American Osler Society, Inc. John P. McGovern Award Lectureship

The Evolution of The Controlled Trial

Sir Richard Doll

John P. McGovern

JOHN P. McGOVERN AWARD LECTURESHIP

Through the generosity of the John P. McGovern Foundation to the American Osler Society, the John P. McGovern Award Lectureship was established in 1986. The lectureship makes possible an annual presentation of a paper dedicated to the general areas of Sir William Osler's interests in the interface between the humanities and the sciences—in particular, medicine, literature, philosophy, and history. The lectureship is awarded to a leader of wide reputation who is selected by a special committee of the Society and is especially significant in that it also stands as a commemoration of Doctor McGovern's own long-standing interest in and contributions to Osleriana.

John P. McGovern Award Lectureships

1. *Our Lords, The Sick* presented by Albert R. Jonsen, Ph.D., April 12, 1986, in San Francisco, California.

2. *To Humane Medicine: Back Door or Front Door?* presented by Edward J. Huth, M.D., April 29, 1987, in Philadelphia, Pennsylvania.

3. *Medicine and the Comic Spirit* presented by Joanne Trautmann Banks, May 3, 1988, in New Orleans, Louisiana.

4. *The 'Open Arms' Reviving: Can we Rekindle the Osler Flame?* presented by Lord Walton, April 26, 1989, in Birmingham, Alabama.

5. *Rx: Hope* presented by E. A. Vastyan, May 8, 1990 in Baltimore, Maryland.

6. *Osler's Gamble and Ours: The Meanings of Contemporary History* presented by Daniel M. Fox, April 10, 1991, in New Orleans, Louisiana.

7. *From Doctor to Nurse with Love In a Molecular Age* presented by William C. Beck, March 26, 1992, in San Diego, California.

8. *The Heroic Physician In Literature: Can The Tradition Continue?* presented by Anne Hudson Jones, May 12, 1993, in Louisville, Kentucky.

9. *'The Leaven of Science': Osler and Medical Research* presented by David Hamilton, May 10, 1994, in London, England.

10. *A Body of Knowledge: Knowledge of the Body* presented by Sherwin B. Nuland, May 10, 1995, in Pittsburgh, Pennsylvania.

11. *Other People's Bodies: Human Experimentation on the 50th Anniversary of the Nuremberg Code* presented by David J. Rothman, April 25, 1996, in San Francisco, California.

12. *The Coming of Compassion* presented by Roger J. Bulger, April 3, 1997, in Williamsburg, Virginia.

13. *Why We Go Back to Hippocrates* presented by Paul Potter, May 6, 1998, in Toronto, Ontario

Cover — Obverse and reverse sides of John P. McGovern Award Lectureship commemorative medal which is presented to each annual lecturer.

John P. McGovern Award Lectureships

14. *Health Care in the Next Millennium* presented by John D. Stobo, M.D., May 5, 1999, in Montreal, Canada.

15. *"Writ Large": Medical History, Medical Anthropology, and Medicine and Literature* presented by Gert H. Brieger, M.D., PH.D., May 17, 2000, in Bethesda, Maryland.

16. *Reflections on American Medical Education* presented by Kenneth M. Ludmerer, M.D., April 18, 2001 in Charleston, South Carolina.

17. *John Shaw Billings as a Historian* presented by James H. Cassedy, Ph.D., April 24, 2002 in Kansas City, Kansas.

18. *The Evolution of The Controlled Trial* presented by Sir Richard Doll, May 23, 2003 in Edinburgh, Scotland.

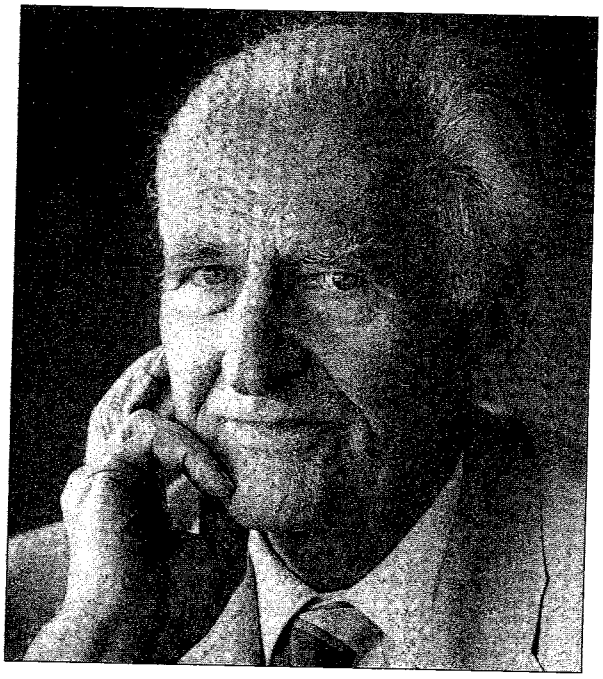
The Eighteenth John P. McGovern Award Lecture

The Evolution of the Controlled Trial

by

Sir Richard Doll

Delivered May 23, 2003
at the Thirty-Third Meeting of the American Osler Society
Edinburgh, Scotland



Sir Richard Doll

Sir Richard Doll

Sir Richard Doll qualified in medicine at London University in 1937, worked as a physician whole time until 1945 and part time until 1969, and served in the Royal Army Medical Corps in France and the Middle East, 1939-45.

He began epidemiological research with Dr Avery Jones in 1946, transferred to the Medical Research Council's Statistical Research Unit under Professor Bradford Hill in 1948, and was director of the unit from 1961-69. From 1969-79 he was Regius Professor of Medicine at the University of Oxford and from 1979-83 Warden of Green College, Oxford. Since 1983 he has been an honorary member of the Imperial Cancer Research Fund's Cancer Studies Unit. He was elected a Fellow of the Royal College of Physicians 1957, Royal Society, 1966, and foreign member of the US National Academy of Sciences 2001.

He has served on many bodies including the UK Medical Research Council, the Scientific Council of the International Agency for Research on Cancer and the World Health Organisation's Advisory Committee on Medical Research and its Committee on Health and the Environment and is currently Chairman of the Management Committee of the UK Childhood Cancer Study and the National Radiological Protection Board's Advisory Committee on Non-ionizing Radiation.

He has received honorary degrees from 13 universities and many awards, including the United Nations Award for Cancer Research 1962, the Charles S Mott Prize for Cancer Research 1979, the Prince Mahidol Award 1992, and the gold medal of the European Cancer Society 2000.

He was knighted in 1971 and made a Companion of Honour in 1996.

His research has included the occupational causes and therapy of peptic ulcer, the long term effects of smoking, ionizing radiation, and oral contraceptives, and the occupational hazard of asbestos.

When, in September last year, I was invited to give the McGovern lecture to the joint meeting of the Osler Societies of America and Japan and the Osler Club of London, I had to accept. For it was not only an honour that I greatly appreciated but it also gave me an opportunity to acknowledge the debt that Green College, Oxford, and the Oxford Medical School both owe to Dr McGovern and Oslerians worldwide. And our debt is great; for it is measured not only in the financial terms that have enabled 13 Norham Gardens to be maintained in part as a memorial to Sir William, housing a collection of Osleriana; but also in intellectual terms, as it has strengthened our capacity for teaching that the practice of medicine is not just the mechanistic process of diagnosis and prescription, but requires a sympathetic ear and an understanding of a patient's concerns. I have, as must be only too obvious, been retired from any responsibility for either the College or the Medical School for many years, but I know I speak for both when I say that the help of Dr McGovern and the continued interest of Oslerians are both greatly appreciated.

Origin of Controlled Trials

The subject that I have chosen for my talk is not one that might appear at first glance to be of great historical interest, but I hope I can persuade you that it is and that those who took the first steps towards the development of what is now an essential element of medical science should be better known than they are. For many doctors, I suspect, the concept of the controlled trial would be dated to the origin of the Medical Research Council's trial of streptomycin in the treatment of pulmonary tuberculosis in 1946, when Bradford Hill persuaded his clinical colleagues to allocate the new treatment to patients by some system of randomisation (Medical Research Council Streptomycin in Tuberculosis Trials Committee, 1948) and that little or nothing has happened since to modify its original form. Both beliefs are incorrect. The practicality, need for, and essential characteristics of the modern controlled trial had, in fact, been debated and even tentatively tested progressively over at least three centuries (Chalmers, 2001) and several major changes have been

made in the last 50 years to the 1946 model. The evolution of the trial has, in truth, been a slow and continuing process, which was speeded up in 1946, just as the evolution of *homo sapiens* that has occurred over several million years may have been qualitatively changed by a mutation in a single gene (protocadherin XY) that facilitated the development of language about 100-150,000 years ago (Sargent *et a*, 2002).

Initially, new treatments were introduced on the basis of trial and error by leaders of the profession who may have acted on theory or adopted folk remedies, and whose improved results, if indeed they were improved, were generally obtained by the displacement of harmful or more harmful remedies or by the psychotherapeutic effect of their personal attention and the treatments were subsequently used on the strength of their teaching.

Van Helmont's Proposal

The first conscious step of which we have record towards ensuring what we should now consider the essential prerequisite of a valid trial—namely, the elimination of bias in the selection of cases for comparison—was taken by John Baptista Van Helmont, a Flemish physician who challenged the orthodoxy of the day in the first half of the 17th century. "If ye speak truth, Oh ye Schooles", he wrote, "that ye can cure any kinde of Fevers without evacuation, but will not fear of a worse relapse; come down to the contest ye Humorists. Let us take out of the Hospitals, out of the Camps, or from elsewhere, 200 or 500 poor people, that have Fevers, Pleurisies, etc. Let us divide them in halves, let us cast lots, that one half of them may fall to my share and the other to yours. I will cure them without bloodletting and sensible evacuation; but do you do as ye know (for neither do I tye you up to the boasting, or of Phlebotomy, or the abstinence from a solutive Medicine) we shall see how many Funerals both of us shall have. But let the reward of the contention, or wager, be 300 Florins, deposited on both sides. Here your business is decided." (Van Helmont, 1662). A fine offer indeed, but there is no record that it was ever taken up. Nor is it clear whether the toss of a coin was to decide who would get each member of a pair of patients or whether this was to be decided by alternation or some other pseudorandom method and the toss of a coin was to decide which group Van Helmont was to treat—so-called cluster randomisation—but it was certainly an impressive proposal that would have eliminated the bugbear of many comparisons that continued to be made into the 20th century: namely, the selection of a group of patients with a relatively good prognosis to test the investigator's favoured method.

Lind and the treatment of scurvy

It was in the UK, however some 100 years later that the realisation of the need to compare like with like began to be taken seriously (Tröhler, 2000) epitomised by Lind's study of the comparative values of the various treatments recommended for scurvy, that was such a terrible hazard of long voyages to distant parts of the world in ships that carried little or no fresh food. When Lind carried out his famous study he had no formal medical qualification (although he did obtain one later in Edinburgh) and had been only an apprentice to a local surgeon before becoming a surgeon in the navy (Tröhler, 2000). In this capacity he had obtained a strong impression of the value of citrus fruits. He was not the first to think that they might be effective—Vasco da Gama's sailors are reported to have demanded oranges on their way back from India 250 years earlier, having seen their beneficial effect when stopping at Mombasa on their way out (see Alvaro Velho and Ravenstein, 1998) but he set out to test his impression scientifically by comparing their effect with that of five other treatments given under similar conditions. In his own words "On the 20th 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 of 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." (Lied, 1753). He then gave five pairs of patients five common treatments for a fortnight and the sixth pair two oranges and a lemon a day for six days before supplies ran out and found that the last pair showed the most improvement, one sailor even becoming fit for duty.

The account of Lind's experiment was published as a treatise in Edinburgh in 1753, but it was 42 years before the Admiralty sanctioned the general use of lemon juice as a prophylactic against scurvy aboard its ships—a year after Lind had died. This is not the place to analyse the reasons for the delay, which were complex and, for the most part, not attributable to administrative incompetence (Trohler, 2000) for I have cited Lind's experiment only as an illustration of the changing attitude to dogmatism and empiricism which marked British medicine in the second half of the 18th century and there were other claims of effective remedies to be taken into account.

Introduction of alternation

In this period an increasing number of doctors in Britain began to undertake numerical analysis of what actually happened to groups of patients diagnosed with different diseases when treated by different methods, rather than detailed accounts of the characteristics and progress of individual patients (Tröhler, 2000). None, I regret to say, had any connection with Oxford, for the enthusiasm for science that led Oxford physicians and natural philosophers to found the Royal Society in the second half of the seventeenth century had waned and the leaders of the movement came, almost to a man, from Edinburgh. Many of them too were military or naval officers. Not having the social status within the medical hierarchy that Louis achieved in Paris a century later (Greenwood, 1936) their work has been largely forgotten and the precedent for insisting on numerical analysis has been attributed to Louis. The records the Edinburgh physicians obtained had, however, revealed the high mortality associated with the customary practice of treating fever by 'anti-inflammatory evacuation' (that is bleeding, giving emetics, and purging) compared with reliance on cordials or the newly introduced cinchona bark, and prepared the way for an experiment in which different treatments were allocated alternatively without any selection by the investigator. The first such study, according to Milne & Chalmers (2001) was a study of the effect of bleeding by Alexander Hamilton and two other Army Surgeons that was conducted in a military hospital in Spain in the course of the war against Napoleon and was reported in Hamilton's doctoral thesis for the University of Edinburgh in 1816. According to Hamilton, "It had been so arranged that this number [that is, 366 sick soldiers] 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 lances. He lost two, I four cases; whilst out of the other third (treated with bloodletting by the third surgeon) thirty five patients died."

Twenty or so years later, according to West (1854) Balfour, another military surgeon with Edinburgh training, carried out a study in which alternation was used to test the efficacy of a claim that very small homeopathic doses of belladonna could prevent scarlet fever, then a serious life-threatening disease in children. "There were" Balfour reported "151 boys of whom I had tolerably satisfactory evidence that they had not had scarlatina. I divided them in two sections taking them alternately from

the list, to prevent the imputation of selection. To the first section (76) I gave belladonna, to the second (75) I gave none; the result was that two in each section were attacked by the disease. The numbers are too small to make any deduction as to the prophylactic power of belladonna, but the observation is good because it shows how we are misled by imperfect observation. Had I given the remedy to all the boys I should probably have attributed to it the cessation of the epidemic." If he had not succeeded in proving, or disproving the efficacy of belladonna as a prophylactic, Balfour had at least demonstrated his own scientific competence in both experimental design and in the interpretation of experimental results.

By the end of the century, alternation had evidently come to be accepted by many leading medical scientists as a standard method of investigation. When, for example, it came to introducing the new anti-toxin treatment for diphtheria, the Danish physician Johannes Fibiger (1898) wrote "In many cases a trustworthy verdict can only be reached when a large number of randomly selected patients are treated with the new remedy and, at the same time an equally large number are treated as usual.... I suggested to Professor Sørensen to treat all patients admitted on the one day with serum, but none of those who were admitted the following day."

Delayed medical response

Despite these good examples, the fact is that as late as the 1930s, when I was a medical student in London, most new treatments were still introduced on the authority of professors or, in the UK, where there were few medical professors, that of eminent consultants who compared the results they obtained in a series of patients with the, almost inevitably, less favourable results that others or they themselves had obtained previously. Sir William relied on such methods himself; but he used them so perceptively that one of his great contributions was to discourage the use of 'nauseous mixtures' and to stress that there was an inherent tendency for many diseases to recover independently of any treatment, as in the majority of cases nature alone was quite competent to restore the patient to health and he was certainly not averse to more formal statistical analysis for, according to Greenwood (1913) he commented that "Karl Pearson's new iatro-mathematical school of medicine has done good work in making the profession more careful about its facts as well as its figures".

Old ways, however, died hard and Bradford Hill, one of Pearson's pupils, still had to lay stress on the importance of concurrent controls in

his series of articles on statistics in medicine that appeared in the Lancet in 1937. These articles, written in simple English, without the use of mathematical formulae, and subsequently republished in book form (Hill, 1937) had an immense influence in the English speaking world on what was, in general, an innumerate profession and gave authority to the use of alternation as an appropriate statistical design.

Randomised Trials

Early studies

By this time, however, some other scientists, stimulated, in all probability by Fisher's work in agricultural science (Fisher, 1925) had already begun to think about the use of randomisation in clinical trials. To Fisher, randomisation was the prerequisite for a scientific estimate of the statistical significance of any difference in the results obtained by two treatments. Emphasizing what he had written previously, he reiterated in his 1935 text that "The purpose of randomisation.....is to guarantee the validity of the test of significance, this test being based on an estimation of error made possible by replication."

The first examples I know of the use of randomisation in clinical trials are two trials of the possible effect of irradiation with ultraviolet light to prevent the common cold, a rather unsatisfactory one by Dora Colebrook (1929) on behalf of the Medical Research Council in the late 1920s and a more satisfactory one by Doull *et* al at Johns Hopkins University a few years later. On the advice of Professor Lowell Reed, Doull *et al (1931*) drew up, in alphabetical order, a list of 373 volunteers from students or members of the staff of the School of Hygiene and Reed then withdrew a die from a bag that contained 122 white die, 66 red die, and 185 black die, which had been thoroughly mixed in a sampling machine. A white die indicated that the next person on the list should be allocated one treatment a week, a red die that he or she should have two treatments a week, and a black die no treatment. The results, in accord with Colebrook's, sadly showed no benefit from the treatment.

A few years later a similar technique was used by Theobald (1937) at St Mary Abbot's Hospital, where there was an obstetric unit affiliated with the London Postgraduate Medical School, to test the ability of a combination of calcium lactate, vitamin A, and vitamin D to prevent toxaemia in pregnancy. This time women presenting at an antenatal clinic withdrew a bead from a box containing equal numbers of blue and white beads and, according to the colour, they either received the special prophylactic treatment or did not receive it. Only 100 women participated

and the number, Theobald thought, was regrettably small; but the results, were nevertheless statistically significant, the women who received the supplements having fewer signs and symptoms suggestive of toxaemia than those who did not. This, he thought, indicated an inadequate level of nutrition, something which was then distressingly common among women in England.*

Meanwhile, another form of randomisation had been reported by Amberson *et al. (1931)*. They divided 24 patients with pulmonary tuberculosis into two equal groups, the members of each group having been "individually matched" in pairs and then tossed a coin to decide which group should be given the newly acclaimed treatment of sodium gold thiosulphate—a procedure reminiscent of what Van Helmont had proposed 260 years before. Their technique of cluster randomisation suffers, however, from two disadvantages: the virtual impossibility of truly being able to match cases and the lack of any means of measuring the relevant random error, because the effective sample size was only two (Armitage, 1982).

Streptomycin for pulmonary tuberculosis

It was not, however, until after the second world war that a trial was carried out that put randomisation on the map and led to its progressive adoption throughout the world for testing the efficacy of new treatments: namely, the Medical Research Council's trial of streptomycin for the treatment of pulmonary tuberculosis, that was designed by Bradford Hill and conducted by Marc Daniels (Medical Research Council's Streptomycin in Tuberculosis Trials Committee, 1948). This was not the first such trial that Hill had designed, for one to test the efficacy of immunization against whooping-cough had been started on similar lines a few months earlier (Medical Research Council Whooping-Cough Immunization Committee, 1951). The results of the vaccine trial were, however, published later and were less dramatic in their impact. The streptomycin trial, which might at first sight have been thought unethical, had been ethical in Britain because the supply of streptomycin was so limited and, what little there

*A further trial that used the toss of a coin to decide which of two treatments individuals should receive was described by Hinshaw & Feldman (1944) but it must be doubted whether it was ever put into effect (Doll *et al., 1999)*.

was, was ear-marked first for conditions that were uniformly fatal: namely, miliary tuberculosis and tuberculous meningitis. The government, strapped for US dollars after the war, was unable to provide more than a small sum to purchase a supply of an expensive foreign drug that could be used for less uniformly fatal conditions. Participation in the trial consequently ensured all the patients of the best treatment currently available with the immediate opportunity of a hospital bed (then in short supply) and a 50% chance of getting in addition a promising new treatment not otherwise available.

To Bradford Hill, the value of randomisation was not, as it was to Fisher, a theoretical justification for the use of statistical tests of significance, but a practicable means of ensuring the absence of bias in the allocation of cases. For he had found, from practical experience, that the use of alternation had at times led to discrimination in deciding whether a patient was eligible for admission, the most seriously ill patients not being admitted to the group that was to receive the investigator's favoured treatment. Randomisation, **after** it had been decided that the patient was eligible for the trial, avoided this pitfall and the avoidance of bias is now considered the basic justification for its use; the statistical advantage over alternation that concerned Fisher being of little practical importance.

Randomisation for clinical trials was not uniformly accepted overnight; some clinicians arguing that it was inefficient, because it did not permit discrimination between patients that could benefit and those that could not. To this Hill (personal communication) used to reply "Tell me the criteria to distinguish such patients and we will build it into the trial." He never had any detailed response to his challenge, but it was always agreed that randomisation could be limited to patients within strata defined by the clinician as having different prognoses—if the clinician could, in fact, define them sufficiently clearly for general medical use. Gradually, however, the use of randomisation spread and it is now so broadly accepted that it has become a necessary element of trials submitted to licensing authorities for the approval of new drugs.

Later developments

Large Trials.

Satisfactory though it was, the streptomycin trial was not the last word in trial design, for there have subsequently been two important developments. First, it has come to be realized that trials may need to be organized on a qualitatively larger scale. Although some of the early tri-

als were collaborative, in that they required the participation of several specialist groups to obtain sufficient patients for the conduct of the trial, the total number of patients in any one trial was seldom more than a few hundred and often less. This was fine, if the benefit of the new treatment was much greater than that of the old, as was the case with the initial trial of streptomycin and has been for some of the trials of cortisone; but it certainly was not when one was dealing with a drug that reduced the fatality rate from say 12% to 10%—an unimportant gain in many instances, but important when the condition was as common as myocardial infarction. For a small gain from the treatment of a common condition may be socially more important than a large gain from the treatment of a rare condition. That it was practicable by international collaboration to carry out a trial on tens of thousands of patients rather than on a few hundred has been demonstrated by the ISIS series of trials of therapy for myocardial infarction (*v.i.*) which have consequently obtained such clear evidence of several modest effects that they have changed medical practice and avoided many thousands of premature deaths worldwide.

Factorial Design.

Such large trials are expensive and time-consuming and there is much to be gained from testing two or more therapies in the one trial. Fortunately this was made possible by the factorial design that began to be introduced in the 1930s (see Doll, 2003) and was illustrated very clearly by Wilson *et al. (1946)* in the mid 1940s when they sought to test simultaneously the separate effects of supplements of cysteine and of reduced dietary fat on the course of infective hepatitis. At that period the standard method used by the scientifically minded to test a new remedy was alternation of new and old therapies and Wilson *et al. (1946)* consequently gave alternate patients a supplement of 5 g cysteine a day, giving alternate patients in each group either a low fat or a high fat diet. When, therefore, the patients given supplementary cysteine were compared with those not given it, each group had had comparable diets, half having had a high fat diet and half a low fat diet and the comparability held with regard to supplementary cysteine when they compared the patients on the two diets. They were thus able, with 103 patients, to show some possible benefit from cysteine in that jaundice, liver enlargement, and biliuria lasted for a shorter time than when it was not given, but that no benefit, judged by the same criteria, was obtained by reducing the fat in the diet.

Six years later, after the introduction of randomisation, I was able to extend Wilson *et al.'s* method to test, in one and the same trial, three ther-

apies for gastric ulcer, by giving successive patients, in random order, one or other of the eight possible combinations of three therapies (a, b and c; a and b; a and c; b and c; a alone; b alone; c alone; or none of them) and showed, as a result, that ulcers healed more quickly if patients were kept in bed in hospital than if they were ambulant at home, but that nothing was gained by medication with ascorbic acid or phenobarbitone (Doll & Pygott, 1952).

The same procedure has now been adapted to large trials but with a better system of randomisation, ensuring that no treating clinician can have any knowledge of what the next combination of treatments is likely to be. Two therapies were, for example, successfully tested in 17,000 patients with myocardial infarcts showing beneficial effects from both aspirin and streptokinase so clearly that in many countries medical practice was quickly altered and both therapies were adopted (ISIS-2 Collaborative Group, 1988) and three therapies were tested a few years later in a trial in 58,000 patients, only one of which (an angiotensin converting enzyme inhibitor) proved to have any detectable benefit (ISIS-4 Collaborative Group, 1995).

The Future

By emphasizing the importance of really large trials to obtain clear and definitive results, I have not wished to imply that these are the acme of controlled trials at which their evolution ends. That trials could be conducted on tens of thousands of patients has been a major development is, I believe, evident; but that is not to say that small trials will not continue to have their place. On the contrary, new remedies will inevitably be tested first in small trials or, at the most, in trials of medium size; but these should be made much more effective by the greater specificity that knowledge of molecular biology should make possible, even so effective that (dare I say it) we may, on some occasions, complete the circle and go back to the use of historical controls.

The introduction of imanitab mesylate, which inhibits the action of the specific tyrosine kinase that is the product of the chromosomal translocation that characterizes chronic myeloid leukaemia, is a case in point. The effect on the progress of the leukaemia in a handful of patients was so dramatic that the drug went straight into clinical use from the trial designed to discover its most appropriate therapeutic dose (Druker *et al.*, 2001) just as the sulphonamides and penicillin had done in the first half of the last century.

Envoi

But leaving the future aside, which I always hesitate to predict, the development that has already occurred is, to my mind, among the most important in the methodology of clinical research in the last 100 years. In this account of it I have restricted myself to the development of the science and, for my knowledge of that, I have been deeply indebted to Sir Iain Chalmers, who has done more than anyone since Bradford Hill to bring home to all sections of the medical profession the need to base their practice on reliable evidence of the balance of benefit and harm from what they do, something that is now enshrined in the work of the worldwide Cochrane Collaboration that Sir Iain initiated. I have consequently said very little about the ethical considerations that necessarily accompany every new step in Medicine, and nothing at all about the growth of regulations that are progressively being attached to the allocation of funds for research. Both need much more thought than they appear to have had, if research in the public interest is to be continued, and their examination is best reserved for another occasion and another speaker with more of Sir William's scholarship, clarity of presentation, and power to persuade than I could possibly hope to achieve.

References

ALVARO VELHO A, RAVENSTEIN RG (1998) *Journal of the first voyage of Vasco da Gamo 1497-1499.* Columbia: South Asia Books.

AMBERSON JB, McMAHON BT, PINNER M. (1931) A clinical trial of sanocrysin in pulmonary tuberculosis. *Amer Rev Tuberc,* **24**, 401-435

ARMITAGE P. (1982) The role of randomisation in clinical trials. *Stats in Med, 1,* 345-352

CHALMERS I. (2001) Comparing like with like: some historical milestones in the evolution of methods to create unbiased comparison groups in therapeutic experiments. *Int J Epidemiol,* **30**, 1156-1164

COLEBROOK D. (1929) *Irradiation and health.* Medical Research Council Special Report Series No.131

DOLL R. (2003) Controlled trials testing two or more treatments. *in press*

DOLL R, PETO R, CLARKE M. (1999) First publication of an individually randomized trial. *Controlled Clinical Trials,* **20**, 367-368

DOLL R, PYGOTT F. (1952) Factors influencing the rate of healing of gastric ulcers: admission to hospital, phenobarbitone, and ascorbic acid. *Lancet,* **1**, 171

DOULL JA, HARDY M, CLARK JH, HERMAN NB. (1931) The effect of irradiation with ultraviolet light on the frequency of attacks of upper respiratory disease (common colds). Am J Hyg, **13**, 460-477

DRUKER BJ, TALPAZ M, RESTA DJ, PENG B, BUCHDUNGER E, FORD JM, LYDON NB, KANTARJIAN H, CAPDEVILLE R, OHNO-JONES S, SAWYERS CL. (2001) Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukaemia. *N Engl J Med,* **344**, 1031 -1037

FIBIGER J. (1898) Om serumbehandlung af Difteri. *Hospitalstidende,* **6**, 309-325 & 337-350.

References

FISHER RA. (1925) *Statistical methods for Research Workers.* Edinburgh, Oliver & Boyd

FISHER RA. (1935) *The design of experiments.* Oliver & Boyd, Edinburgh

GREENWOOD M. (1913) On methods of research available in the study of medical problems, with special reference to Sir Almsroth Wright's recent utterances. *Lancet,* **1,** 164

GREENWOOD M. (1936) *The medical dictator and other biological studies.* Williams & Norgate

HAMILTON AL. (1816) *Dissertatio Medica Inauguralis De Synocho Castrensi.* Edinburgh, J Ballantyne, 1816

HAMILTON D. (1994) The Leaven of Science. Osler and Medical Research. John P McGovern Award Lecture delivered 10 May 1994 at the Twenty-Fourth Annual Meeting of the American Osler Society, London and Oxford

HILL AB. (1937) *Principles of medical statistics.* The Lancet, London

HINSHAW HC, FELDMAN WH. (1944) Evaluation of chemotherapeutic agents in clinical tuberculosis: a suggested procedure. *Amer Rev Tuberc,* **50,** 202-213

ISIS-2 (Second International Study of Infarct Survival) COLLABORATIVE GROUP. (1988) Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet,* **2,** 349-360

ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group. (1995) ISIS-4: A randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. Lancet 1995; 345: 669-685

References

LIND J. (1753). 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.* Edinburgh: Printed by Sands, Murray and Cochran for A Kincaid and A Donaldson, 1753, pp.145-146

MEDICAL RESEARCH COUNCIL STREPTOMYCIN IN TUBERCULOSIS TRIALS COMMITTEE (1948) Streptomycin treatment for pulmonary tuberculosis. *Brit med J*, **2**, 769-782

MEDICAL RESEARCH COUNCIL WHOOPING-COUGH IMMUNIZATION COMMITTEE (1951) The prevention of whooping-cough by vaccination. *Brit med J*, **1**, 1463-1471.

MILNE I, CHALMERS I. (2001) Tackling bias in assessing the effects of health care interventions: early contributions from James Lind, Alexander Lesassier Hamilton, and T. Graham Balfour. *Proceedings of the Royal Society of Edinburgh* **31** *suppl 9, pp. 46-48*

SARGENT CA, BLANCO P, AFFARA NA. (2002) Did the hominid-specific regions of X-Y homology contain candidate genes potentially involved in a critical event linked to speciation? In: *The Speciation of Modem Homo Sapiens.* Ed. T.J. Crow, Oxford University Press, Oxford.

THEOBALD GW. (1937) Effect of calcium and vitamins A & D on incidence of pregnancy toxaemia. *Lancet*, **1** ,1397-1399