Volume 56 • Number 1 • Spring 1982

I-9

Manuscripts (typewritten, in duplicate, with text and footnotes double spaced), review copies of books, and correspondence relating to the publication of papers should be addressed to:

The Editor
Bulletin of the History of Medicine
1900 E. Monument Street
Baltimore, Maryland 21205

Contributors of articles will receive 25 reprints without covers free of cost. An order form for additional reprints, giving the cost, is sent with the galley proof.

The *Bulletin* is published quarterly: March, June, September, December. One volume of at least 500 pages is published during the year.

Subscription price: Individuals—$18 per year
Institutions—$31 per year

Foreign subscribers add $2.50 for postage. Current single issues are $5.00 each.

New subscriptions are entered to begin with the first issue of the current volume and should be sent to:

The Johns Hopkins University Press
Baltimore, Maryland 21218

Claims for replacement of missing numbers must be received within 3 months (6 months in the case of foreign subscribers) following publication of the number.

Members of the American Association for the History of Medicine receive the *Bulletin* on payment of $18.00 annual dues, of which $13.00 is for a year's subscription. Do *not* send changes in address of members and application for membership to The Johns Hopkins University Press, but to the Secretary-Treasurer:

Dr. Robert P. Hudson
Dept. History and Philosophy of Medicine
University of Kansas Medical Center
Rainbow Blvd. at 39th St.
Kansas City, Kansas 66103

The Fielding H. Garrison Lecture*

CETERIS PARIBUS: THE EVOLUTION OF THE CLINICAL TRIAL

Abraham M. Lilienfeld**

A major characteristic of contemporary medicine is the use of clinical trials for evaluating prophylactic or therapeutic agents. Such trials now constitute a major research area with its own methods and problems. The extent of their importance is indicated by the fact that, in 1979, a Society for Clinical Trials with its own scientific journal was established in the United States. Much of this new emphasis on clinical trials has resulted from the recognition of the value of experimentation in clinical medicine and also because of governmental regulatory agencies. Increasingly, the latter have begun to require trials before biological products are licensed and before a drug is allowed on the market. However, this paper will not focus on these regulatory aspects but on the development of the idea of the clinical trial, a subject which has not received the historical attention it deserves. Some previous work has been done by J. P. Bull, who has provided a general chronological overview of the topic, and by Harry Dowling and Ulrich Tröhler, who have discussed specific aspects in different periods, but no one has taken the approach proposed here.[1]

To trace the evolution of the clinical trial, this paper will first describe such a trial as currently conceived, and then indicate its major components and characteristics, including their role in the structural framework of the trial. Each of these components developed almost independently of each other and therefore has its own history. The paper will then discuss the development and recognition of the importance of these characteristics.

The clinical trial, which is schematically presented in Figure 1, is a systematic experiment in which individuals are randomly allocated to two (or more) groups, one of which is known as a control and the other as an experimental group. The experimental group is administered the new drug

* Presented at the 54th annual meeting of the American Association for the History of Medicine, Toronto, Canada, 14 May 1981.

** Recipient of Career Award No. 5K06 GM13901 from the National Institute of General Medical Science. Research for this paper was supported in part by Grant No. 1RO1 LM03114 from the National Library of Medicine.

[1] J. P. Bull, "The historical development of clinical therapeutic trials," *J. Chronic Diseases*, 1959, *10*: 218–48; Harry F. Dowling, "The emergence of the cooperative clinical trial," *Trans. Studies Coll. Phys. Philadelphia*, 1975, *43*: 20–29; Ulrich Tröhler, "Quantification in British Medicine and Surgery, 1750-1850, With Special Reference to Its Introduction into Therapeutics" (PhD. thesis, University of London, 1978).

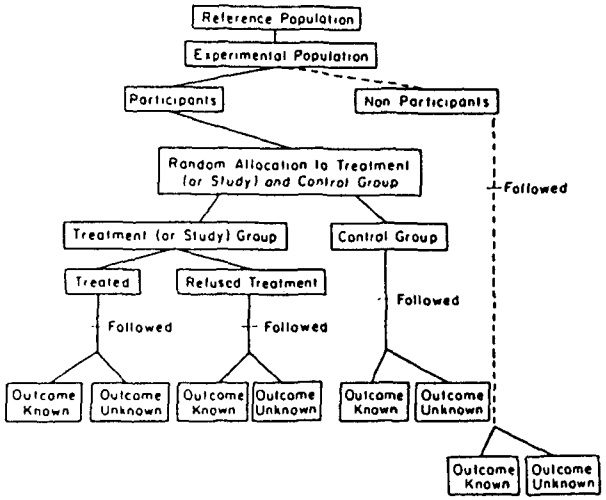

Fig. 1. Outline of Randomized Clinical Trial

being tested and the control (or comparison) group is given the drug in current use; if no such drug exists, then a placebo, an inert substance such as a sugar pill or a saline injection, is used.

To simplify this presentation, clinical trials will be discussed in terms of evaluating the efficacy of a drug in the treatment of a disease. However, there are several types of clinical trials. These include: 1) Therapeutic trials, in which a therapeutic agent or procedure is given to relieve the symptoms and/or improve the survivorship of those with the disease. Examples are laser treatment for diabetic retinopathy and simple mastectomy for breast cancer. 2) Intervention trials, in which the investigator intervenes before a disease has developed in individuals with certain characteristics that increase their risk of developing the disease. The use of anti-hypertensive drugs to reduce the risk of developing stroke and of physical exercise for decreasing the risk of myocardial infarction are examples of these. 3) Preventive or prophylactic trials, in which an attempt is made to determine the efficacy of a preventive agent or procedure. Examples are BCG vaccination or isoniazid for the prevention of tuberculosis.

Clinical trials can be conducted in a single hospital or medical center, or amongst a population group. However, in order to obtain a sufficient number of subjects, it has become increasingly necessary to have collaborative studies involving many centers, hospitals or population groups; these are known as *multi-center trials*.

Clinical trials have the following characteristics:

1. They are *comparative* in nature. A comparison is usually made between an experimental and a control group. This discussion will be limited to comparative trials. Some do use the term "clinical trial" to describe a study in which the investigator determines the effect of a drug, vaccine, and so on, *on individual subjects, e.g., to determine if a vaccine results in the* production of antibodies or if a drug has a toxic effect. These will not be considered as clinical trials for this paper.

2. *Randomization.* Subjects are allocated to the experimental and control groups in a random manner, usually by the use of a table of random numbers. Randomization is used mainly to assure comparability of the experimental and control groups for a variety of factors, except for the one being studied. Randomization also provides the means by which the investigator avoids the introduction of conscious or subconscious bias into the process of allocating individuals to the experimental or control group, thereby increasing degree of comparability. This is best expressed by the term *ceteris paribus*, i.e., "all other factors being equal."

3. *Blind or masked assessment.* Since bias can be introduced consciously or subconsciously in the assessment of the effect of the treatment by the subject, the investigator or the statistician, blinding or masking procedures have been introduced. There are *single blind* studies in which the subjects do not know whether they are in the experimental or study group, but the observer of the subject does. *Double blind* studies are those in which both the subject and the observer of the subject do not know who is in each of these groups. Lastly, there are *triple blind* studies in which the subject, observer and the epidemiologist or statistical analyst do not know who is in which group. After the final results have been analyzed in terms of a comparison between the two groups, the trial is then unblinded in order to derive inferences. Again, the major goal is to achieve *ceteris paribus*.

Since, historically, these attributes of the trial have been introduced independently, the development of each will be described.

Comparative Statistics, Studies and Trials. The earliest recorded account of a comparative study is found in the Old Testament in the first chapter of the Book of Daniel:

> 1. In the third year of the reign of Jehoiakim king of Judah came Nebuchadnezzar king of Babylon unto Jerusalem, and besieged it ...
> 3. And the king spoke unto Ashpenaz his chief officer, that he should bring in certain of the children of Israel, and of the seed royal, and of the nobles ...
> 5. And the king appointed for them a daily portion of the king's food, and of the wine which he drank that they should be nourished for three years ...
> 8. But Daniel purposed in his heart that he would not defile himself with the king's food, nor with the wine which he drank; therefore he requested of the chief of the officers that he might not defile himself ...
> 10. And, the chief of the officers said unto Daniel: "I fear my Lord the king, who

has appointed your food and your drink; for why should he see your faces sad in comparison with the youths of your own age?" ...

11. Then Daniel said to the steward ...

12. Try thy servants, I beseech thee, ten days; and let them give us pulse (leguminous plants) to eat and water to drink ...

13. Then let our countenances be looked upon before thee, and the countenances of the youths that eat of the king's food ...

14. So, he hearkened unto them and tried them in this matter, and tried them for ten days ...

15. And at the end of ten days their countenances appeared fairer and they were fatter in the flesh, than all the youths that did eat of the king's food.[2]

Thinking in comparative statistical terms does not seem to appear again until the fourteenth century, when, in a letter to Boccaccio, Petrarch noted:

> I once heard a physician of great renown among us express himself in the following terms: ..... I solemnly affirm and believe, if a hundred or a thousand of men of the same age, same temperament and habits, together with the same surroundings, were attacked at the same time by the same disease, that if the one half followed the prescriptions of the doctors of the variety of those practicing at the present day, and that the other half took no medicine but relied on Nature's instincts, I have no doubt as to which half would escape.[3]

The development of vital statistics was initiated in 1662 by Graunt's book, *Natural and Political Observations Made Upon the Bills of Mortality.* This work contained the elements of comparative statistics.[4] Populations and mortality statistics were compared for different countries, ages, sexes, and for rural and urban areas. Such statistical comparisons continued during the seventeenth and eighteenth centuries. The idea of comparing mortality among groups became an important aspect of the controversy over the efficacy of smallpox inoculation, as has been pointed out by Genevieve Miller.[5] Thus, James Jurin published several papers between 1723 and 1727, in which he compared the mortality from natural smallpox, which was one death for every five or six cases, with that of those inoculated, where it varied from one death in sixty cases to one in forty-eight.[6]

A classical example of the use of such comparative statistics in the evaluation of therapeutics is the study by Lind in 1747 endeavoring to assess the merits of several existing treatments for scurvy. Lind described his study as follows:

[1] J. J. Slotki, *Daniel, Ezra, Nehemiah, Hebrew Text and English Translation with Introductions and Commentary* (London: Soncino Press, 1951), pp. 1–6.

[2] S. J. Witkoski, *The Evil That Has Been Said of Doctors: Extracts from Early Writers*, trans. with annotations T. C. Minor (Cincinnati, 1889; Reprint from the *Lancet-Clinic*), p. 55.

[3] J. Graunt, *Natural and Political Observations Mentioned in a Following Index, and Made Upon the Bills of Mortality* (London, 1662; Reprinted in Baltimore by the Johns Hopkins Press, 1939).

[4] Genevieve Miller, *The Adoption of Inoculation for Smallpox in England and France* (Philadelphia, University of Pennsylvania Press, 1957), pp. 114–18.

[5] *Ibid.*, p. 118

> On the 20th of May, 1747, I took twelve patients in the scurvy, on board the *Salisbury* at sea. Their cases were as similar as I could have them. They all in general had putrid gums, the spots and lassitude, with weakness of their knees. They lay together in one place, being a proper apartment for the sick in the fore-hold; and had one diet common to all, *viz.*, water-gruel sweetened with sugar in the morning; fresh mutton-broth often times for dinner; at other times puddings, boiled biscuit with sugar, etc.; and for supper, barley and raisins, rice and currants, sago and wine, or the like. Two of these were ordered each a quart of cyder a-day. Two others took twenty-five gutts of *elixir vitriol* three times a-day, upon an empty stomach; using a gargle strongly acidulated with it for their mouths. Two others took two spoonfuls of vinegar three times a-day, upon an empty stomach; having their gruels and their other food well acidulated with it, as also the gargle for their mouth. Two of the worst patients, with the tendons in the ham rigid (a symptom none of the rest had), were put under a course of sea-water. Of this, they drank half a pint every day, and sometimes more or less as it operated, by way of a gentle physic. Two others had each two oranges and one lemon given them every day. These they ate with greediness, at different times, upon an empty stomach. They continued but six days under this course, having consumed the quantity that could be spared. The two remaining patients took the bigness of a nutmeg three times a-day, of an electuary recommended by a hospital-surgeon, made of garlic, mustard-seed, *rad. raphan.*, balsam of Peru, and gum myrrh; using for common drink, barley-water well acidulated with tamarinds; by a decoction of which, with the addition of *cremor tartar*, they were gently purged three or four times during the course.
>
> The consequence was, that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the end of six days fit for duty. The spots were not indeed at that time quite off his body, nor his gums sound; but without any other medicine, than a gargarism of *elixir vitriol*, he became quite healthy before we came into Plymouth, which was on the 16th of June. The other was the best recovered of any in his condition; and being now deemed pretty well, was appointed nurse to the rest of the sick.[7]

Another somewhat later example of the use of comparative statistics is the report by Robert Robertson in 1776 on the use of bark in the treatment of continuous fever.[8] When serving as ship's surgeon on the *Juno* of the British Royal Navy, Robertson's stock of bark lasted from April to December 1776, after which he had to utilize other therapeutic methods. A comparative study thus inadvertently occurred when he compared the case fatality percentages in these two periods, as shown in Table 1. Robertson was not alone in making this sort of use of statistics. Other British physicians during the eighteenth century who utilized arithmetic methods included William Black, John Millar, Gilbert Blane, John Pringle, John Ferriar, John Coakley Lettsom, and others. Tröhler has very aptly designated this group as "arithmetic observationists."[9]

[7] J. Lind, *A Treatise of the Scurvy* (Edinburgh: Sands, Murray and Cochran, 1753), pp. 191–93.

[8] Tröhler, "Quantification in British Medicine," pp. 235–36.

[9] *Ibid.*

Table 1 Results of Treatment of Continuous Fever, Robert Robertson, 1776

Treatment With	No. Treated on Board Ship	Died on Board	
		No.	%
Bark (April 4–Dec. 31, 1776)	216	1	0.4
All Other Methods	296	6	2.0

Such comparative studies continued into the nineteenth century. The comparisons of vital statistics gradually became more refined, paralleling developments of statistical methods. Thus, life tables or tables of mortality were increasingly used for comparing the mortality experience of communities and other groups. The hygienic and sanitary reform movement utilized vital statistics as a justification for proposed reforms since they provided an index of the health of the community. The work of Louis René Villermé, who compared the mortality experience of different social districts in Paris, and that of William Farr, who developed many of the methods still used today, represent examples of the use of comparative statistics.[10] During the nineteenth century, statistical methods began to be used to describe and investigate many human activities. These developments have been previously discussed and need not be presented here except to emphasize that both the sanitary and statistical movements in the nineteenth century were a product of the scientific and social milieu in which clinical trials gradually developed.[11]

In the early part of the nineteenth century in the Parisian School, the concept of comparative studies in general, and more specifically their use in evaluating treatment as well as in observational epidemiologic studies, received their most systematic exposition. P. C. A. Louis outlined this development in his exposition of "la méthode numérique."[12] It is best exemplified by his statement:

> In any epidemic, for instance, let us suppose five hundred of the sick, taken *indiscriminately*, to be subjected to one kind of treatment, and five hundred others, taken in the same manner, to be treated in a different mode; if the mortality is greater among the first than among the second, must we not conclude that the treatment was less appropriate, or less efficacious in the first class, than in the second?[13]

[10] See, for example, M. J. Cullen, *The Statistical Movement in Early Victorian Britain: The Foundations of Empirical Social Research* (New York: Barnes and Noble, 1975), pp. 135–46; Louis René Villermé, "De la mortalité dans les divers quartiers de la ville de Paris, et des causes qui la rendent très différente dans plusieurs d'entre eux, ainsi que dans les divers quartiers de beaucoup de grandes villes," *Ann. d'Hyg. Pub.*, 1830, *3*: 294–341; John M. Eyler, "The conceptual origins of William Farr's epidemiology: numerical methods and social thought in the 1830s," in Abraham Lilienfeld (ed.), *Times, Places, and Persons: Aspects of the History of Epidemiology* (Baltimore: The Johns Hopkins University Press, 1980), pp. 1–21.

[11] The best discussion of this topic is Cullen, *Statistical Movement*.

[12] David E. Lilienfeld and Abraham M. Lilienfeld, "The French influence on the development of epidemiology," in Lilienfeld (ed.), *Times, Places and Persons*, pp. 28–38.

[13] P. C. A. Louis, *Researches on the Effects of Bloodletting in Some Inflammatory Diseases and on the Influence of Tartarized Antimony and Vesication in Pneumonitis*, trans. C. G. Putnam (Boston: Hilliard, Gray, 1836), pp. 59–60.

The term "indiscriminately" is emphasized because it suggests that per. Louis had the concept of randomization in mind, a concept which will be discussed below. Many other examples of Louis's utilization of comparative studies can be found in his works.[14]

In addition to this, Louis clearly enunciated the idea that when groups are compared they should be similar with respect to the many factors that may influence the disease under consideration.

> I come now to therapeutics, and suppose that you have some doubt as to the efficacy of a particular remedy: How are you to proceed? Will you compare two cases in which the remedy has been employed with two similar ones in which it has not? Surely not; for you know that the inference drawn from them would be of no general value. You would take as many cases as possible, of as similar a description as you could find, and would count how many recovered under one mode of treatment, and how many under another; in how short a time they did so; and if the cases were in all respects alike, except in the treatment, you would have some confidence in your conclusions; and if you were fortunate enough to have a sufficient number of facts from which to deduce any general law, it would lead to your employment in practice of the method which you had seen oftenest successful.[15]

It is of more than passing interest that many of Louis's students, including Elisha Bartlett, William Farr, William Budd, and H. I. Bowditch, became involved in the development of both the hygienic and statistical movements in England and the United States.[16]

During the nineteenth century, probably the most sophisticated clinical trial of a preventive type was conducted by Ignaz Semmelweis.[17] As is well known, Semmelweis observed mortality differences from puerperal fever in the two Divisions of the Lying-In Hospital in Vienna. Physicians and medical students were assigned to the First Division and pupil midwives to the Second Division. Mortality was noticeably higher in the First Division. Between 1841 and 1847, this difference had been noted by a number of observers and many explanations had been proposed. Semmelweis evaluated each of the proposed hypotheses using methods and reasoning processes that were quite sophisticated. Semmelweis noted that a consequence of the great emphasis on the study of pathology in the Vienna Medical School was that professors, assistants and students were in frequent contact with cadavers, while midwives had no such contact. Semmelweis stated, "That the cadaveric particles clinging to the hands are not entirely removed by the ordinary method of washing the hands with soap, is shown by the cadaveric odor, which the hand retains for a longer or shorter time."[18] From contaminated

[14] Lilienfeld and Lilienfeld, "French Influence," pp. 30–31.

[15] P. C. A. Louis, "The applicability of statistics and the practice of medicine," *London Med. Gaz.*, 1837, *20*: 490–91.

[16] Lilienfeld and Lilienfeld, "French Influence," pp. 33–35.

[17] Ignaz Philipp Semmelweis, "The etiology, the concept and the prophylaxis of childbed fever" (1861), trans. F. P. Murphy, *Med. Classics*, 1941, *5*: 350–773.

[18] *Ibid.*, p. 392.

Table 2. *Comparison of Maternal Mortality From Childbed Fever in the Physicians' and Midwives' Divisions of the Lying-In Hospital in Vienna in 1841–1848*

	Physicians' Division			Midwives' Division		
		Deaths			Deaths	
Year	Births	No.	%	Births	No.	%
1841	3036	237	7.7	2442	86	3.5
1842	3287	518	15.8	2659	202	7.5
1843	3060	274	8.9	2739	169	6.2
1844	3157	260	8.2	2956	68	2.3
1845	3492	241	6.8	3241	66	2.03
1846	4010	459	11.4	3754	105	2.7
1847				3306	32	0.9
January–May	2134	120	5.6			
	Intervention introduced in May					
June–December	1841	56	3.04			
1848	3556	45	1.27	3219	43	1.33

hands, therefore, he believed, cadaveric particles were being introduced during the examination of women in labor.

In May 1847, Semmelweis began the use of "chlorina liquida." Each student in the First Division washed his hands with this before an examination. This substance was later replaced by chlorinated lime, which was less expensive. This method of prophylaxis immediately reduced the mortality from puerperal fever in the First Division to a rate similar to that in the Second Division, as shown in Table 2. It is somewhat puzzling that Semmelweis did not discuss his reasons for using "chlorina liquida" or chlorinated lime as a disinfectant.

Semmelweis also reviewed the mortality experience in the hospital prior to 1840 when, as he expressed it, "medicine, still indulging in theoretical speculation, got along without anatomical principles."[19] He found that, in that period, the death rates from puerperal fever were similar for patients attended either by physicians and medical students, or by midwifery students. In 1840, when pathology became an essential part of the course of study in Vienna, the death rate in the Physician's Division increased as shown in Table 3.

Another area for comparative studies in the nineteenth century was cholera therapy. For example, comparisons of different methods of treatment of cholera in the 1845 epidemic were reviewed and analyzed by the Treatment Committee of the Medical Council in Great Britain in a 1855 report to the General Board of Health.[20] It is of more than passing interest

[19] *Ibid.*, p. 397.

[20] British Parliamentary Papers: *Reports on the Epidemics of 1854 and 1856 and Other Reports on Cholera With Appendices 1854–56. Report on the Results of the Different Methods of Treatment Pursued in Epidemic Cholera (1855)* (Shannon: Irish University Press, 1970), pp. 657–684.

Table 3. *Comparison of Maternal Mortality from Childbed Fever in the Physicians' and Midwives' Divisions of the Lying-In Hospital in Vienna in 1833–1858*

Years	Physicians' Division (%)	Midwives' Division (%)
1833–1840	6.56	5.58
1841–1846	9.92	3.38
1847–1858	3.57	3.06

that the necessary information was obtained by a survey of metropolitan hospitals and medical practitioners. The degree of sophistication displayed in evaluating the data is indicated by the following statement in the report:

> In order to judge correctly of the value of this evidence, it is necessary to examine as far as may be possible, the degree of severity of the cases brought beneath this test. The only means at our command (on the present occasion at least) to make this examination is to consider the relative proportion which the cases of collapse bear to the number of deaths of their own classes, respectively.[21]

The members of the treatment committee were also concerned about the question of dosages of the different medications but did not have adequate information by which to evaluate this.

In the 1860s, Joseph Lister utilized comparative statistics to evaluate the effects of antisepsis on the mortality from amputations.[22] He analyzed the hospital records of the Glasgow Royal Infirmary for two of the years prior to the "antiseptic period" ("the hospital records are unfortunately imperfect for one of the three years immediately preceding the antiseptic period") and for three years during the antiseptic period. The findings are presented in Table 4. Lister noted that "These numbers are, no doubt, too small for a satisfactory comparison; but when the details are considered they are highly valuable with reference to the question we are considering."[23]

The idea of comparative experimentation was also expressed by Claude Bernard in 1865 in his *Introduction to the Study of Experimental Medicine:*

> Especially in therapeutics, the need of comparative experiment has always struck physicians endowed with the scientific spirit. We cannot judge the influence of a remedy on the course and outcome of a disease if we do not previously know the natural course and outcome of the disease ... To be valid, comparative experiments have therefore to be made at the same time and on as comparable patients as possible... I shall recall only one recent example concerning the treatment of pneumonia. Comparative experiment showed, in fact, that treatment of pneumonia by bleeding, which was believed most efficacious, is a mere therapeutic illusion ..... To learn we must necessarily reason about

[21] *Ibid.*, p. 669.

[22] Joseph Lister, "On the effects of the antiseptic system of treatment upon the salubrity of a surgical hospital," *Lancet*, 1870, *1*: 4–6, 40–42.

[23] *Ibid.*, p. 40.

Table 4. Perce[illegible]aths Among Amputations Before and After Antiseptic Period

Year	No. of Amputations	Deaths	
		No.	Percent
	Before Antiseptic Period		
1864	17	7	41
1866	18	9	50
	35	16	46
	During Antiseptic Period		
1867	7	0	0
1868	17	3	18
1869	16	3	19
	40	6	15

what we have observed, compare the facts and judge them by other facts used as controls.[24]

His use of the word "controls" without any explanation suggests that the term was in general use. However, no more direct use of this term "controls" applied to one group in an experiment has been found until 1890 when Hankin used the term "control-mice" in describing a series of immunization experiments against tetanus in mice.[25] Immunized mice were inoculated with tetanus and did not show any symptoms whereas two control-mice died in thirty-six hours.

It has been suggested that the original meaning of the word "control" was "check," and that it was derived from "counter-roll" (or the French *contre-rolle*), which meant a duplicate register or account that was used to verify a first-made account.[26] In 1893 the New English Dictionary defined "controls" as "a standard of comparison used to check the inferences deduced from an experiment by application of the Method of Difference" referring to one of John Stuart Mill's methods of experimental inquiry.[27]

In the 1890s, based on bacteriological and immunological studies, Emil Behring and his colleagues developed diphtheria antitoxin serum and evaluated its efficacy on thirty patients, six of whom died.[28] They compared this result with the usual mortality of fifty percent for the disease and with the previous year's mortality in the same hospital of sixty-six percent. In 1898, Johannes Fibiger in Blegdam Hospital, Denmark, improved the quality of the controlled trial of anti-diphtheria serum by using alternate controls: every

[24] Claude Bernard, *An Introduction to the Study of Experimental Medicine*, trans. Henry Copley Greene (1st ed. 1865, New York: Dover Publications, 1957), pp. 194–95.
[25] E. H. Hankin, "A cure for tetanus and diphtheria," *Nature*, 1890, *43*: 122.
[26] Edwin G. Boring, "The nature and history of experimental control," *Amer. J. Psych.*, 1954, *67*: 573–89.
[27] *Ibid.*, p. 574.
[28] E. Behring, O. Boer and Kossel, "Zur behandlung diphtheriekranker menschen mit diphtherieheilserum," *Deut. Med. Woch.*, 1893, *19*: 349–418.

Table 5. Study Designs Used in Serum Treatment of Lobar Pneumonia in Boston City 1923–1929

Years	Study Design
1923–1924	No alternation.
1924–1929	Control series used: serum given to alternate cases as soon as diagnosis was made.
1925–1926	Only cases entering hospital within 4th day of disease onset used in study.
1928–1929	Cases admitted to two of the medical services were treated, leaving those on the two other services as controls.

other patient was treated and every other one served as a control, in the sequence—treated, control, treated, control and so on. He then determined whether the treated and untreated groups were comparable with respect to age, symptoms and severity of the disease.[29] The use of alternate controls marked a conceptual advance toward the goal of *ceteris paribus*.

The major areas of medical research interest in the early twentieth century were clearly in the prophylaxis or treatment of infectious diseases, including acute respiratory infections or colds and pneumonia. A review of the published literature of the period from 1900 to 1930 indicates the beginning of an appreciation of the need for controls and a gradually increasing refinement in experimental design. The serum treatment of lobar pneumonia is an example of therapy in which different approaches were employed during this period of time. This is shown by the different methods of conducting clinical trials that were used at the Boston City Hospital, which were summarized by Maxwell Finland and are shown in Table 5.[30] There was general dissatisfaction amongst the medical profession with the long series of reports on the results of the serum treatment of pneumonia until the advent of the sulfonamides.

However, a certain refinement of design was developed using the alternate control method, as shown in a report by John Wyckoff and others.[31] In 1930, these investigators attempted to assess the therapeutic value of digitalis in the treatment of pneumonia. Use of the drug had been initiated in 1916 and had become extensive. In their report the investigators stated, "The analysis of hospital records of previous years with the study of digitalis in untreated and treated groups does not constitute a properly controlled series." These investigators decided that:

Patients should be selected for the digitalis or nondigitalis groups in the following way: Patients were received into the "pneumonia series" according to the

[29] Johannes Fibiger, "Om serum behandlung of difteri," *Hospitalstidende*, 1898, *6*: 309–25.
[30] Maxwell Finland, "The serum treatment of lobar pneumonia," *New Eng. J. Med.*, 1930, *202*: 1244–47.
[31] John Wyckoff, Eugene F. Dubois and I. Ogden Woodhill, "The therapeutic value of digitalis in pneumonia," *JAMA*, 1930, *95*: 1243–49.

date and hour of admission, and alternate patients were treated with serum. The combination of serum and digitalis therapy led to the grouping of patients into four classes, selected only by the time of admission, and termed arbitrarily A, B, C and D. The treatment of the patients in these four classes was as follows:

Class A received neither serum nor digitalis.
Class B received serum only.
Class C received digitalis only.
Class D received both serum and digitalis.

This system of classification operated in each ward independent of other wards, so that factors of general care and nursing might be the same for each class of treatment.[32]

The investigators found that the mortality among digitalis-treated patients was about twenty-five percent higher than among the untreated group and therefore concluded that the routine administration of digitalis to patients with lobar pneumonia was dangerous. This led to the discarding of digitalis as a method of treatment of pneumonia.

This issue of serum treatment of pneumonia stimulated the first collaborative trial that has been found reported in the literature. It was described in 1934 by the Therapeutic Trials Committee of the Medical Research Council in Great Britain.[33] The degree of collaboration was not as rigorous as current collaborative trials; its extent can be noted from the phrases used in the report. "Certain principles of selection were laid down so as to make the data derived from the centres homogeneous, and to exclude from the comparison patients in whom the serum could not be expected to have any effect." The method agreed upon by the collaborating centers in London, Edinburgh and Aberdeen was that alternate cases of lobar pneumonia taken simply in the order of their admission into the hospital should be used respectively for serum treatment and controls. The study in London included various London County hospitals and St. Bartholomew's Hospital. "That these limitations were desirable was agreed upon by all the workers at a preliminary conference on the subject."[34]

It is worth noting that the Therapeutic Trials Committee was formed by the Medical Research Council in 1931 at the request of the Association of British Chemical Manufacturers.[35] The chairman of this committee defined its object as being the supplying of manufacturers with clinical reports on new therapeutic agents which they were preparing to place on the market. His statement is worth quoting since it shows the broad conceptual base of the Committee's interests, which included economic and social considerations as well as scientific ones, and these remain relevant today. He stated:

[32] *Ibid.*, p. 1243.
[33] "The serum treatment of lobar pneumonia: A report of the Therapeutic Trials Committee of the Medical Research Council," *Lancet*, 1934, *1*: 290–95.
[34] *Ibid.*, p. 291.
[35] Annotations, "Clinical trials of new remedies," *Lancet*, 1931, *2*: 304.

Anyone accustomed to read continental medical and technical literature must notice the frequency with which pharmacological and clinical reports regarding the new products put out by private firms appear. In this country it is much more difficult, for reasons which are well understood, to secure for publication clinical reports on privately owned remedies, and the British manufacturer is consequently at a disadvantage. The committee removes that disadvantage, and by reason of its impartial standing strengthens the claims of any remedy of which it approves. The committee acts at the request of the manufacturer, who is required to submit with his product all available information regarding it. If the substance appears likely to be a useful remedy, supplies are distributed to a number of clinicians in different parts of the country, and in accordance with the reports received, the manufacturer is advised whether the product is of value or not. If the remedy is approved the committee supplies clinical reports which the manufacturer may use for the purpose of introducing his product. On the other hand, the committee may refuse to deal with any product if the description and statement supplied by the manufacturers make it seem that the product is not likely to be worth investigating. In the event of a product being tried and reported upon unfavourably the manufacturer is advised not to proceed, and the committee reserves the right to publish the reports . . . It is in the nature of things that the best work done by such a committee, that is the investigations in relation to the drugs which eventually are reported upon unfavourably, will not be made public. I describe this as the best work, since a careful and comprehensive opinion on a product will benefit the public and the medical profession by keeping off the market valueless drugs which might otherwise be brought into use on insufficient evidence, and will save the manufacturer the expense and effort of launching an article which is destined ultimately to be unsuccessful. A good product, on the other hand, would be expected, in any case, eventually to gain acceptance, though the favourable report of an impartial committee of a public body will be a great advantage to the British manufacturer in marketing the remedy.[36]

The announcement in 1931 of the formation of this committee in both the *Lancet* and the *British Medical Journal* contains one of the first published uses of the term "clinical trial."[37] In reviewing the annual reports of the Medical Research Council, it was found that the term "clinical trial" appeared in the body of the annual report for 1928-29. It must have been an accepted term since it was used without any particular note.

In addition to infectious diseases, an area of importance for the development of clinical trials was the investigation of nutritional diseases, particularly pellagra. Classical comparative studies were conducted by Joseph Goldberger in the form of a preventive diet for pellagra.[38] These comparative trials were conducted during 1914-17 in two orphanages in Jackson,

[36] Cited in C. H. Hampshire, "Pharmacy in retrospect and prospect," *Lancet*, 1933, *2*: 275.
[37] Annotations, "Clinical trials," p. 304.
[38] J. Goldberger, C. H. Waring and W. F. Tanner, "Pellagra prevention by diet among institutional inmates," *Pub. Health Reps.*, 1923, *38*: 2561–68. For a more detailed discussion of Goldberger's work, see Milton Terris (ed.), *Goldberger on Pellagra* (Baton Rouge: Louisiana State University Press, 1964).

Mississippi, and two wards of a Georgia State Sanitarium. They were generally similar to those that were conducted by Semmelweis.

Randomization. Another feature of the clinical trial is that it incorporates use of the idea of randomization. This concept, although alluded to by investigators in the nineteenth century, was first formally introduced in agricultural, not medical, research by R. A. Fisher in 1923 at the Rothamsted Experimental Station in England as part of experimental design. Fisher had developed a new method of analysis for agricultural field experiments—the analysis of variance, a method that depended upon randomization. He introduced this concept with the statement, "If all the plots are undifferentiated *as if the numbers* had been mixed up and written down in random order . . ."[39] This was based on Fisher's view that in applying statistical tests in analyzing the experimental data, it was necessary to assume that the observations had been independently made. He perceived that random allocation of treatments to agricultural plots would simulate independence and therefore the statistical analysis would be valid. Previously, agricultural experiments had been conducted using systematic designs in which the fertility of adjoining plots, for which different treatment had been applied, clearly was not independent.

Joan Fisher Box points out in a discussion of Fisher's work that randomization was not readily accepted either by experimenters or mathematicians.[40] It was not until 1926 that the first randomized block design was conducted in agriculture. Other experimenters such as W. S. Gossett, as late as 1936, expressed the view that "Since the tendency of deliberate randomization is to increase the error, a balanced arrangement . . . is best."[41] However, investigators concerned with experimentation gradually adopted the concept of randomization.

In clinical trials, the first mention of a formal method of random assignment of patients is found in a report on a clinical trial of sanocrysin in the treatment of pulmonary tuberculosis.[42] Sanocrysin was a gold compound that had first been introduced in 1924. Various gold compounds were then being tried throughout the world as therapeutic agents and differences of opinion developed as to the value of such treatment. In 1931, J. Burns Amberson and his colleagues reported the results of a trial which had been started in 1926 at the *W. H. Maybury Sanitorium*, Northville, Michigan because ". . . none of the clinical reports submitted up to late in 1926 seemed

[39] Ronald Aylmer Fisher and W. A. Mackenzie, "Studies in crop variation: II. The manurial response of different potato varieties," *J. Agric. Sci.*, 1923, *13*: 315 and Ronald Aylmer Fisher, "The arrangements of field experiments," *J. Ministry Agric. Great Britain*, 1926, *33*: 503–13. Both are contained in *Collected Papers of R. A. Fisher*, ed J. H. Bennet, 5 vols. (Adelaide: University of Adelaide, 1971, 1972). They are in vol. 1, no. 32 and vol. 2, no. 48, respectively.

[40] Joan Fisher Box, "R. A. Fisher and the design of experiments, 1922–1926," *Amer. Stat.*, 1980, *34*: 1–7.

[41] *Ibid.*, p. 3.

[42] J. Burns Amberson, Jr., B. T. McMahon and Max Pinner, "A clinical trial of sanocrysin in pulmonary tuberculosis," *Amer. Rev. Tuberculosis*, 1931, *24*: 401–35.

fully conclusive as to the value of the drug." More specifically, the report stated that:

> An examination of some of the clinical reports does not seem to justify all the conclusions that have been drawn for or against sanocrysin. Many so-called control cases do not serve such a designation, and various interpretations of clinical observations are open to question. It is difficult, indeed, in appraising a therapeutic agent for tuberculosis, to distinguish natural fluctuations of the disease from those induced by a substance or procedure under trial. The follow-up of large groups of cases during a period of at least five years after treatment usually yields valuable, if limited, information provided there is not too much disparity in the size of comparable groups. Another method is an intensive study of smaller, carefully selected, and closely comparable groups of cases, according to a prearranged plan. This also is satisfactory for limited purposes but requires special facilities and personnel. Neither of these methods have been applied very rigidly in the case of sanocrysin.[43]

These investigators carefully selected a small group of cases, twenty-four to be exact. They then divided these twenty-four patients into two approximately comparable groups of twelve each. The cases were:

> . . . individually matched, one with another, in making this division. Obviously the matching could not be precise, but it was as close as possible, each patient having previously been identified independently by two of us. Then, by a *flip of a coin*, one group became identified as Group I (sanocrysin-treated) and the other as Group II (control). The members of the separate groups were known only to the nurse in charge of the ward and to two of us. The patients themselves were *not aware of any distinction in the treatment* administered.[44] (Italics added).

They further stated, "We report finally on only twelve treated cases and eleven 'controls.' Ordinarily this small number would be insufficient to afford much information."[45] However, they found that the sanocrysin-treated cases became worse and that the one patient in the drug-treated cases who died had died from gold poisoning.

This report is interesting since it contains the first mention of the use of random allocation in clinical trials, by flipping of a coin although, in this instance, it was done for groups rather than for individuals. In addition, it is the first report discussing the actual matching of individuals for a group of characteristics, as well as including the first mention of the fact that the patients were not aware of the method of treatment.

Studies of immunization against acute respiratory disease had been conducted using vaccines of varying composition since 1921.[46] Controls

[43] *Ibid.*, p. 402.

[44] *Ibid.*, p. 403–4.

[45] *Ibid.*, p. 430.

[46] Ann I. Von Sholly and William H. Park, "VII. Report on the prophylactic vaccination of 1536 persons against acute respiratory diseases, 1919–1920," *J. Immunology*, 1921, *6*: 103–15; Fergus R. Ferguson, A. F. C. Davey and W. W. C. Topley, "The value of mixed vaccines in the prevention of the common cold," *J. Hygiene*, 1927, *26*: 98–109; George E. Rockwell, Herman C. Powell, "Oral immunization to colds," *J. Immunology*, 1935, *28*: 475–83.

were selected for comparative purposes in these trials, but the selection of such controls apparently was not done systematically. In 1938, Harold Diehl and co-workers reported such a trial conducted at the University of Minnesota, stating that "At the beginning of each year of the study, students were assigned at random and without selection to a control or experimental group."[47] In this paper Diehl mentioned that Mogath and Berkson at the Mayo Clinic in 1936–37 had conducted a similar study of oral vaccination, in which "The experimental and controls groups were carefully equated and treated in a like manner."[48] None of these vaccines proved to be effective.

The concept of randomization gradually gained acceptance in clinical medicine after 1940, particularly as chemotherapeutic agents became available for treatment of such diseases as tuberculosis, where the biological and clinical complexity of the disease led to the development of increased refinement and rigor in the clinical trials.

H. Corwin Hinshaw and William H. Feldman of the Mayo Clinic, in 1944, pointed out the difficulties in interpreting the results of previous studies of various methods of treating tuberculosis.[49] They emphasized that the principles of objective experimentation applied to animals had to be utilized in therapeutic experimentation in humans. They further suggested a procedure for such experimentation which emphasized the need for making the experimental and control groups as nearly similar as possible. They also pointed out that "Some procedures of chance could be resorted to in deciding which patients are to receive treatment with the drug." Further, "In our own studies now under way, pairs of patients who had as nearly comparable disease as possible were selected and the toss of a coin decided which members were to receive treatment with the drug. The remaining one is considered the control."[50] The Medical Research Council of Great Britain's study of streptomycin treatment of tuberculosis, which was reported in 1948, appears to be the first in which patients were allocated to the experimental and control groups by random sampling numbers.[51]

Blind or Masked Assessment. The third characteristic of the modern clinical trial is the blind or masked assessment of the effect of the agent. As was mentioned earlier, the purpose of blind assessment, although varying with the disease or end point under study, is to prevent it becoming known which patient receives the treatment and thereby influencing or biasing the assessment of the effect.

[47] H. S. Diehl, A. B. Baker and D. W. Cowan, "Cold vaccines: an evaluation based on a controlled study," *JAMA*, 1938, *111*: 1168.
[48] *Ibid.*, p. 1172.
[49] H. Corwin Hinshaw and William H. Feldman, "Evaluation of chemotherapeutic agents in clinical trials: A suggested procedure," *Amer. Rev. Tuberculosis*, 1944, *50*: 202–13.
[50] *Ibid.*, p. 205.
[51] Medical Research Council, "Streptomycin treatment of pulmonary tuberculosis," *Brit. Med. J.*, 1948, *2*: 769–82.

Insofar as it can be ascertained, the first double *blind* trial was the previously mentioned one conducted by Amberson and others in 1931 evaluating the efficacy of sanocrysin, where the controls received intravenous injections of distilled water.[52] It was not until 1938 that the report by Diehl and his colleagues of the use of a placebo (saline solution) in their study of the prophylactic value of cold vaccines appeared; they referred to the saline solution as a placebo.[53] The Diehl report mentioned that in a Mayo Clinic trial of an oral vaccine, conducted during 1936–37, lactose-filled capsules were used for the control groups.

In 1944, Hinshaw and Feldman suggested that "administration of inert preparations to the control patients to exclude the factor of mental suggestion would also be to some advantage."[54] When streptomycin became available for the treatment of tuberculosis in 1948 in the United States, it was not possible initially for the participating groups to utilize controls because the investigating physicians considered it to be unethical to withhold treatment. They therefore decided to compare the results of the streptomycin treatment with the individual's immediate progress before the initiation of treatment with streptomycin.

In the meantime, the British Medical Research Council initiated the cooperative trial, discussed above.[55] Since streptomycin was in short supply in Great Britain, no ethical problems were involved and a collaborative, controlled trial was started in 1946. Although a placebo was not used, the patients were not notified of the group to which they had been assigned; each group was in different wards. Blindness of assessment was assured by having two radiologists read the x-ray films independently without knowing to which patient they belonged. The results of change in x-rays at six months and one year were reported. Thus, in 1948, the essential prototype of the modern randomized double-blind clinical trial had evolved, with all of the components now seen as forming part of such a trial having been assembled.

In 1951, Otho B. Ross commented on the need for adequately controlled clinical trials for evaluating therapy.[56] He reviewed 100 articles from leading medical journals published between January and June 1950, that dealt with some procedure or form of treatment that was either recommended or condemned. He found that forty-five percent had no controls, eighteen percent had inadequate controls, in ten percent controls were impossible, and twenty-seven percent were well-controlled. Of interest was his quotation from a book on experimental sociology published in 1945: "Effective control is the key to the entire experimental procedure. It is essential too for the accuracy of conclusions. Without proper control we cannot be certain that the causal nexus which we seek to establish is a real

[52] Amberson, McMahon and Pinner, "A clinical trial," p. 406.
[53] Diehl, Baker and Cowan, "Cold vaccines," p. 1168.
[54] Hinshaw and Feldman, "Evaluation of chemotherapeutic agents," p. 205.
[55] Medical Research Council, "Streptomycin treatment," p. 769.
[56] Otho B. Ross, Jr., "Use of controls in medical research," *JAMA*, 1951, *145*: 72–75.

Table 6 ...ion of Clinical Trials: Historical Summary

Historical Introduction	
A) Comparative Studies	
(No Systematic Intervention)	
1. Depend on previous experience (Memory—uncontrolled)	"Arithmetic Observationists"—end of 18th century.
2. Comparison of results of different treatments:	
a) Reported in literature	Jurin, Smallpox inoculation—1721–1727
b) Analysis of records	Jurin, Smallpox inoculation—1721–1727
c) Historical comparisons	Lister—1870 Diphtheria antitoxin—1890s
d) Concurrent comparisons	Cholera treatment—1850s
B) Systematic Trials	
1. Alternate controls	Fibiger, Diphtheria antitoxin—1898
2. Comparison with previous experience of patient	Streptomycin for tuberculosis—1940s
3. Concurrent controls—not randomized	Lind, Scurvy—1747 Semmelweis, Puerperal fever—1848 Goldberger, Pellagra—1914
4. Randomized controls	Diehl, Cold vaccine—1938 Medical Research Council, Streptomycin, tuberculosis—1947

one."[57] This is mentioned to indicate how ideas from one discipline can be transferred to another, a subject of considerable historical interest which should be explored more fully.

By way of summary, the major developments in the evolution of clinical trials are shown in Table 6, classified by certain methodological considerations. The evolutionary path has been a tortuous and lengthy one. Although there are issues that are still being debated, the double blind randomized clinical trial which is the basic condition of *ceteris paribus* has become a part of contemporary medicine. Its more rapid acceptance in the past two decades has probably been influenced largely by external regulatory action and requirements of agencies that fund the research. This part of the story will require separate examination.

[57] E. Greenwood, *Experimental Sociology: A Study in Method* (New York: King's Crown Press, 1945).

WHY A PHYSIOLOGIST?—THE CASE OF HENRY P. BOWDITCH*

W. Bruce Fye

Henry Pickering Bowditch is widely acknowledged as a pivotal figure in the professionalization of American physiology in the nineteenth century.[1] In order to understand why he plays this role, more needs to be known about the circumstances which led Bowditch to devote his life to physiology, rather than following a more traditional pattern for a physician of pursuing his scientific interests as an avocation and sideline to his practice of medicine. This paper will discuss the individual and institutional factors which led Bowditch to choose a career as a professional physiologist and these will be documented by quotations from correspondence between Bowditch and his family members and teachers in the formative years of his career from 1860 to 1870.[2]

Why is it important to study the career of Henry Bowditch as a physiologist? In many respects Bowditch was to American physiology what William H. Welch was to American pathology.[3] They both were the dominant figures in the early years of their respective scientific fields and led the way in the professionalization of these subjects in the United States. Moreover, Bowditch and Welch served as role models for a new type of medical educator—the full-time medical scientist whose main commitments were to teaching and original research. Their successful introduction of the European concept of the full-time medical scientist in their own medical schools would eventually lead to the adoption of this type of instructor for the pre-clinical sciences in all American medical schools. Furthermore, it would provide a model for the development of a related concept, the full-time clinical faculty member.[4]

* Presented at the 54th annual meeting of the American Association for the History of Medicine, Toronto, 11 May 1981. Financial support for research on this topic was provided in part by the Robert Wood Johnson Foundation, Clinical Scholars Program, Johns Hopkins University, 1976–1978.

[1] The standard biographical sketch of Bowditch is Walter B. Cannon, "Henry Pickering Bowditch," *Memoirs Nat. Acad. Sci.*, 1922, *17*: 183–96. See also Frederick W. Ellis, "Henry Pickering Bowditch and the development of the Harvard laboratory of physiology," *New Engl. J. Med.*, 1938, *219*: 819–28. A recent overview of Bowditch's career is Everett Mendelsohn, "Henry Pickering Bowditch," *DSB*, 1973, 2: 365–68.

[2] Much of this paper is based upon correspondence and other archival materials preserved at the Francis A. Countway Library of Medicine, Harvard University, Boston, Massachusetts. Unless otherwise noted, manuscript material cited in this paper is from the Bowditch papers housed at the Countway Library. The assistance of Mr. Richard Wolfe, curator of rare books and manuscripts, and his staff is gratefully acknowledged.

[3] For an interpretive and well documented survey of Welch's career, see Donald Fleming, *William H. Welch and the Rise of Modern Medicine* (Boston: Little, Brown, 1954).

[4] Lewellys Barker, "Medicine and the Universities," *Amer. Med.*, 1902, *4*: 143–47. See also Florence Sabin, "The extension of the full-time plan of teaching to clinical medicine," *Science*, 1922, *56*: 1–18.