61st Regiment, the remainder of my own (the 42nd) Regiment. Neither Mr. Anderson nor I ever once employed the lancet. He lost two, I four cases; whilst out of the other third [*treated with bloodletting by the third surgeon*] thirty five patients died [21].

The last decades of the nineteenth century witnessed the publication of other prospective surgical studies using alternate allocation. These included catheterization for urethrotomies, capsulotomy following removal of cataracts, and pediatric hernia management [22–24]. The goals of these researchers were twofold: (1) to

make firmer distinctions among different interventions and (2) to demonstrate impartiality. In 1893, W. T. Bull explained how alternate allocation reduced bias when comparing a spring truss to a skein wool truss for the treatment of pediatric hernias:

In children under the age of 1 year the worsted or so-called ‘hank truss’ has been extensively tried. This truss has been very highly praised by some, and as strongly condemned by others. During the past year an attempt has been made to give it an impartial trial, and alternate cases up to the age of 1 year were treated by the ‘hank’ and the light spring truss. The results in 240 cases carefully followed up led us to discard the hank truss as a routine method of treatment, although there are still a few cases—for example, very young and ill-nourished infants where it fills a useful but temporary place [24].

Although prospective studies comparing groups of patients emerged into the professional surgery landscape, individual patient outcomes remained powerful guides for surgical management. After all, Bull still advocates for the use of the hank truss in “a few cases” based on select patient characteristics. In fact, despite the emergence of prospective controlled studies, case series feature prominently in the body of published surgical evidence well into the late nineteenth and early twentieth centuries [25]. Indeed, these retrospective studies helped surgeons select operative techniques amidst the variability of their patient populations. They also led, however, to protracted debates about competing techniques and to the propagation of now-defunct operations including treatments for ptosis, constipation, and autonomic nerve dysfunction [26]. Surgeons tended to publish case series that promoted their own opinions, leading to unresolved debates in areas such as radical mastectomy and prostate surgery [27, 28]. Biased results continued to highlight the need for more carefully designed investigations.

1.5 The First Randomized Clinical Trials

In order for the randomized trial to become the gold standard in guiding medical practice, its various constituents, including controls, blinding, quantification, and randomization, needed to undergo their own evolution [25]. The randomization component in particular is important because it eliminates selection bias, balances treatment groups with respect to confounders, and forms the basis of statistical tests which assume equality of treatments. After R.A. Fisher demonstrated the utility of randomization and novel statistical analysis techniques in agricultural research in the 1920s, researchers began adapting this method in medicine.

The impetus for randomized trials also depended on the interaction between professional interests and the regulatory environment. Scandals surrounding drug safety and the for-profit pharmaceutical industry in the 1930s prompted clinical trials in medicine. Journalists, consumer protection organizations, and federal regulators began mounting a campaign for stronger regulatory authority by publicizing a list of harmful products including radioactive beverages and ineffective “cures” for diabetes and tuberculosis [29]. The Food and Drug Administration (FDA) began

to require random assignment and control groups in pharmaceutical testing. As surgeons were relatively unaffected by these controversies, they enjoyed greater freedom in the early twentieth century to “adopt, adapt, or invent” through personal experience and case studies [28]. After all, surgery physically rearranges body tissues, and the end product is visible proof that an intervention has taken place. A purported “magic pill” is much more vulnerable to skepticism.

In 1931, American researchers published an article in the *American Review of Tuberculosis* depicting the first randomized controlled trial with blinding and placebo controls. Amberson and his colleagues used a coin flip randomizing tuberculosis patients to receive either sanocrysin (a gold compound) or distilled water. The resulting data demonstrated that all of the patients receiving sanocrysin suffered adverse systemic drug effects, with no evidence of therapeutic benefit at follow-up [30]. In the very same journal issue, Brock published a study arriving at very different conclusions, that sanocrysin had “an outstanding clinical effect on exudative tuberculosis in white patients,” although “very little effect in limiting the progression of disease in black patients” [31]. In comparison to Amberson’s trial, Brock’s study was demonstrably weaker; he observed 46 patients who were given varying dosages of sanocrysin, did not have an untreated control group, and did not control for baseline differences in treatment setting and disease stage between black and white patients [32]. Clinicians recognized that Amberson’s randomized controlled trial provided stronger evidence, and thus gold therapy for tuberculosis fell into disrepute throughout America.

The first multicenter trials addressing the treatment of pulmonary tuberculosis with streptomycin were published in the United Kingdom in 1948 and the United States in 1952. The British study included 107 patients from 7 centers and concluded that streptomycin-treated patients experienced significantly better outcomes compared to control patients. The Veterans Administration and the United States Armed Services added to the body of evidence from multicenter trials, with good success [33].

One of the earliest randomized controlled trials related to surgery was anesthesiologist Henry Beecher’s 1955 investigation of three different anti-emetics for post-operative vomiting [34]. The year 1958 saw the launch of several randomized controlled trials on surgical procedures, including the management of upper gastrointestinal bleeding, prophylactic surgery for esophageal varices, internal mammary artery ligation, and radical mastectomy [35–38]. Perhaps the largest and most well-known early randomized controlled trial in surgery was performed by C. Goligher’s team in Leeds and York in 1959. The study randomized 634 carefully selected patients to one of three operations for duodenal ulcers. This trial helped lay the design foundation for future trials in surgery. On the importance of random assignment, the study remarks:

This method of randomization may strike some as very impersonal, but we would point out that during the time the trial has been in progress surgical opinion throughout the country on the choice of elective operation for duodenal ulceration has been so divided that in any large hospital several different methods were already in use. Which one would be performed on an individual patient has depended largely on the personal predilection of the particular

surgeon to whom he happened to be referred and not on any accurate knowledge of the relative late results. Our trial has merely organized somewhat this pre-existing system of random usage in order to extract more reliable information from it [39].

Randomized trials aimed, therefore, to settle the conflict of “divided” opinions in the country regarding surgical management of specific diseases and to set a precedent of basing treatments on proven effectiveness rather than on individual surgeon preference. Yet the opening line of the quote clearly articulates how foreign and potentially controversial this methodology was to surgeons in the 1950s and 1960s, many of whom questioned the ethics of denying patients the treatment perceived to be most efficacious.

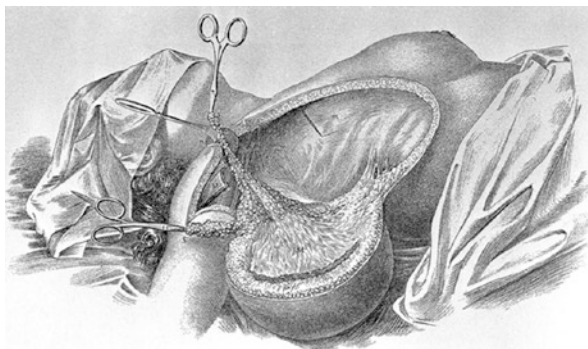
Despite the lack of “high-quality evidence” provided by carefully designed investigations, procedures such as vagotomy and subtotal gastrectomy, among many others, came to be regularly practiced. How did this occur? And how did the standard of proof transform from expert opinion to more standardized trials? A focused history of clinical research in breast cancer surgery offers a lens through which to understand this phenomenon.

1.6 The History of Clinical Research in Breast Cancer Surgery

Propelled by the theory that breast cancer spread centrifugally in the plane of subcutaneous tissues and lymphatics, radical mastectomy remained a mainstay of surgical treatment throughout the first half of the twentieth century [40]. William Halsted did much to pioneer the radical mastectomy, performing the first “Halsted mastectomy” in 1882 [41]. He used clinical and pathologic findings from a series of 210 cases, of which he marked 42% as 3-year cures, results that surpassed those of other surgeons at the time [42]. A 1924 review of 20,000 cases of breast cancer by British statistician Janet Lane-Clayton reported that radical mastectomy offered 43.2% three-year survival rates compared to less than 30% survival from more conservative operations [43]. Halsted’s operation peaked in popularity after World War II as the American Cancer Society pushed for early detection and removal of breast cancers. Some surgeons, believing Halsted’s operation to be insufficient, pushed the envelope even further through “superradical” operations such as removing ribs, deep lymph nodes, limbs, and even internal organs to eradicate cancer cells [27] (Fig. 1.2).

Several case series in Europe began to cast doubt on this prevailing theory, reporting that, in stage I and II breast cancers, more conservative operations resulted in similar survival rates compared to radical mastectomy [45–47]. Similarly, a few American physicians such as Barney Crile presented retrospective data indicating that less radical procedures resulted in equal or better results with fewer side effects compared to the Halsted approach [48]. Moreover, radiotherapy pioneered at the Curie Institute in Paris emerged as a new modality for treatment, and case series demonstrated that when it was either used alone or in combination

Fig. 1.2 Original drawing of the radical mastectomy reported by William S. Halsted in 1894. (Source: William Stewart Halsted, Surgical papers, Wellcome Collection) [44]



with more conservative surgeries, radiation appeared to lend similar or better results compared to radical mastectomy [49, 50]. Physician-historian Barron Lerner points out the strength of a “natural experiment” by Smith and Meyer which showed that simple mastectomies performed during World War II due to staff shortage—therefore representing patients with similar disease severity compared to patients treated in peacetime—led to similar results between simple and radical mastectomies [27, 51].

Radical mastectomy remained standard of care, however. It was viewed as unethical to deprive patients of the ostensibly superior Halsted radical procedure. Surgery carried a strong culture of reliance on expertise gained from firsthand operative experience rather than on biostatistics. Even proponents of conservative surgery such as Barney Crile did not advocate for randomized trials; they felt their personal operative records held sufficient proof of the merits of their approach. Surgeons were concerned that randomized trials would impede on their authority to make individualized decisions for their patients [27].

As providers in the United States debated whether to perform randomized controlled trials to study breast cancer surgery, these very trials were initiated in Europe. Beginning in 1951, researchers used alternate allocation to compare simple mastectomy and radiation with radical mastectomy, showing that the more radical procedure afforded no additional survival benefit [52]. In 1958, radiotherapists Diana Brinkley and J. L. Haybittle launched a randomized controlled study in Cambridge comparing simple mastectomy to radical mastectomy, with all patients receiving radiotherapy. Five- and ten-year survival was equivalent between the two groups [36].

Despite the apparent need for more rigorous studies, it was not until 1971 that the first randomized controlled trial on breast cancer surgery began in the United States. Bernard Fisher began enrolling breast cancer patients to compare radical mastectomy with simple mastectomy [53]. At 25-year follow-up, the study found there was no significant survival advantage gained from performing radical mastectomy to remove occult positive nodes at the time of initial surgery or from radiation therapy. These findings further supported the notion that outcomes from breast cancer surgery relied not on radicality but rather on adequate control of local disease and treatment of secondary tumor spread.

Physician-historian David S. Jones points out that RCTs often were not required or even relevant to promulgate changes in surgical practice [25, 38]. The rates of radical mastectomy had already fallen from 50% in 1972 to 3% in 1981, long before the publication of Fisher's trial [54]. Operative management of breast cancer was already shifting toward more conservative methods; therefore, the RCT made its impact in tandem with other factors such as patient empowerment and new understandings of disease models [27, 54]. The history of breast cancer surgery research illustrates the evolution from empiric clinical gestalt informing decisions to the use of rigorous trials to support or refute longstanding theories. Randomized controlled trials were not foundational to the move away from radical mastectomies, however, and historically such trials have not shaped surgical practice nearly to the same extent as they have shaped medicine.

1.7 Challenges in the Uptake of RCTs in Surgery

Randomized evaluations of surgical techniques are rare, and many interventions have been widely adopted without rigorous evaluation. In the 1990s, an estimated one half of interventions in internal medicine were based on evidence from RCTs, compared to fewer than 25% of surgical interventions [55–57]. In the latter half of the twentieth century, only 3.4% of all articles in the leading surgical journals were randomized controlled trials [58]. This gradually rose to an estimated 10% by 2006 [59, 60].

There are several reasons for the large-scale delay in the uptake of RCTs in surgery [25, 38]. One major reason stems from the blurred lines between clinical practice, innovation, and research, a phenomenon explored by scholars of surgical history. Sally Wilde and Geoffrey Hirst describe how early twentieth century surgeons constantly combined theories about the body with empirical observations in the operating room to innovate new techniques [61]. Surgeries are not controlled by a regulatory body such as the FDA and can be performed without first undergoing extensive evaluation; therefore, regulatory factors are not an impetus to devote the funding and institutional organization required to support large-scale randomized controlled trials [62]. In a survey of surgeons who had published papers describing innovative surgeries, Reitsma and Moreno found that 14 of 21 surgeons confirmed that their work was research, but only 6 had sought IRB approval, and only 7 mentioned the innovative nature of the procedure in the informed consent document [63]. These findings demonstrated a clear need for education and possibly some minimal criteria that define experimentation in performing surgical procedures. In 2009, the IDEAL Collaboration endorsed several suggestions geared toward improving the assessment of surgical innovations, including the use of prospective databases and registries as well as increasing the number of prospective studies with adequate statistical control techniques [64].

Skeptics also felt that surgery was inherently not amenable to standardization, particularly in comparison to medication, where pills maintain the exact same

chemical composition and dose throughout a trial. In contrast, operations comprise hundreds of steps that individual surgeons continually refine and innovate for each particular patient in the hopes of achieving better outcomes [65]. Unlike a clinical trial testing a new medication, variation in surgical skill and experience will allow some surgeons to achieve an adequate result more quickly, whereas other surgeons may need to perform the procedure multiple times to attain the same results.

The ethics of randomized controlled trials in surgery also carries complexities; for instance, establishing evidence using randomized controlled trials would not be ethical in some procedures due to the risk of harm in the nonoperative group. Moreover, studies with placebo sham surgeries have been viewed as unethical because the benefits cannot outweigh the risks of an invasive procedure [66].

1.8 Ethics and Regulation of Clinical Trials

The development of ethical standards with respect to medical experimentation has been an ongoing concern [67]. Military surgeon Walter Reed utilized some of the first written informed consents (in English and Spanish) for his yellow fever trials in Cuba at the turn of the century [68]. The Nuremberg Code of 1949, issued in reaction to Nazi experimentation, was the first document to set out ethical principles based on informed consent. These principles were revised and released by the World Medical Association in 1964 as the Declaration of Helsinki [69]. When thousands of children were born with birth defects as a result of pregnant women taking the drug thalidomide for morning sickness, the 1962 Kefauver-Harris Amendment to the Food, Drug, and Cosmetic Act set forth legal requirements for “adequate and well-controlled investigations” prior to a drug’s approval by the FDA [29, 70]. In 1966, as Henry Beecher was about to publish his exposé on unethical clinical research practices, the US Surgeon General requested that hospitals and universities establish review boards [71].

One of the most infamous clinical trials where ethical principles were lacking was the “Tuskegee Study of Untreated Syphilis” conducted by the US Public Health Service from 1932 to 1972 [72, 73]. It involved nearly 400 black men with late-stage syphilis. When penicillin was found to be an effective cure for syphilis in 1946, the subjects enrolled in the study were not offered this treatment and were not informed of their diagnosis. Jean Heller of the Associated Press broke the story of this study in 1972, revealing that the trial did not have a formal protocol. The magnitude of the risks taken with the subjects involved led many to believe that the Public Health Service had “played” with human lives [72]. The Tuskegee study performed a key role in creating institutions and practices that govern the use of human volunteers in US biomedical research today, but it also introduced a level of distrust between patients and physicians and made the public wary of participating in clinical studies.

In the wake of the tragic disregard for ethical principles in the Tuskegee Study, the National Research Act was signed into law in 1974, culminating in the creation

of the Belmont Report. It put forth three basic ethical principles: respect for persons (to protect autonomy as well as those with diminished autonomy), beneficence (to maximize possible benefits and minimize possible harm), and justice (to divide benefits and burdens of research equally among individuals) [69, 74]. The principles of the Belmont Report have been incorporated into every aspect of human research and are the basis for ethical regulations in practice today.

To date, compared to medical therapies and devices, new surgical techniques have arguably escaped the same type of scrutiny imposed by the FDA and Institutional Review Boards (IRBs) [75]. Designation of a surgical innovation as “experimental” has largely been left to the discretion of the surgeon. To address the concerns that potentially harmful operations could be developed without rigorous evaluation, the American College of Surgeons formulated guidelines in 1995 for the evaluation and application of emerging procedures, urging that new technologies require earlier and continued IRB review, scrutiny of the research protocol, along with a thorough description of informed consent of subjects [63, 76].

The history of clinical research in surgery sheds light on the tension between innovation and strict control. Not long ago, it was common practice for surgeons to take novel operations and technology and apply them to patients after minimal study. The gold standard for building evidence has now assumed the form of a rigorous, expensive, multi-year process. Spurred appropriately by the desire to protect patients from unethical conditions, researchers have foregone rapid innovation in favor of safety. As surgeons navigate ways to improve surgical care, the scientific community will continue to reevaluate the balance between innovation and regulation.

1.9 Recent Times

As clinical trials became more complex, they required additional regulation and administration. Clinical trials at academic centers often ran from specific medical departments. Clinical trial offices (CTOs) emerged over the last two decades to encompass administrative activities related to clinical trials, ranging from protocol development to billing compliance. Their main goal was to enhance institutional research capabilities. A review of CTOs at eight academic health centers in 2008 revealed, however, that there was little uniformity in the structure of functions designated to the CTOs across institutions; some were gatekeepers on all budgeting and billing, and others provided educational or liaison services, while still others wielded monitoring and auditing responsibilities for compliance [77]. This review points to the challenge that institutions face when defining the structure of clinical trial administration. CTOs will become increasingly important as there is added pressure on academic organizations to focus their billing and compliance activities, increase communication between researchers, consolidate education and training, decrease costs and infrastructural redundancy, and increase visibility of trials.

To validate their results and share resources, randomized trials need collaboration and organization among multiple institutions. The Early Breast Cancer Trialists' Collaborative Group was a multinational effort to compile results from randomized trials of adjuvant endocrine and cytotoxic treatments [78]. The creation of the National Cancer Institute (NCI) in 1937 designated the start of federally sponsored medical research in the United States and developed into the National Institutes of Health in the post-World War II years. With the aim of facilitating cross-institutional collaboration, the NCI created several cooperative cancer research groups including the National Surgical Adjuvant Breast and Bowel Project. Collaborative research arising from these groups has helped demonstrate that breast conservation surgery is often better than radical mastectomy, for instance [79]. These multicenter trials are attractive for a number of reasons, including large-scale patient recruitment to achieve the needed numbers, attention to regulatory and ethical issues, as well as marketing strategies. Mega-trials often include thousands of patients with significant heterogeneity in demographics, clinical characteristics, comorbidities, and associated therapies. One downside to applying the findings of a mega-trial to daily medical practice is that group-averaged data are transferred to individual care often with weak demographic and clinical associations [80]. For these reasons, the enthusiasm for international mega-trials has waned somewhat, and researchers have started to focus their energies in more individualized patient-centered research—not all that dissimilar from the patient-particular practices of Hippocratic physicians 2000 years ago. While there may never be a “perfect” trial, clinicians and researchers will continue to employ old, new, and yet-to-be-invented modalities to test the therapeutic potential of interventions in the everlasting goal of providing the best possible care to patients.

References

1. Majno G. *The healing hand: man and wound in the ancient world*. Boston: Harvard University Press; 1975.
2. Nutton V. *Ancient medicine*. New York: Routledge; 2004.
3. Flannery M. Avicenna [updated 12 Oct 2018]. <https://www.britannica.com/biography/Avicenna>. Accessed 5 Dec 2018.
4. Sirasai N. *Avicenna in renaissance Italy: the canon and medical teaching in Italian universities after 1500*. Princeton: Princeton University Press; 1987.
5. Siraisi NG. *Medieval and early renaissance medicine*. Chicago: University of Chicago Press; 1990.
6. Packard F. *The life and times of Ambroise Paré*. 2nd ed. New York: Paul B. Hoeber; 1925.
7. Paré A. *The apologie and treatise of Ambroise Paré*. Birmingham, AL: Classics Of Medicine Library; 1984.
8. Ambroise Paré et l'examen d'un malade. 2011. https://commons.wikimedia.org/wiki/File:James_Bertrand-Ambroise_Paré.jpg.
9. Solís C. Bartolomé Hidalgo de Agüero's 16th century, evidence-based challenge to the orthodox management of wounds. 2011. <http://www.jameslindlibrary.org/articles/bartolome-hidalgo-de-agueros-16th-century-evidence-based-challenge-to-the-orthodox-management-of-wounds/>. Accessed 9 Dec 2018.

10. Lind J. A treatise of the scurvy in three parts. Containing an inquiry into the nature, causes and cure of that disease, together with a critical and chronological view of what has been published on the subject. London: A. Millar; 1753.
11. Bilguer M. A dissertation on the inutility of the amputation of the limbs. London: Baldwin, Becket and de Hondt; 1764.
12. Alanson E. Practical observations on amputation, and the after-treatment. 2nd ed. London: Joseph Johnson; 1782.
13. Poisson D, Larrey D, editors. Recherches de statistique sur l'affection calculeuse, par M. le docteur Civiale. Comptes rendus hebdomadaires des séances de l'Académie de Sciences [Statistical research on (urinary) stone, by M. Dr Civiale]. Paris: Bachelier; 1835.
14. Cheselden W. The anatomy of the human body. 5th ed. London: William Bowyer; 1740. <http://www.jameslindlibrary.org/cheselden-w-1740/>. Accessed 14 Dec 2018
15. Tröhler U. "To improve the evidence of medicine": the 18th century British origins of a critical approach. Edinburgh: Royal College of Physicians of Edinburgh; 2000.
16. Morabia A. Pierre-Charles-Alexandre Louis and the evaluation of bloodletting. *J R Soc Med.* 2006;99(3):158–60.
17. Ackerknecht EH. Medicine at the Paris hospital, 1794–1848. Baltimore: Johns Hopkins Press; 1967.
18. Tröhler U. Statistics and the British controversy about the effects of Joseph Lister's system of antisepsis for surgery, 1867–1890. 2014. <http://www.jameslindlibrary.org/articles/statistics-and-the-british-controversy-about-the-effects-of-joseph-listers-system-of-antisepsis-for-surgery-1867-1890/>. Accessed 9 Dec 2018.
19. Pernick M. A Calculus of suffering: pain, profession, and anesthesia in 19th century America. New York: Columbia University Press; 1985.
20. Chalmers I. Why the 1948 MRC trial of streptomycin used treatment allocation based on random numbers. *J R Soc Med.* 2011;104(9):383–6.
21. Hamilton AL. *Dissertatio Medica Inauguralis De Synocho Castrensi*. Edinburgh: J Ballantyne; 1816. Accessed 9 Dec 2018.
22. Hill B. Clinical lectures on the stricture of the urethra. *BMJ.* 1879;2:883–6.
23. Fox LW. Immediate capsulotomy following the removal of cataract. *JAMA.* 1894;22:837–9.
24. Langton J. A discussion on the treatment of hernia in children. *BMJ.* 1899;2:470–2.
25. Jones DS. Surgery and clinical trials: the history and controversies of surgical evidence. In: Schlich T, editor. *The Palgrave handbook of the history of surgery*. Montreal: Macmillan Publishers; 2018. p. 479–501.
26. Barnes BA. Discarded operations: surgical innovation by trial and error. Costs, risks and benefits of surgery. New York: Oxford University Press; 1977. p. 109–23.
27. Lerner BH. *The breast cancer wars: fear, Hope, and the pursuit of a cure in twentieth-century America*. New York: Oxford University Press; 2011.
28. Wilde S. See one, do one, modify one: prostate surgery in the 1930s. *Med Hist.* 2004;48:351–66.
29. HILts P. *Protecting America's health: the FDA, business, and one hundred years of regulation*. Chapel Hill: University of North Carolina Press; 2003.
30. Amberson J, McMahon B, Pinner M. A clinical trial of sanocrysin in pulmonary tuberculosis. *Am Rev Tuberc.* 1931;24:401–35.
31. Brock B. The sanocrysin treatment of pulmonary tuberculosis in the white and negro races. *Am Rev Tuberc.* 1931;24:436–45.
32. Diaz M. Lessons from using randomization to assess gold treatment for tuberculosis. 2004. <http://www.jameslindlibrary.org/articles/lessons-from-using-randomization-to-assess-gold-treatment-for-tuberculosis/>. Accessed 10 Dec 2018.
33. Marshall G, Crofton J, Hurford J, Brent R, Ross I, Gilmour WS, et al. Streptomycin treatment of pulmonary tuberculosis: a medical research council investigation. *Br Med J.* 1948;2(4582):769–82.
34. Knapp MR, Beecher HK. Postanesthetic nausea, vomiting, and retching; evaluation of the antiemetic drugs dimenhydrinate (dramamine), chlorpromazine, and pentobarbital sodium. *J Am Med Assoc.* 1956;160(5):376–85.

35. Conn HO, Lindenmuth WW. Prophylactic Portacaval anastomosis in cirrhotic patients with esophageal varices—a preliminary report of a controlled study. *N Engl J Med.* 1962;266:743–9.
36. Brinkley D, Haybittle JL. Treatment of stage-II carcinoma of the female breast. *Lancet.* 1966;288:291–5.
37. Cobb L, Thomas G, Dillard D, Merendino K, Bruce R. An evaluation of internal-mammary-artery ligation by a double-blind technic. *N Engl J Med.* 1959;260:1115–8.
38. Crenner C. Placebos and the progress of surgery. In: Schlich T, Crenner C, editors. *Technological change in modern surgery.* Rochester: University of Rochester Press; 2017. p. 156–84.
39. Goligher JC, Pulvertaft CN, Watkinson G. Controlled trial of vagotomy and gastro- enterotomy, vagotomy and antrectomy, and subtotal gastrectomy in elective treatment of duodenal ulcer: interim report. *Br Med J.* 1964;1(5381):455–60.
40. Handley WS. *Cancer of the breast and its operative treatment.* London: Murray; 1906.
41. Ellis H. *A history of surgery.* London: Greenwich Medical Media; 2001.
42. Halsted WS. The results of radical operations for the cure of carcinoma of the breast. *Ann Surg.* 1907;46:1.
43. Lane-Clayton J. *Reports on public health, medical subjects.* London, H.M. Stationary Office; 1924.
44. William Stewart Halsted, surgical papers. Wellcome Collection. <https://wellcomecollection.org/works/bpn6hxgq?query=L0004968>.
45. McWhirter R. Simple mastectomy and radiotherapy in treatment of breast cancer. *Brit J Radiol.* 1955;28:128–39.
46. Williams I, Murley R, Curwen M. Carcinoma of the female breast: conservative and radical surgery. *Brit Med J.* 1953;2:787–96.
47. Mustakallio S. Treatment of breast cancer by tumor extirpation and roentgen therapy instead of radical operation. *J Facult Radiol.* 1956;6:23–6.
48. George Crile J. Simplified treatment of cancer of the breast: early results of a clinical study. *Ann Surg.* 1961;153:745–61.
49. Lee BJ. The therapeutic value of irradiation in the treatment of mammary cancer: a survey of five-year results in 355 cases treated at the memorial hospital of New York. *Ann Surg.* 1928;88(1):25–47.
50. Grace EJ. Simple mastectomy in cancer of the breast. *Am J Surg.* 1937;35:512–4.
51. Holman E. Sir William Osler, William Stewart Halsted, Harvey Cushing: some personal reminiscences. *Surgery.* 1965;57:589–601.
52. Kaae S, Johansen H. Simple mastectomy plus postoperative irradiation by the method of McWhirter for mammary carcinoma. *Ann Surg.* 1969;170(6):895–9.
53. Fisher B, Jeong J, Anderson S, Bryant J, Fisher E, Wolmark N. Twenty-five-year follow-up of a randomized trial comparing radical mastectomy, total mastectomy, and total mastectomy followed by irradiation. *N Engl J Med.* 2002;347(8):567–75.
54. McPherson K, Fox M. Treatment of breast cancer. In: Bunker J, Barnes B, Mosteller F, editors. *Costs, risks and benefits of surgery.* Oxford: Oxford University Press; 1977. p. 308–22.
55. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. A-Team, Nuffield Department of Clinical Medicine. *Lancet.* 1995;346(8972):407–10.
56. Howes N, Chagla L, Thorpe M, McCulloch P. Surgical practice is evidence based. *BJS.* 1997;84(9):1220–3.
57. Kenny SE, Shankar KR, Rintala R, Lamont GL, Lloyd DA. Evidence-based surgery: interventions in a regional paediatric surgical unit. *Arch Dis Child.* 1997;76(1):50.
58. Wente MN, Seiler CM, Uhl W, Buchler MW. Perspectives of evidence-based surgery. *Dig Surg.* 2003;20(4):263–9.
59. Gelijs AC, Ascheim DD, Parides MK, Kent KC, Moskowitz AJ. Randomized trials in surgery. *Surgery.* 2009;145(6):581–7.
60. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet.* 1996;347(9007):984–5.

61. Wilde S, Hirst G. Learning from mistakes: early twentieth-century surgical practice. *J Hist Med Allied Sci.* 2009;64(1):38–77.
62. McLeod RS. Issues in surgical randomized controlled trials. *World J Surg.* 1999;23:1210–4.
63. Reitsma AM, Moreno JD. Ethical regulations for innovative surgery: the last frontier? *J Am Coll Surg.* 2002;194(6):792–801.
64. McCulloch P, Altman D, Campbell WB, Flum D, Glasziou P, Marshall J, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet.* 2009;374:1105–12.
65. Love J. In comment: drugs and operations: some important differences. *JAMA.* 1975;232:37–8.
66. Macklin R. The ethical problems with sham surgery in clinical research. *N Engl J Med.* 1999;341(13):992.
67. Lederer SE. *Subjected to science: human experimentation in American before the second world war.* Baltimore: The Johns Hopkins University Press; 1995.
68. Cutter L. Walter reed, yellow fever, and informed consent. *Mil Med.* 2016;181(1):90–1.
69. Moreno JD. *Undue risk: secret state experiments on humans.* New York: W.H. Freeman; 2000.
70. Krantz J. New drugs and the Kefauver-Harris amendment. *J New Drugs.* 1966;6(22):77–9.
71. Jones D, Grady C, Lederer S. “Ethics and clinical research”—the 50th anniversary of Beecher’s bombshell. *N Engl J Med.* 2016;374:2393–8.
72. Jones J. The Tuskegee syphilis experiment—a moral astigmatism. In: Harding S, editor. *The racial economy of science: toward a democratic future.* Bloomington: Indiana University Press; 1993. p. 275–86.
73. Reverby S. *Examining Tuskegee: the infamous syphilis study and its legacy.* Chapel Hill: University of North Carolina Press; 2009.
74. OHRP. The Belmont report [updated 15 Mar 2016]. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>. Accessed 16 Dec 2018.
75. Burger ISJ, Goodman SN. Ethical issues in evidence-based surgery. *Surg Clin N Am.* 2006;86:151–68.
76. ACo S. Statement on issues to be considered before new surgical technology is applied to the care of patients. *Bull Am Coll Surg.* 1995;80:46–7.
77. Rubin E, Lazar D. *Clinical trials office: what is new in research administration?* Washington, DC: Association for Academic Health Centers; 2009.
78. EBCTCG. *Treatment of early breast cancer. Vol 1. Worldwide evidence, 1985–1990.* Oxford: Oxford University Press; 1990. <http://www.jameslindlibrary.org/early-breast-cancer-trialists-collaborative-group-1990/>.
79. Kelahan A, Catalano R, Marinucci D. The history, structure, and achievements of the cancer cooperative groups. *Managed Care Cancer.* 2001:28–33.
80. Volpe M, Tocci G. Fewer mega-trials and more clinically oriented studies in hypertension research? The case of blocking the renin-angiotensin-aldosterone system. *J Am Soc Nephrol.* 2006;17(4 (Suppl 2):S36–43.