OF PLAGUES, BLIGHTS, AND BLOODLETTING: HISTORICAL HIGHLIGHTS OF THE RANDOMIZED CONTROLLED TRIAL

Mary Lou Greenfield, RN, MPH, MS
The University of Michigan
Department of Anesthesiology

Abstract: *This article provides a broad review of the history of research studies and the use of comparative statistical methodology. This knowledge is important in understanding the current clinical research context and in helping to define how to improve clinical research in the future. Sentinel historical developments leading to the randomized clinical trial (RCT) and the key elements of the modern RCT are discussed*

Medical research is characterized by the use of experiments and clinical trials to evaluate prophylactic or therapeutic agents. It represents the value that we place on experimentation in the medical field, and it must meet the standards required by regulatory agencies. According to Meinert (Clinical Trials, 1986) the elements of a clinical trial are: a planned experiment to assess the efficacy of a treatment by comparing outcomes in a group of patients treated with the test treatment with the outcomes observed in a comparison group receiving a control treatment.

The first written account of a comparative study appears in the Old Testament (Book of Daniel, Verses 12-15) (Table 1). The King of Babylon conquered Daniel and the Israelites; he wanted to support the royal families from the Israelites so he wanted to make sure they were well fed. The king wanted the Israelites to eat the food his men ate. This was against the Jewish dietary restrictions, so Daniel suggested an experiment to the king: that his men would eat the food they were used to and at the end of 10 days the king

TABLE 1
Historical Highlights of the Randomized Controlled Trial

Bible: First written account of a comparative study: Book of Daniel (Old Testament, Verses 12-15)
1300 ad: The poet Petrarch used comparative statistical terms in a letter
1662: John Graunt was the first person to work with comparative statistics
1727: James Jurbin did the first comparison of mortality rates for treatment and control groups
1747: James Lind conducted the first experiment with concurrently treated control groups
1787-1872: Pierre Charles Alexander Louis refined comparative studies in mortality and conducted studies involving treatment
1839: William Farr set up a system to routinely compile the number and causes of deaths in England and Wales; he defined exact population at risk, chose appropriate comparison groups, and provided confounding explanations
1855: John Snow became the first epidemiologist; he charted the frequency and distribution of cholera and ascertained a cause of the outbreak
1847: Semmelwies conducted the first prevention study
1860s: Joseph Lister used comparative statistics to study the effects of antisepsis on mortality from amputation
1863: Gull and Sutton used a placebo treatment
1865: Claude Bernard used the word “control”
1898: Johannes Fibiger used the words “alternate controls”
1930: John Wycoff refined the alternate control method
1923: Fisher and MacKenzie used randomization in an agriculture experiment
1931: First use of “clinical trials” in Lancet and British Medical Journal
1924: First use of randomization in a clinical trials by Amberson in a pulmonary tuberculosis trial
1948: First use of “random numbers sampling” for treatment and control allocations by the Medical Research Council
1931: First use of blinding in a clinical trial

could compare how the Israelites looked with how the king's men looked. At the end of 10 days, the Israelites looked better so the king allowed them to continue to eat what they wanted to eat.

Little happened until the middle ages, when thinking in comparative statistical terms began. In 1300 ad, the poet Petrarch wrote a letter to his friend Boccaccio complaining about physicians. He said that if you took 100,000 men of the same age, temperament, habits, and surroundings who had the same disease, and you divided them into two groups, and one group got no medicine and the other group was treated with medicine, the group not treated by the physicians would escape harm or death.

Ambroise Pare, who is considered the father of modern surgery, provided the next written account of comparative studies in 1536 at the Siege of Turin. In attending wounded soldiers, Dr. Pare was using the currently accepted treatment of cauterizing wounds with boiling oil. He ran out of oil and created a poultice composed of egg yolk, rose oil, and turpentine. He demonstrated that this poultice produced a superior outcome to oil.

John Graunt, an English haberdasher who is considered the father of demography, was the first person to work with comparative statistics. Graunt studied death records back to 1532. In 1662 he wrote a treatise called The Nature and Political Observations Made Upon the Bills of Mortality. He developed the life table and presented mortality in terms of survivorship; compared population and mortality rates for counties, ages, sexes, and rural and urban areas; classified death rates according to causes; and looked at overpopulation. Graunt observed that the death rates were higher for urban than for rural dwellers. He observed that the male death rate was higher than the female death rate, and that the male birth rate was higher than the female birth rate.

Between 1723 and 1727, James Jurbin helped settle the debate about the use of vaccinations for small pox during the epidemics that were rampant in England. Through observational studies, he compared the unvaccinated death rate to the vaccinated death rate and found that the rate of death was 1:5-6 for people who were not vaccinated and 1:48-60 for people who were vaccinated.

In 1747, James Lind, a surgeon in the British navy, wrote A Treatise of the Scurvy. Today, we know that scurvy is caused by a Vitamin C deficiency but at the time, they did not know this. Scurvy was characterized as "putrid gums, spots and the lassitude, and weakness of the knees." Sailors suffered the most from scurvy. Dr. Lind was the first person to experiment with concurrently treated control groups, with treatments that were being used at the time for scurvy: vinegar, elixir vitriol gargle, seawater, oranges and lemons, bigness of nutmeg, and cider. He studied 12 men in 6 groups. The group who ate oranges and lemons had the most sudden and best effects.

In 1776, Robert Robertson, another surgeon in the British navy, conducted an inadvertent comparison study. The current practice was to treat fever with bark, which was mixed with water into a brew to make a tea. We know now that the bark contains aspirin. At nine months he ran out of bark and was able to compare the case fatality rate between the two time periods; it was almost five times higher without the bark (0.4% with bark and 2% without bark).

Pierre Charles Alexander Louis (1787-1872) refined comparative studies in mortality and also conducted studies involving treatment. The best example was his work with bloodletting for the treatment of inflammatory diseases, particularly pneumonitis. He found that bloodletting reduced pneumonitis duration in survivors, but it increased overall short-term mortality.

Perkins tractors, long rods made of a special combination of metals, were developed in America in 1799. People believed that if you rolled the rods up and down the affected part of the body, they could cure crippling rheumatism, pain in the joints, gout, wounds, pleurisy, and inflammatory tumors, and could be used to sedate violent cases of insanity. A doctor named Haygarth decided to test this. He used imitation tractors made of wood on five patients with chronic rheumatism the first day, and the metal tractors the second day. He did this without telling the patients, thus conducting the first blinded study. All patients improved "a very little" regardless of treatment.

During the 19th century, the British were instrumental in advancing comparative statistics. They refined comparative vital statistics, continued development of life tables and survival plots, used statistics to justify the hygienic and sanitary reform movement, and began to use statistical methods to describe and investigate other human activities besides mortality rates.

In 1839, William Farr, who was responsible for medical statistics in the Office of the Registrar General for England and Wales, set up a system for routine compilation of the number and causes of deaths. He used the data collected to learn about illness and to evaluate problems that affected public health. Farr looked at mortality patterns among single and married people and workers in different occupations, and drew associations between sea level elevation and cholera deaths, and imprisonment and mortality. He set up good methodology: he defined the exact population at risk, chose appropriate comparison groups, and began to consider alternate explanations for the effects of mortality (e.g., age among workers who were dying in different occupations).

Florence Nightingale, a pioneer in nursing, was a reformer of hospital sanitation methods and the first nurse statistician. She plotted the incidence of preventable deaths in the Crimean War and used new statistical analyses. She:

Innovated the collection, tabulation, interpretation, and graphic display of descriptive statistics to convince people who were in power and authority to reform health care

Demonstrated that social phenomenon can be objectively measured and subjected to mathematical analysis

Proved statistics to be an organized way of learning, which led to improvements in medical and surgical practices.

Nightingale developed a model hospital statistical form to collect consistent data and statistics. She was a fellow of the Royal Statistical Society and an honorary member of the American Statistical Society.

During the cholera epidemic (1853-1854), more than 50 million people in India and Europe died. No one knew the cause of cholera or how it was transmitted. The "miasmatic" (bad smell) theory was popular. People were undernourished, overworked, lived in crowded conditions, and were surrounded by filth and sewage. In London, many people used water that was pumped directly from the Themes River, where raw sewage was dumped.

British physician John Snow, the first epidemiologist, said that the, "most terrible outbreak of cholera which ever occurred... within 250 yards of the spot where Cambridge joins Broad Street, there were upwards of 500 fatal attacks of cholera in 10 days." Farr's data enabled Dr. Snow to formulate and test hypotheses about the cholera epidemic. He felt that the risk of cholera depended on the water supply. Dr. Snow looked at the number of deaths from cholera by pump supply, and observed that the Broad St. pump had 114 deaths per 10,000 houses between 1853 and 1854, a nearby pump had only 37 deaths per 10,000 houses, and the rest of London had 59 deaths per 10,000 houses. He charted the frequency and distribution of cholera and ascertained a cause of the outbreak: the water supply. Dr. Snow studied the distribution, the outcome, and the determinants of disease.

Dr. Semmelwies, a Hungarian physician, who was studying at the Vienna Medical School and Lying-In Hospital in Vienna, conducted the first class prevention study in Austria in 1847. Dr. Semmelwies observed mortality rates from childbed fever in two divisions of the hospital: Division I, which physicians and medical students attended, and Division II, which midwifery students attended. Between 1841 and 1846, he observed that the physicians' division had a 9.92% mortality rate from childbed fever, compared to 3.38% for the midwives' division. He hypothesized that the mortality was related to the new emphasis on the study of pathology at the Vienna Medical School; physicians would work on cadavers and then examine women during labor without washing their hands.

Dr. Semmelwies hypothesized that the contact with cadavers and then examination of women during labor was contributing to higher rates of mortality in the physicians' division. In 1847, he introduced the use of "Chlorina Liquida" for hand washing among physicians and medical students. Between 1847 and 1848, the physicians' division reduced its childbed fever mortality rate one third (to 3.57%).

In the 1860s, Scottish physician Joseph Lister studied the effects of antisepsis on mortality from amputation. From 1864 to 1866, he observed that 16 of 35 amputees died from serious infections. Dr. Lister paid attention to Louis Pastour's findings about how disease is transferred, which was thought to be something in the air. He also learned that a solution called carbolic acid cleared up the decaying process. He made a solution of carbolic acid and water and began to treat amputees with this. From 1867 to 1869, Dr. Lister observed that only 6 of 40 amputees died.

In 1863, Sir William Gull and his colleague Sutton used a placebo treatment. They wanted to study the importance of placebo in assessing the natural history of the course of a disease and the possibility of a spontaneous cure. They treated 44 rheumatic fever patients with mint water. They demonstrated that too much importance was being attached to the use of medicines and that the use of medicines was questionable where there was a natural tendency to get better with "no treatment."

French physician Claude Bernard used the word "control" in 1865 in an Introduction to the Study of Experimental Medicine. Danish physician Johannes Fibiger, who received the Nobel Prize for his work in cancer in 1926, used "alternate controls" in 1898. He did early work in developing a diphtheria serum, injecting the serum into diphtheria patients who were admitted into the clinic every other day. He determined whether this assignment resulted in similar groups.

Early twentieth century efforts built upon the arithmetic observationalists of the eighteenth century. The major interest in medicine was the treatment of infectious diseases, especially diseases of the lungs (colds, pneumonia, flu, tuberculosis), which killed many people. Sometimes a scientific emphasis was lacking and there was great debate about findings.

In the early 1900s, digitalis was a popular treatment for lobar pneumonia. In 1930, English physician John Wycoff studied all the cases for treatment of lobar pneumonia for the past 20 years. He wrote, “The analysis of the hospital records of previous years with the study of digitalis in untreated and treated groups does not constitute a properly controlled series.” Dr. Wycoff refined the alternate control method, using four groups of classifications:

Class 1: no serum, no digitalis
Class B: serum only
Class C: digitalis only
Class D: both serum and digitalis.

He found that the death rate was about 25% higher among digitalis-treated patients than among untreated patients. That landmark study that stopped the use of digitalis for lobar pneumonia.

The first use of “randomization” was in 1923, when Fisher (who developed the Fisher Exact Test and Analysis of Variance) and MacKenzie conducted an analysis of agricultural field experiments. They were trying to plot different areas for planting different crops and said: “If all the plots are undifferentiated as if the numbers had been mixed up and written down in random order...”

Randomization was first used in clinical trials in 1924, in a study by Dr. Amberson of sanocrysin in the treatment of pulmonary tuberculosis in Northville, Michigan. Sanocrysin contained gold, which was through to be helpful in treating tuberculosis. Dr. Amberson divided 24 patients into 2 groups and assigned treatment based on “a flip of a coin.”

Dr. Amberson also used blinding to help prevent staff from knowing which patients received treatment and minimize bias in the outcome of the study. He said, “The patients themselves were not aware of any distinction in the treatment administered.” He also said, “The staff taking care of the patients did not know their treatment assignments.” Dr. Amberson also had independent reviewers select patients for the study. It was an, “... intensive study of smaller, carefully selected, and closely comparable groups of cases, according to a prearranged plan.” The results of this trial were published in 1931.

The Lancet and British Medical Journal were the first to use the term “clinical trials” in 1931, in a report of the Medical Research Council’s Therapeutic Trials Committee. The council was trying to provide the chemical companies and physicians with information on how trials should be designed, considering the economic and social impact of clinical trials.

Rigor in the conduct of research met with much resistance. However, in the 1940s, people began trying to study drugs that were biologically and clinically complex, especially chemotherapeutic agents for tuberculosis. This increased complexity required physicians and people who were supporting research to increase the refinement and rigor of clinical trials, which gradually became accepted in clinical medicine.

Until 1948, most randomization involved flipping a coin. A study of streptomycin in the treatment of tuberculosis that year used “random numbers sampling” to allocate the treatment and control groups.

Conclusion

Humankind has been using experiments and studies to evaluate prophylactic or therapeutic agents throughout history. These studies represent the values that we have placed on such studies in medical science. A broad historical review provides an idea of where we have been so that we understand the context in which we now work, and helps us define where we want to go in the future to improve our clinical research.