

Introductory note to: Bull JP (1951). A study of the history and principles of clinical therapeutic trials. MD Thesis, University of Cambridge.

In 1947, I was appointed to work at the Medical Research Council's Industrial Injuries and Burns Research Unit, based at the Birmingham Accident Hospital. The director of the new Unit, John Squire, like myself, had done some personnel research in the Services during the war. I also had experience with mepacrine control of Malaria in the Tropics and the treatment of an outbreak of Schistosomiasis in West Africa with intravenous Tartar Emetic. We were both keen to develop clinical research at the Accident Hospital to parallel Leonard Colebrook's bacteriological studies on Burns.

The science teaching at Burton Grammar School was excellent and had encouraged my interest in the history of Science. I thought the history of the noting and testing of treatments would be a promising topic for my MD Thesis. No formal supervision was possible. Having already had experience of working on original papers for my Cambridge "Part II", it was not a big step to work up references – using particularly *Index Medicus* at the Royal Society of Medicine Library.

During late 1950s, Everett Evans, director of the burns unit at the Massachusetts General Hospital in Boston, USA, visited our unit in Birmingham. During his visit he noticed my thesis on the history of the clinical trial lying on my desk. In his capacity as editor of the *Journal of Chronic Diseases*, he invited me to prepare an article based on the thesis, and this was published not long after (Bull JP 1959. The historical development of clinical therapeutic trials. *Journal of Chronic Diseases* 10:218-248).

John P Bull
Birmingham, April 2007

**A STUDY OF THE HISTORY AND PRINCIPLES
OF CLINICAL THERAPEUTIC TRIALS**

MD Thesis, University of Cambridge, 1951

John P Bull

**Medical Research Council Industrial Injuries and Burns Research Unit
Birmingham Accident Hospital, Bath Row, Birmingham, 15.**

CONTENTS

	Page
<u>Introduction</u>	3
<u>Historical Review</u>	3
The Ancient World.....	3
Greece and Rome	5
The Middle Ages.....	8
The Renaissance.....	10
The Seventeenth Century	12
The Eighteenth Century	15
The Nineteenth Century	23
The Twentieth Century.....	30
<u>Principles</u>	41
The Basis of Comparison	41
General clinical experience	
Previous special experience	
The same patient	
Concurrent controls outside the trial	
Controlled series	
Criteria for inclusion of cases	48
Criteria of Success	49
The role of Statistics.....	51
Estimation of optimum dosage	59
The Technique of the Controlled Trial	61
The agent	
The disease	
The patient	
The control therapy	
Randomisation	
Investigation facilities	
Organisation of staff	
The Pilot trial	
<u>Discussion</u>	67
<u>Summary</u>	73
<u>References</u>	74
<u>Appendix</u>	

A Study of the History and Principles of Clinical Therapeutic Trials

INTRODUCTION

Deliberate experiments designed to assess upon patients the value of therapeutic procedures are an important feature of modern medicine. These clinical trials link laboratory results with medical practice and so provide a reliable basis for advance in treatment. They are seldom easy, usually costly of time and effort and sometimes dangerous to patients. The danger may be to patients treated in the trial, or to patients badly treated as a result of reliance upon an inadequate trial. The efficiency of these experiments is thus a matter of some importance.

A study of the problem tackled in the past, by what methods and with what success may yield lessons for the future. It may show some of the pitfalls and limitations of the clinical therapeutic experiment as well as its triumph in certain circumstances. The conditions for each success may be seen and, further, the best methods to use when those conditions are present.

With these aims, a review of the history of clinical trials will first be made, followed by an account of the principles exemplified and some suggestions for future developments.

HISTORICAL REVIEW

The Ancient World

The methods of therapeutic investigation used by the ancient Egyptians are unknown. The outstanding lessons from records discovered are first the extreme antiquity of systematic medicine, a papyrus in the British Museum gives prescriptions believed to date from c. 2000 B.C. (British Museum 1930), the time of the building of the great Pyramids, and second the high quality of treatment for certain lesions. Whilst medical conditions were typically treated with a combination of ritual exhortations and fantastic mixtures of herbs and natural products (Bryan 1930) surgical treatment and, in particular, minor

surgery was at a much higher level. The Edwin Smith papyrus (Breasted 1930, Ranke 1933) dating from c.1600 B.C., refers to bandaging and stitching of wounds and, for instance, the treatment recommended for a dislocated jaw by reduction is that of modern surgery. We do not know how such methods of treatment were worked out, but it is interesting to note that the types of lesion so excellently treated were all of simple aetiology, mostly accidental injuries, with mechanisms of production and treatment not dissimilar to those of structural engineering of which the ancient Egyptians were such masters. The success of the treatments would be unequivocal, so that therapy developed from simple mechanical principles could readily be tested by 'trial and error'. The simplicity of such surgical conditions can be contrasted with the complexity of aetiology, diagnosis and criteria of cure of medical conditions likely to have been prevalent in the Nile valley. Even with modern aid the differentiation of tropical fevers is not easy and one of the characteristic Egyptian diseases, schistosomiasis, possibly the A'ia disease of the papyri, has only recently been the subject of satisfactory therapeutic trials.

The surviving records of Babylon and Assyria show a state of medical knowledge similar to that of ancient Egypt (Dawson 1930) and there is much evidence of interchange with Egyptian medicine (Jastrow 1917). An elaborate Pharmacopoeia was used but much of the therapy was irrational or ritualistic. Exorcism of hostile powers was the aim of the treatment but incantations were combined with practical measures such as the use of poultices, purgatives and enemas. No elaborate trial would be necessary to test the effects of these simple treatments and the persistence of the associated irrational ritual suggests a low level of scientific criticism. Not all the ritual, perhaps, should be dismissed; it is known for instance from excavations that public hygiene was highly developed in Babylon and the frequent recommendation of purification by water or fire may indicate that lessons learnt in hygiene were applied to therapeutics.

According to Herodotus it was a custom of the Babylonians to exhibit their sick in a public place so that passers-by might advise suitable treatment based on their experience of similar cases. It seems unlikely that successful treatments would be fully validated on such a system and the records do not suggest that orthodox medicine benefited by these public clinical trials.

As that of other ancient civilisations the medicine of the Hebrews was closely allied to religion. Public hygiene, in which the Jews excelled, was administered by the priests and enforced by religious sanctions. The knowledge of preventive measures against infectious diseases was perhaps a product of life in tribal communities, large enough to have an administrative structure and small enough for each member to be known personally. The pressure of hard living conditions and the ever-present danger of communicable disease would favour the growth of a sound knowledge of hygiene. There is no evidence of any deliberate experiments and little is known of Ancient Hebrew therapeutics as distinct from preventive medicine. It is possible that the lost book of the Wisdom of Japheth which emphasised the value of observation in medicine would have thrown light on methods of advance in treatment (Gordon 1942).

Greece and Rome

Modern medicine is often said to have begun in Greece. The wealth of original observation and cautious deduction recorded in the Hippocratic books is one of the outstanding achievements of mankind. The study by the Coan school of natural history of disease gave rich results in diagnosis and prognosis but there was less success in therapeutics. Surgical treatment, especially of minor complaints, reached a high standard as it had done many centuries before among the Egyptians. In contrast, medical therapeutics were designated by a priori theories. These theories, though less speculative than those of other schools, were over-simple derivations from a general philosophy of nature. The therapy based on them was intended to assist the natural powers of healing by simple exercises and diets. These were rationally deduced from the postulates, but there

is no suggestion that experimental trial and judgment by results was used. Thus the Regimen of Acute Diseases urges physicians to enquire into the best treatments and goes on to give reasons in favour of barley gruel for fevers, but characteristically the gruel is recommended for being smooth and soft and not for its observed effects upon patients, nor is any comparison offered with alternative treatments.

Medicine in contrast to more theoretical subjects has the advantage of enforcing a continual testing of theory by practice. This was recognised by the writer of the Precepts 'One must attend in medical practice not primarily to plausible theories, but to experience combined with reason'. The natural sequel in therapeutics would seem to be the clinical trial which aims to test the 'day to day impressions' and to substitute a deliberate advance for the accumulation of 'plausible theories'. The Hippocratic writers with their confidence in general philosophy did not draw this conclusion. A study of the successors of the Hippocratic school emphasised different aspects of this paradox between a stated reliance upon observation and an actual trust in a priori theory. The Dogmatists developed the theory of humours and qualities into a rigid formalism (Singer 1928). The Empirics on the other hand, discounted theory and relied upon practical tests only. In this way knowledge of drugs was advanced, in particular poisons and their antidotes were widely studied. Attalos and other rulers tested the effect of poisons upon criminals by a kind of reversed 'clinical trial' and the famous antidote of Mithridates comes from this era. Extreme empiricism with its undue emphasis on 'cures' and its neglect of general principles of aetiology and diagnosis led to extravagant polypharmacy and hindered therapeutic progress.

Roman medicine derived directly from that of Greece, many of its leaders being of Greek origin. The arguments of the conflicting schools continued but the outstanding men favoured an eclectic compromise. Celsus, after reviewing the history of Classical medicine, states fairly the Empiric view of the development of therapeutics, '....careful men noted what generally answered the better, and

then begun to prescribe the same for their patients. Thus sprang up the Art of Medicine which from the frequent recovery of some and the death of others, distinguished between the pernicious and the salutary'. Celsus agrees that, '....nothing adds more to a really rational treatment than experience' and concludes that, '.... The Art of Medicine ought to be rational but to draw instruction from evident causes, all obscure ones being rejected from the practice of the Art, although not from the practitioner's study'. Like other early writings De Medicina cites little evidence to support the claims made for its treatments and though these include sound use of some drugs, in particular local application of astringents, many others are only 'sympathetic' remedies and it seems unlikely that either the experience of 'careful men' or consideration of 'evident causes' could justify giving Ox spleen for splenic enlargement or Pole reed for injuries by splinters.

In the first century A.D. Dioscorides wrote the earliest scientific account of medical botany. He recognised natural families of plants and classified their medicinal qualities. His work was the foundation of Herbals for sixteen centuries (Garrison 1922), but did not describe methods of testing therapeutic action. He recommended Mandragora wine for insomnia and as an analgesic draught which could be used in surgical operations (Gunther 1933). Dioscorides was himself a military surgeon and had, presumably, opportunities for trying such analgesic properties. Some other important pharmacologically active plants, for example Digitalis and Atropa are not mentioned and others such as Salix are recommended for a multitude of complaints often not including what is now known to be their most valuable action. It would be unreasonable to expect an exhaustive account in a pioneer work but the absence of these important parts of materia medica may explain some of the impotence of medicine in medieval times when Dioscorides was followed slavishly.

The advances in surgical technique in Roman times were probably related to experience with gladiatorial and military injuries (Garrison 1922). Galen himself

served as physician to the gladiators and became one of the great founders of scientific medicine. His experimental approach was well shown in physiology but, though he tried to deduce therapeutics rationally from knowledge of disease and understanding of remedies, inadequate testing permitted general theories such as that of treatment by contraries to over-ride practical results upon patients.

The Middle Ages

After the fall of the Roman empire the scientific trend in European medicine was arrested. The newly developing Moslem culture took up and preserved medical teaching from classical and early Christian sources. The doctrines of Islam favoured conservatism in therapeutics and, in particular, the objection to touching the human body delayed advances in Anatomy and Surgery (Garrison 1922). The authority of Galen was in general accepted though his teachings were often mixed with astrology. Arabian merchants dominated the spice and drug trade so it is not surprising that there were numerous pharmacological experiments (Mettler 1947). Rhazes (860-932) was outstanding in this field and Avicenna (980-1037) in his encyclopaedic 'Canon' gives some interesting rules for the testing of drugs (Neuburger 1910). He suggests that in the trial of a remedy it should be used in its natural state upon uncomplicated disease, that two opposed cases be observed and that study be made of the time of action and of the reproducibility of the effects. These rules imply a very modern approach but there seems to be no record of their detailed application though it is possible that documents not yet translated may throw more light on this.

In Europe the supernaturalism of medicine of the monastic period was countered in the eleventh and twelfth centuries by the School of Salerno. Here a simple rational therapy was based on direct study of disease. The famous 'Regimen sanitatis' gives sensible advice on hygiene and diet, considerable detail on blood letting but no suggestion of methods of testing remedies (Harington 1607). Roger of Palermo in his 'Practica' (c.1170) recommended sea-weed for Goitre

and Mercury salves for skin diseases and one can only surmise that some kind of simple trial had shown their usefulness (Garrison 1922).

Latin translations of Arabic scholars such as Avicenna became available in the Thirteenth century and, at the same time, the Roman Church which dominated intellectual life was rejuvenated by the Franciscan movement. An outstanding exponent of this medieval renaissance was the Franciscan Roger Bacon (1214-1292). In his 'De erroribus medicorum' (c.1268) he points out inconsistencies in current medical teaching but excuses some of the defects 'for it is exceedingly difficult and dangerous to perform operations on the human body, wherefore it is more difficult to work in that science than in any other. So that physicians are always to be excused since needs must be that they have deficiencies. For the operative and practical science which do their work on insensate bodies can multiply their experiments till they get rid of deficiency and errors, but a physician cannot do this because of the nobility of the material in which he works; for that body demands that no error be made in operating upon it, and so experience (the experimental method) is difficult in medicine. Therefore physicians are to be excused for their defects more than are workers in the sciences'. Bacon favoured appeal to experience and mathematical demonstration but does not appear to have applied these principles to therapeutics since his own recipes given at the end of 'De erroribus' are on traditional lines.

In the Fourteenth and Fifteenth centuries the attempt was made rather to reconcile Aristotelian dialectics with Arabian medicine than to continue the lead of the Thirteenth century theorists by advancing knowledge by experiment. This failure was in keeping with the medieval intellectual atmosphere and its emphasis on Faith, Authority and philosophical idealism (Allbutt 1901). Quite apart from the intellectual controversies, the practitioners of many crafts continued to make sound observations and steady progress. The use of water power was greatly developed, methods of agriculture improved and, with the introduction of the modern type of ship's rudder long sea journeys became

possible (Lilley 1948). Similar progress presumably occurred in therapeutics. Something of this is suggested by Henri de Mondeville (1260- 1320) who defended cleanliness in the management of wounds against the ‘coction’ and laudable pus’ of the Arabian commentators on Galen (Garrison 1922). His claim that ‘Wounds dry much better before suppuration than after it’ implies direct observation and appeal to experience rather than to authority.

The Renaissance

It is not appropriate to discuss here the various aspects of the revival of learning which, beginning in Italy in the 15th and 16th centuries laid the foundation of modern sciences. The renewed study of Latin texts by Petrarch (1300-74) and his successors, of Greek originals by Boccaccio (1313-75) and others, the invention of gunpowder (c.1330) and of printing (1440-50), the fall of Constantinople (1453), the discovery of America (1492) and Magellan’s circumnavigation of the world (1519-22) all played a part in this enormous development of theory and practice (Dampier-Whetham 1930). Two contrasting threads of this story are important in the history of clinical trials. In common with all other branches of thought medicine was affected by the revival of Classical Humanism. Early medical Humanists such as Leonicensus (1428-1524) and Linacre (1460-1524) provided new and accurate translations of Hippocrates and Galen; they and their followers were thus in an excellent position to attack contemporary medical teaching based on indirect annotations of the classics but their criticism of the new empirical medicine could be just as stringent. This empiricism was a continuation of the mixed science and magic of the alchemists. Paracelsus (1493-1541), its great exponent, was a pupil of Leonicensus and from him acquired a lifelong respect for Hippocrates but his public burning (1528) of the works of Galen and Avicenna shows his violent opposition to current medical orthodoxy. He wrote and spoke in the language of the common people and from them collected information on folk medicine (Garrison 1922). Being also an expert alchemist he made experiments in the use of metals in therapy. There are no records of his methods of trial but he says ‘experience has shown that

Mercury is the sovereign and only remedy for the cure of all ulcers tainted with the great pox' (Cumston 1926). The implied scientific approach was not always sustained; Paracelsus also originated of the earliest sympathetic remedies. The response of the Humanists to these new ideas was decidedly hostile. One of them, Fernel (1497-1558) also an admirer of Hippocrates, attacked the use of Mercury not because it was ineffective but because of its empirical origin. Fernel based his own therapeutics on the doctrine of contraries and favoured Guaiac for syphilis (Sherrington 1946).

Meanwhile Leonardo da Vinci (1451-1519) had worked out the theory of the modern scientific experiment. He saw the importance of mathematical demonstration and says 'I shall test experiment before I proceed further, because my intention is to consult experience first and then with reasoning show why such experience is bound to operate in such a way. And this is the true rule by which those who analyse the effects of nature must proceed; and although Nature begins with the cause and ends with the experience, we must follow the opposite course, namely Begin with the experience and by means of it investigate the cause' (Richter 1939). Leonardo did not use these methods in Therapeutics, and was scornful of doctors, probably with justification. The few recipes he gives in the Notebooks are of traditional type (MacCurdy 1938). Ambroise Paré (1510-90) the greatest figure in Renaissance surgery, popularised the revolutionary anatomical teaching of Vesalius (1514-64) and also made one of the earliest reported clinical trials, albeit an unintentional one. In 1537 while serving with the Mareschal de Motegni he was responsible for the treatment of the wounded, after the capture of the castle of Villaine. They were so numerous that, he says, 'at length my oil lacked and I was constrained to apply in its place a digestive made of yolks of eggs, oil of roses and turpentine. That night I could not sleep at any ease, fearing that by lack of cauterisation I would find the wounded upon which I had not used the said oil dead from the poison. I raised myself early to visit them, when beyond my hope I found those to whom I had applied the digestive medicament feeling but little pain, their wounds neither swollen nor

inflamed, and having slept through the night. The others to whom I had applied the boiling oil were feverish with much pain and swelling about their wounds. Then I determined never again to burn thus so cruelly the poor wounded by arquebuses'. (Packard 1925). The other recorded instance of Paré's use of such methods of trial is perhaps less fortunate. One of his later masters, Charles IX had a bezoar stone supposed to be a universal antidote. Paré criticised this claim and suggested it be tried on a convict. A prisoner agreed and instead of being hanged was given poison and the bezoar stone. He died after about seven hours and Paré did an autopsy which confirmed that death was due to corrosive sublimate. Similar tests of antidotes upon criminals were made by other rulers about this time, and in view of the commonness of poisoning such ruthlessness was perhaps understandable.

During the 16th century there was an increasing interest in Natural History, stimulated by the numerous voyages of discovery. There was a corresponding development of systematic Botany and the work of Fuchs (1501-66) and Valerious Cordus (1515-44) provided first hand descriptions and classification of plants (Gibson 1919). This resulted in improvements in Materia Medica and the earliest Pharmacopoeias date from this time. There were also several treatises on purges and it appears that some kind of testing of pharmacological properties was performed; this seems to have aimed at systematic rather than curative therapy.

The Seventeenth Century

The seventeenth century saw great growth of the theory and practice of scientific method. Medicine was less affected by this than were Physics, Chemistry and Biology, and there was little development of therapeutic trials. As in the preceding periods practical men did note and learn from comparative observations. For example in the first expedition to India by the newly formed East India Company in 1600 there were four ships; on one of them only, that of General James Lancaster, was lemon juice provided and this ship was almost

free from scurvy whereas the others were badly affected during the slow voyage. The company was sufficiently far-sighted to supply all its ships with lemon juice for subsequent voyages (Drummond and Wilbraham 1939). This lead was not taken up by medical men for other diseases, in fact it was not until the next century that a corroborative trial of this same treatment for scurvy was made.

Francis Bacon (1561-1626), the great protagonist of the inductive method in science, devoted a section of his *De augmentis scientiarum* (1628) to Medicine. He discussed the dangers of crude empiricism to which doctors were driven by the demand for 'cures' and said of therapeutics, '....this part of phisic which treat of authentic and positive remedies, we note as deficient; but the business of supplying it, is to be undertaken with great judgement, and, as by a committee of physicians chose for that purpose'. How this 'committee' was to proceed Bacon did not specify.

The outstanding contribution to Physiology made by Harvey's 'De motu cordis' (1628) inspired others to attempt over-facile simplifications of medicine based on general principles analogous to those discovered by Newton. A similar result was produced by Boyle's brilliant discoveries which led in the later iatrochemical school to speculation which were often far removed from sound chemical principles (Shryock 1948). There was failure to apply the applauded scientific methods to test the hypotheses; instead reliance was still placed in a priori reasoning. Boyle (1627-91) was himself very sceptical of current medicine though very interested in drugs and diseases. He made numerous suggestions for imitating medicinal waters and cheapening drugs and, in 'The Usefulness of Natural Philosophy' (1663), he says, 'Another way by which the naturalist....may help to lessen the changeableness of cures is by showing there hath not yet been sufficient proof of their having any medicinal virtues at all..' Boyle also recommended experiments on animals and with Wren he performed the early blood transfusions. In 1657 they succeeded in transfusing animal blood into a human subject as is recorded by Pepys but this was hardly intended as

therapy and when extended as such by Denis in France it was soon discredited (Brown 1948).

An outstanding new therapeutic agent introduced to Europe about this time was Cinchona bark. The precise circumstances of its discovery are still disputed but it is clear that it has been used for many years previously by the Peruvian Indians before it came to the attention of Spanish colonisers about 1630. Its success in the treatment of fevers must have impressed the Jesuit missionaries for they brought the bark to Europe in 1632. It had a hostile reception by orthodox medicine. This was partly due to suspicion of the Jesuits but a more solid objection was that its acceptance meant an overthrow of Galenical theory. As Ramazzini later observed, cinchona did for medicine what gunpowder had done for war (Garrison 1922). No careful trials of the value of this drug seem to have been made and there is even doubt as to the nature of some of the bark used since apart from fraudulent preparations, there seems to have been confusion between cinchona with its anti-malarial action and quina-quina, the source of Peruvian balsam (Haggis 1941). Jesuits bark, discounted by doctors, became the property of quacks. One of these, Talbor (1641-81), achieved great fame with his remedy for fevers and became Physician to Charles II (Scott 1939). Sydenham (1624-89) broke away from orthodoxy and, though a nihilist in many fields of therapy, used and recommended Peruvian bark for intermittent fevers. His methods of trial are unfortunately not known, but his scepticism of the newly developing sciences of Anatomy and Physiology (Payne 1890) make it unlikely that his techniques were derived from the contemporary scientific trends of the Royal Society. His reliance upon clinical experience contrasts with Boyle's confidence in experiment. The marriage of these two approaches to produce a satisfactory clinical trial did not occur in the seventeenth century, though it appears that Locke (1632-1704), a friend of both Boyle and Sydenham, had something of this sort in mind in his projected work 'Ars medica' on the philosophy of medicine (Osler 1900). Therapeutics, instead, were at the mercy of fashion and quackery. The status of Antimony provides an example; it had first been publicised at the

beginning of the century by extravagant claims made in 'The Triumphal Chariot of Antimony'; this book was probably written by Thold but was attributed to a fifteenth century monk, 'Basil Valentine'. By the middle of the century Antimony had fallen into disrepute but in 1657 it was used in treating an illness of Louis XIV. The king recovered and Antimony again became popular for the treatment of fevers (Haggard 1932).

By contrast surgical treatment made more steady progress. Wiseman (1622-76), the outstanding English surgeon of the century, accepted as self-evident the virtues of bleeding, purging, vomiting, sweating, and salivation but nevertheless he made acute observations on simple therapeutic advances. He describes vividly successful treatment of oedema of the legs with laced stockings, the recurrence of the trouble when treatment was interrupted and the further improvement on replacing the appliance (1676). This kind of simple trial with limited mechanical aim has frequently contributed to practical surgical techniques.

The Eighteenth Century

In the eighteenth century the new advances in scientific knowledge and method began to be applied to therapeutics. This was achieved often in the face of opposition from the orthodox theorists who still tried to impose a priori systems upon medical treatment. It is not surprising therefore to find that the advances were made sporadically by persons of independent mind often working in the provinces away from the fashionable systematists. In default of a scientific background new treatments were borrowed from folk medicine but the need of adequate trials of their effects was increasingly realised.

Inoculation as a preventative of smallpox was introduced from Constantinople by Maitland (1668-1748) and Lady Mary Wortley-Montague. They persuaded King George I to permit a trial upon six Newgate convicts in 1721 (Maitland 1722). All survived the operation and were released, one in whom the inoculation failed was found to have had smallpox before, another was exposed

to infection after the treatment and was found to be immune. The results were thought to be conclusively in favour of the inoculation and it became widely practised. The trial has since been criticised because it was uncertain which of the subjects had previously suffered from the disease and the trial was designed to test the safety of the procedure rather than its effectiveness (Creighton 1894). Subsequent experience soon showed that inoculation could have a considerable mortality and that complete protection could not be assured. These variable results were obtained with different modifications of technique which the early trial could not have been expected to test but it would appear that the numbers treated and the precision of the trial were inadequate to give a fair picture of the effects of the operation.

Other treatments continued to be proposed on very slender evidence. Dover published his 'Ancient Physician's Legacy to his Country' in 1733 and claimed extravagant cures of Gout, Dropsy and Diabetes and in the same year Bradley published a rejoinder accusing Dover of quackery; neither performed careful trials of the recommended treatments. Cures for the stone were in great demand and a Mrs. Stephens was given a Parliamentary grant of £5,000 for her secret remedy. The very able experimental physiologist Stephen Hales (1677-1761) was on the committee which examined the claims for the treatment. Patients supposedly cured were questioned and the grant was supported. Hales may have had some misgivings since he later experimented with the mixture and decided that only the 'soap lees and lime of eggshells' which it contained among many other ingredients could have had any solvent effect upon calculi (1740); he does not seem to have considered the necessity of careful trials upon patients.

Herberden (1710-1801) was critical of many traditional treatments and in his 'Antithriaka' (1745) he dealt the final blow to the spurious antidote of Mithridates famous in various forms for two thousand years. He based his attack on the variable constitution and the inconsistencies and unreasonableness of the supposed properties of Theriac rather than upon a demonstration of its ineffectiveness. Another popular remedy was publicised by Bishop Berkeley in

‘Siria’: A chain of philosophical reflexions and Inquiries concerning the virtues of Tar Water’ (1744). From anecdotal evidence of Tar water curing Smallpox and Scurvy he goes on to quote his own use of it in the Irish epidemics; the transition is then made through Chemistry and Physics to Philosophy and Theology. Though no cooperative results are given to substantiate the claims, tar water became so popular that a special warehouse was opened in London for its distribution (Clark-Kennedy 1929). Hales also examined this remedy and cautiously attributed discrepancies in its effects to variations in method of preparation (1747).

At this time an investigation of quite different quality was made by a ship’s surgeon, James Lind (1716-94). Appalled by the ravages of scurvy in Anson’s recent circumnavigation of the world when three-quarters of the men died from the disease, Lind planned a comparative trial of the most promising scurvy ‘cures’. On the 20th May, 1747, he says, ‘I took twelve patients in the scurvy, on board the Salisbury at sea. The cases were as similar as I could have them....they lay together in one place.... and had one diet common to all.’ To two of them was given a quart of cyder a day, to two an elixir of vitriol, to two vinegar, to two oranges and lemons and to the remaining two ‘an electuary recommended by an hospital surgeon’. ‘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 other....was appointed nurse to the rest of the sick’. Apart from the cyder, which seemed to do a little good, the other remedies were ineffective. In spite of this apparently conclusive demonstration Lind (1753) himself continued to cling to other theories; in discussing the treatment of scurvy later in the book he recommends fruit and vegetables but he gives priority to ‘pure dry air’ and says again ‘hence the first step....is change of air’. Others seemed no less doubtful and there were several suggestions for cheap alternatives supposedly equal to fruit. A conclusive trial of fruit as an anti-scorbutic is sometimes attributed to Captain Cook who used it on his second voyage in 1772; but this cannot be sustained since many different prophylactics

were used, amongst the beer-wort of which Cook wrote to Sir John Pringle that it was 'one of the best anti-scorbutic sea-medicines yet found out'. (Cook 1776). The British Navy did not supply lemon juice to its ships until 1795.

A close association existed at this time between general science, exploration and medical research. Pringle was one of the founders of preventive medicine and, when awarding Cook the Royal Society's Medal, praised him as much for his achievement in preventing scurvy as for his geographical discoveries. Sir Joseph Banks, who succeeded Pringle as President of the Royal Society, had himself accompanied Cook on his first voyage round the world and became a great supporter of scientific work; among his many correspondents were John Hunter and Edward Jenner. Another feature of the intellectual life of the Eighteenth century was the rise of medical and scientific societies where the members met to discuss a wide range of subjects. The Lunar Society in Birmingham included the leaders of Science, Engineering and Medicine of that rapidly growing industrial area. Contact was maintained with continental and American workers and in this atmosphere of scientific triumphs it is not surprising that along with genuine advances less authentic ones were introduced. Many experiments were made on static electricity and electric shock treatment became popular under the name of 'Franklinism'. How little this name was justified was shown by Franklin's letter to Pringle in 1757 describing his tests of electricity upon paralysed patients who came to request it. The letter recounts the temporary improvement which Franklin cautiously suggests may have been due to the exercise of the journey in coming for treatment or the 'spirits given by the hope of success'.

Similar evidence of popular and even medical credulity was shown by the success of Perkins' Tractors. These were metallic rods which were supposed by some electrical influence to cure a great variety of diseases. The treatment was recommended by several distinguished doctors, including Nathan Smith the founder of Yale Medical School, and an 'Institute of Perkinism' was founded in

London (Haggard 1932). Haygarth (1740-1827) of Bath submitted the method to a devastating clinical trial described in his book 'Of the Imagination as a cause and as a cure of disorders of the Body' (1801). On five patients he used imitation 'tractors' made of wood and all but one of the patients were relieved. The following day he repeated the treatment using instead a pair of genuine tractors and obtained identical results. Haygarth aptly quotes Lind's comment on the fictitious Scurvy remedies used at the siege of Breda 'an important lesson in physic is here to be learnt, viz. the wonderful and powerful influence of the passions of the mind upon the state and disorders of the body. This is too often overlooked in the cure of diseases...'

Several new medical Schools were founded in the Eighteenth century and another important development was in hospital construction. This, after having 'approached perfection in the fifteenth century' (Garrison 1922) had been seriously neglected in the sixteenth and seventeenth. The number of hospitals actually diminished as a result of the disbanding of religious houses in the Reformation and little was done to replace the loss in spite of increasing populations. In the eighteenth century the position began to be corrected by the founding of new hospitals both in London and in the growing provincial cities. William Withering (1741-99) became Physician to the Birmingham General Hospital at its foundation in 1778. Here were more patients upon which to continue his trials of the preparations of digitalis first brought to his notice as a constituent of a folk remedy for dropsy. In his 'Account of the Foxglove' (1785) he gives his experience of the drug in 163 cases. There is a masterly description of his establishment of correct dosage. At first, he thought it necessary 'to bring on and continue vomiting', then he persisted until nausea was caused, later he aimed at either diuresis, sickness or purging, but having noticed slowing of the pulse in some cases his final method was not to repeat the drug too quickly and to continue 'until it either acts on the kidneys, the stomach, the pulse or the bowels'. Withering's achievements in deciding the types of patients who would benefit from digitalis are equally remarkable when it is remembered that

virtually nothing was then known of the pathology of different kinds of oedema. After describing cases apparently with hypertension and malignant ascites as showing little improvement he says, 'on the contrary if the pulse be feeble or intermitting, the countenance pale, the lips livid, the skin cold, the swollen belly soft and fluctuating, or the anasarcaous limbs readily pitting under the pressure of the finger, we may expect the diuretic effects to follow in a kindly manner'. In assessing the value of this treatment Withering relied upon close observation of his patients using all the appropriate clinical methods then available comparing the results with the same patients' previous condition and sometimes with their relapses on discontinuing the drug; this technique is particularly appropriate for such a problem.

The originality of Withering's discovery was disputed by his fellow member of the Lunar Society, Erasmus Darwin (Roddie 1936), but Darwin's sporadic use of the drug in apparently excessive dosage (Pearson 1930) bears no comparison with the other's systematic study of a long series of patients. Another acquaintance of Withering's was Thomas Fowler who had succeeded him at Stafford Hospital (founded 1772). Here Fowler made his study of arsenic solutions in the treatment of a variety of diseases. In his 'Medical Reports of the effects of Arsenic' (1786) he gives a good account of toxic effects and claims that arsenic is a possible substitute for Peruvian bark in the treatment of ague, of which disease he quotes 247 cases treated by his method. He claims to have cured two-thirds of these patients, but it appears that his methods of observation were unequal to the difficult task of assessing a remedy for such a notoriously intermittent and relapsing disease, which even with the modern advantages of microscopic diagnosis has only recently been the subject of satisfactory clinical trials.

The history of Surgery in this period is dominated by the work of John Hunter (1728-93), and it was he who established modern surgery on a scientific basis. There is an incident in his early days as a military surgeon during the Belle Isle

campaign (1760-1) which is reminiscent of the experience of Paré two centuries earlier. It was current practice to search for and if possible remove the missiles of gunshot wounds. Hunter describes five cases in which this was not done and explains that the 'neglect rather arose from accident than design'. The patients were Frenchmen who had hidden since being wounded and, with only superficial dressings, they all recovered. Hunter modified his treatment of gunshot wounds accordingly and writes, 'This practice has arisen from experience, for it was found that balls when obliged to be left, seldom or ever did any harm when at rest and when not in a vital part' (1793). As with Paré and the cauterisation of wounds the necessities of military surgery forced this unintentional trial upon Hunter. Being the outstanding men they were, they both drew the correct lessons from these experiences.

At the end of the Eighteenth century Jenner published his famous studies on Vaccination (1797-8). This again was a procedure derived from folk medicine having been suggested by the belief among country people that infection with Cowpox prevented subsequent attack by Smallpox. Jenner's original account cites fourteen persons who, having had Cowpox, did not take Smallpox when inoculated subsequently. In a further ten patients artificial infection with Cowpox is described, and four of these did not take Smallpox when inoculated. A reasonable *prima facie* case was thus provided in favour of vaccination but the trial was not so conclusive as has sometimes been claimed. The method was the uncertain one of arm-to-arm infection, the numbers in whom protection was demonstrated were few, this protection was against artificial inoculation and not the naturally occurring disease, and no account was taken of the possibility of natural or previously acquired immunity. George Pearson in 1798 published another smaller trial in which he made a detailed study of the results of inoculating five persons with Smallpox, of whom three had previously had Cowpox and two had not. The results confirmed the protection given by Cowpox but Pearson was guarded in his conclusions and recommended 'well-directed observation in a thousand cases of inoculated Cowpox' (1798). Studies of this

magnitude were soon available since Smallpox and inoculation hospitals provided ample material (e.g., Woodville 1799), but the conditions of observation and the techniques whereby mixed infections of Cowpox and Smallpox were transmitted prevented clear and consistent conclusions to be drawn. A smaller trial, the counter-part of Pearson's was made by Waterhouse (1800), the first doctor to use the method in America. He vaccinated nineteen boys, twelve of these he afterwards inoculated with Smallpox, also two others who had not been vaccinated. These two only took the smallpox infection (Haggard 1932).

A study of the relative merits of different treatments for syphilis was made by John Pearson. He dedicated his book to Thomas Fowler and in it compared the claims and his experiences of various herbal and chemical remedies 'to ascertain whether any other substance than Mercury be a true and certain antidote' (1800). As a surgeon to the London Lock Hospital he had a wide knowledge of the disease and he tried out, apparently not very systematically, any likely remedies. He gives the details of thirty-one patients in support of his opinion that guaiac, China root, sarsaparilla and other treatments recommended as alternatives were only of value when used in addition to mercury. As to mercury itself, he states that its effectiveness was demonstrated in 'not less than twenty thousand cases' of which he had personal experience. He was not blind to its disadvantages and looked forward to further discoveries since 'it were highly desirable to acquire a medicine equally potent as an antivenereal, and not possessing certain active properties peculiar to that mineral'. This study, based on simple clinical assessment of results, seems to have been highly successful, particularly so since there was at that time little understanding of the pathology of the disease nor were there accurate tests of cure and, as Pearson was aware, symptoms due to syphilis were frequently confused with those of excessive treatment with mercury.

The Nineteenth Century

Early in the nineteenth century a further emphasis began to be placed on the need for a careful statistical approach in the evaluation of remedies. This was chiefly directed to criticism of extravagant claims. Cobbett, for instance, in his pamphlet 'The Rush Light' (1800) threw doubt upon Rush's evidence for the value of bleeding and purging in Yellow Fever, and was one of the first to appeal for the application of statistics to such problems (Shryock 1948). Theoretical statistics was also advancing rapidly; the *Théorie Analytique de la Probabilité* appeared in 1810 and in it Laplace, after reviewing the application of statistical methods in many fields, says, 'La même analyse peut-être étendue aux divers résultats de la Médecine'. This approach was developed by P. C. A. Louis (1787-1872) chiefly to establish the diagnostic features and the natural history of diseases such as Typhoid Fever and Phthisis. In his 'Essay on Clinical Instruction' (1834) he lays down admirable rules for the use of his 'Numerical Method' in the assessment of therapy. 'As to different methods of treatment, if it is possible for us to assure ourselves of the superiority of one or other among them in any disease whatever, having regard to the different circumstances of age, sex and temperament, of strength and weakness, it is doubtless to be done by enquiring if under these circumstances a greater number of individuals have been cured by one means than another. Here again it is necessary to count. And it is, in great part at least, because hitherto this method has been not at all, or rarely employed, that the science of therapeutics is still so uncertain; that when the application of the means placed in our hands is useful we do not know the bounds of this utility'. Louis goes on to detail some of the necessary precautions....' 'in order that the calculation may lead to useful or true results it is not sufficient to take account of the modifying powers of the individual; it is also necessary to know with precision at what period of the disease the treatment has been commenced; and especially we ought to know the natural progress of the disease, in all its degrees, when it is abandoned to itself, and whether the subjects have or have not committed errors of regimen; with other particulars'. The method was not likely to be easy; 'The only reproach which can be made to the Numerical Method...is

that it offers real difficulties in its execution. For, on the one hand, it neither can, nor ought to be applied to any other than exact observations, and these are not common; and on the other hand, this method requires much more labour and time than the cost distinguished members of our profession can dedicate to it. But what signifies this reproach, except that the research of truth requires much labour, and its beset with difficulty'. (1834). The most famous example of Louis' use of this method is in his 'Recherches sur les effets de la Saignée' (1835). He studied the effects of bleeding upon 78 cases of pneumonia, 33 causes of erysipelas and 23 cases of inflammation of the throat; he found no appreciable difference in mortality or in duration or type of symptoms or signs between patients bled and not bled and between patients bled at different stages of the disease. This result, which was contrary to orthodox teaching of the time caused an uproar among French physicians, but it dealt a fatal blow to the 'depletive' treatment then in vogue.

A quite different approach to research in therapeutics was also developed in the early nineteenth century. This was based on tests in animals and its great exponent was Magendie (1783-1855). Similar methods had been used sporadically for many years but now many new chemical substances were being prepared and many alkaloids identified, particularly by French scientists. Magendie and his co-workers tested these substances upon animals to investigate their toxicities and pharmacological actions. In this way halogen compounds, strychnine, emetine, quinine and morphine among other drugs were introduced to medicine. Magendie, in his preface to the 'Formulaire' (1821) states 'la manière d'agir des médicaments et les poisons est la même sur l'homme et sur les animaux' and, with confidence in this belief, trials upon man were limited to confirmatory tests upon healthy persons of the pharmacological actions observed in animals. Such an approach is highly appropriate for drugs used to obtain a physiological response or for symptomatic treatment but it is inadequate for the assessment of curative value. There is a great difference between establishing that emetine has an emetic action and assessing its therapeutic value in

dysentery. For most drugs pharmacological investigation is a valuable and perhaps essential prerequisite for a clinical trial but it cannot be a satisfactory substitute.

The rapidly increasing knowledge of chemistry, physiology and pharmacology made the need for good clinical trials increasingly evident. Several attempts were made to clarify the requirements of a satisfactory trial. An American physician, Elisha Bartlett (1804-55) in 'An Essay on the Philosophy of Medical Science' (1844) which he dedicated to P. C. A. Louis, said that in therapeutic investigations, cases which are to be compared must have equal disturbing factors of location, social class and the like, they should be susceptible of a clear and positive diagnosis, there must be no selection of cases and the method of treatment must be clearly defined. The certainty of results, he said, will be 'in proportion to the fixed and uniform character of the compared facts and to the greatness of their numbers'. Bartlett was well aware of the dangers of unbalanced use of the method. Countering the accusation that it tended to therapeutic nihilism, he points out that certain orthodox procedures though they had received no formal trial had been 'established by a series of observations of such vast extent as to compensate in a good degree for the absence of the other conditions'. It is interesting to note that in this category along with quinine for intermittent fever, opium for colic and calomel for syphilis he includes bleeding for acute pleurisy, though he warns that 'even in these cases it is only by a faithful adherence to the rules and methods which have been described that the exact value of the several remedies can be ascertained'. Bartlett dealt also with the danger of losing sight of the individual in the statistical group, 'No acquaintance, however perfect, with the laws of pathology and therapeutics, can ever remove, or in any degree diminish, the necessity of a thorough and discriminating study and knowledge of the single instances which unite to make up the materials of the law'.

The influence of Louis was particularly important in the development of scientific medicine in America (Osler 1908); one of his admirers, Oliver Wendell Holmes (1809-94) in a lecture given in 1860 analysed the errors which led to over-medication. He cited incapacity for sound observation, inability to weigh evidence, counting of only favourable cases, the 'post hoc ergo propter hoc' fallacy and false induction from genuine facts with failure of correction by experience. Holmes further laid part of the blame on the public which 'insists on being poisoned'.

Clinical trials on the lines of Louis' suggestions were made in Great Britain by Bennett in Edinburgh and Sutton in London. Bennett in 1865 reported his experience over twenty years of 129 cases of pneumonia treated on 'restorative' principles. He analysed his patients carefully in respect of sex, age, severity, mortality and duration of illness and was able to compare his results favourably with other series treated by more heroic methods. He also showed that his treatment gave good results in the hands of others and concluded that bleeding and the use of Tartar Emetic should be abandoned. Also in 1865 Sutton published a trial made upon Rheumatic fever patients under the care of Sir William Gull. At the time claims were made for a multitude of treatments for this condition and the trial consisted of the careful observation of twenty cases receiving only Mint-water. The results demonstrated the great natural variation in the disease and the marked tendency to spontaneous cure. Sutton concludes, 'A perusal of the above cases tends to show that the best treatment for Rheumatic fever has still to be determined, and will also convince the reader....that it is absolutely necessary to understand the natural progress of the disease before any conclusion can be arrived at concerning the operation of remedies. The cases show that too much importance has been attached to the use of medicines, especially in those acute cases where the tendency to a natural cure is the greatest'.

Parallel with these developments in Medicine, Surgery was making enormous strides since the introduction of general anaesthetics (1842-7). The precise priorities of the discoveries are still disputed but it is clear that all the early trials of anaesthetics by Long, Wells, Morton and Simpson were made on very few patients (Duncan 1947). The unconsciousness of the patient and his subsequent recovery were sufficiently evident not to demand formal control cases nor large series. Some of the subtler details did require more careful study. As it happened none of the early agents had marked long term ill effects nor was individual variation of great importance so that the small numbers and short period of study of the early trials did not lead to error on these scores. Research into the best methods of administration was made by Snow (1813-58), who in 1858 published 'On Chloroform and other Anaesthetics'. Here he summarised a wealth of observations both experimental and clinical which could only have been made by someone who, like Snow, combined a scientific approach with the great experience in one field made possible by specialism.

The other great development of Nineteenth century Surgery was in antiseptic and aseptic methods. Lister (1827-1912) introduced his new techniques initially as rational procedures to avoid sepsis and supported his arguments by the evidence of a few carefully followed cases. In 1870 he published his statistics for amputation operations comparing 35 cases before the use of antiseptics with 40 cases treated by the new method. He showed a mortality of 43% in the former and 15% in the latter but was diffident about drawing conclusions saying, 'These numbers are, no doubt, too small for a satisfactory statistical comparison...'. A modern comment might be that the numbers are not at fault since the X^2 test shows them to be highly significant; what is more open to question is the adequacy of the comparison with previous experience since so many relevant features such as selection of cases for operation must have also changed. Had it been possible, a careful comparative trial of rival methods might have prevented the bitter and profitless controversy which raged for many years on the subject of the importance and technique of prevention of infection at operation.

The increasing knowledge of bacteriology led to many new applications of immunity to therapeutics. The earliest trial was of Pasteur's vaccine for the prophylaxis of anthrax in animals. The experiment is vividly described by the 'Times' correspondent (1881); sixty sheep were used, twenty-five were inoculated and infected, twenty-five were infected only, and ten were neither inoculated nor infected. Care was taken that the infecting virus was given in equal doses, the sceptical Colin shaking the phial himself and the injections being alternated between protected and unprotected animals at the suggestion of other critical observers (Vallery-Radot 1922). The results could hardly have been more conclusive; of the infected animals all the uninoculated died and all the inoculated survived; and the discovery was immediately hailed as a triumph and applied practically.

Pasteur's method for immunisation against Rabies was more difficult to validate conclusively. It was recognised that the disease was fatal once a person was infected, and Pasteur was able to show that his vaccines gave recovery in patients believed to have been infected. The claim for the effectiveness of the injections was doubtless correct but it has subsequently been very difficult to appraise modifications of Pasteur's technique since treatment cannot morally be withheld though evidence of genuine infection is often scanty (Harvey and Acton 1922).

Increasing knowledge of metabolism and nutrition was followed by corresponding advances in therapy. One of the earliest dietetic trials was made by Takaki in the Japanese Navy. Study of official records showed him that until 1883 Beri-beri accounted for one-half of the fatalities and much of the invaliding. After making a review of the possible causes in this same way as Lind 150 years earlier he decided that the food must be at fault. A more generous diet was introduced in 1884 and from then Beri-beri began to decline and was

eliminated by 1886 (Takaki 1906) This was achieved in ignorance of Vitamins and success was attributed rather to increased protein allowances.

In 1891, treatment of myxoedema by thyroid extract was introduced by Murray. Kocher had described the development of myxoedema in man following extirpation of the Thyroid gland and Horsley had produced similar results experimentally in monkeys; Murray's confidence in the result of treating one case is thus understandable and was fully supported by subsequent experience.

The continued rapid growth of Bacteriology at the end of the nineteenth century led to the preparation of a therapeutic serum for Diphtheria. Behring, Boer and Kossel (1893) described the early trials of this, quoting thirty cases treated of whom six died and comparing this result with the then usual mortality of about 50% and with the previous year's experience at the same hospital when eleven died of thirty-two patients. The authors said that more cases would be required to prove the value of the new treatment. Statistically their figures are significant had the cases been comparable throughout.

The following year the rival French school published a study of three hundred diphtheria patients treated by serum (Roux, Martin and Chaillou 1894). A mortality of 26% was observed after various corrections for anomalous cases, and this was compared with a 51% mortality during the previous four years at the same hospital and with a mortality of 60% in concurrent cases at another hospital. The series were compared as far as possible in respect of type of illness and severity and it was shown that the serum-treated patients had a lower mortality in each group.

An attempt to eliminate the uncertainties of inadequately controlled trials of this treatment was made by Fibiger (1898) in Denmark. He used the serum for alternate cases of diphtheria and analysed the 'treated' and 'untreated' series for comparability in age, symptoms and severity. He then compared the results,

within age and severity groups, for mortality, extent of diphtheritic membrane, pyrexia, paralysis and albuminuria. He was able to show that the mortality was less among the serum-treated, that the membrane disappeared sooner and that pyrexia, albuminuria, and paralysis was unaltered. This excellent trial was less valuable than it otherwise might have been because though adequate numbers were studied, the disease at the time happened to be very mild, the overall mortality in 465 cases studied being only 8%.

Other studies of immunity at the end of the nineteenth century led to the introduction of prophylactic inoculation for typhoid fever. Wright (1900) reported the comparative statistics of incidence of the disease in various units of the Army in India. Of more than eleven thousand men about three thousand had received inoculations; less than 1% of these developed typhoid. Among the remaining eight thousand unprotected men the incidence was 2.5%. The circumstances of this trial prevented randomisation of the prophylaxis among the men and the possibility of differences in susceptibility and exposure cannot therefore be excluded. A more evident difficulty arose from the necessary reliance upon many different observers to diagnose the disease and record its presence or absence. These observers were army doctors who would inevitably vary in skill and in interest in the trial. While the results clearly suggest a degree of protection achieved by the inoculation they cannot be considered conclusive.

The twentieth century

The enormous development of organic chemistry in parallel with the previously mentioned growth of bacteriology laid the foundation for chemotherapy, the greatest contribution of our age to medical treatment. At the turn of the century Ehrlich began his search for trypanocidal Agents. His success with arsphenamine in vitro led to its trial first for syphilis in the rabbit, then for the human disease. From the animal experiments Ehrlich believed he had achieved his aim of a 'magna therapia sterilans' and the early trials of arsphenamine were made on this assumption. Wechselmann (1910) reported nine patients, each treated with a

single dose of the new drug. In several, rapid clinical improvement occurred but failure in others was attributed to the use of too small a dose. Though the Wassermann test had been introduced four years earlier it does not seem to have been used routinely in these early trials of '606'. Relapses soon showed that persistent treatment was necessary, and as further chemotherapeutic substances were developed increasingly stringent tests of their effectiveness were used. Such criteria are listed by Moore and others (1936) and include the rate of disappearance of surface organisms, of clinical healing of lesions and of sero-reversal; the proportion of clinical and serological relapses; the occurrences of spinal fluid changes and the number of resistant cases.

Meanwhile several established drugs were subject to a renewed trial. Digitalis which had been studied so profitably by Withering in the eighteenth century was assessed in more modern terms by Mackenzie (1908). The general approach was almost unchanged, a wide range of carefully studied cases was reviewed in which standard preparations of the drug had been used along with the latest techniques of examination. The results confirmed and elaborated Withering's findings.

A trial of emetine by injection for the treatments of amoebic infections was made by Rogers (1912), who published a favourable report based on three patients who were not able to take ipecachuana by mouth. All made quick recoveries on a low dosage. The small number of patients, the vagaries of the usual clinical course and the absence of a long follow-up make this trial unconvincing but other workers soon confirmed the results.

The first World War raised new problems of large scale prophylaxis and treatment in medicine and surgery. Routine tetanus anti-serum for the wounded was introduced at the end of 1914. The incidence and severity of tetanus fell dramatically (Andrews 1922), though it is conceivable that change of terrain or of surgical procedures may have contributed to this result. The evidence in

favour of prophylactic anti-typhoid and anti-cholera inoculation and the value and limitations of the X^2 method of testing such data were reviewed by Greenwood and Yule (1915). In the treatment of wound shock the use of gum-saline solutions was a major advance Bayliss (1919). This was based on sound physiological theory and its value was first demonstrated upon experimental shock in animals. Its freedom from immediate toxicity was also tested and its effects in the resuscitation of a series of patients were carefully observed (Drummond and Taylor 1919); unfortunately the delayed effects upon the liver which have since caused gum-saline to be abandoned were overlooked.

The aftermath of the war provided both the stimulus and the opportunity for accurate clinical trials of the treatment of rickets. In Vienna the disease became very common in children's hospitals and was there believed to be of infective origin. The diet given in treatment was grossly deficient in vitamins. A team of British research workers was thus able to make a conclusive controlled trial of vitamin supplements and irradiation, with corrections for other factors such as season of year and exposure to sunlight (Chick et al. 1922). An exceptional type of control was available for some of the observations. A child treated for scurvy and rickets was found to have a twin brother elsewhere in the hospital with whom progress could be compared. Photographs of the two show the enormous improvement following vitamin therapy (Chick and Dalyell 1921). Another problem repeatedly attacked was the assessment of prophylaxis against the common cold and other respiratory infections by immunisation (von Sholly and Park 1921, Ferguson, Davey and Topley 1927). These trials, though having unspectacular results developed controlled investigations upon large numbers of persons to a new stage. For instance in the second study the element of suggestion was minimised by treating all the subjects similarly so that only the research workers know which had received the saline and which the vaccine injections. The first part of this trial included only volunteers, but in the second part, in order to increase the numbers of observations, non-volunteers acted as controls. It is interesting that the two parts gave similar negative results so that in

this particular example of a trial upon a fairly homogenous student population the less stringent comparison between treated volunteers and control non-volunteers seems to have been sufficient.

The two great therapeutic advances of the post-war decade were of a physiological type. Banting and Best's introductions of insulin for the treatment of diabetes (1922) was initially based upon experiments on pancreatectomised dogs; careful trial followed of the effects upon the blood sugar, glycosuria, ketonuria and general condition of diabetic patients (Banting et al. 1922). The other outstanding development was the demonstration of the value of liver treatment for pernicious anaemia by Minot and Murphy (1926). These workers had previous wide experience of the disease and knew that they might expect spontaneous remissions in about one third of cases. They gave the liver diet to forty-five patients in relapse and found that all improved and remained cured so long as they continued treatment; three who stopped treatment relapsed.

Haematological studies showed that the first response was the rise in reticulocytes and that clinical improvement could be recognised a few days later. This trial shows how much can be achieved without formal control cases when a markedly successful treatment is tried upon a chronic condition and the results are cautiously interpreted.

The treatment of lobar pneumonia by sera was the subject of several extensive and careful trials. Cecil (1928) treated several hundred cases in certain wards with serum and in others without. Typing of the pneumococi made precise comparisons possible and the mortality results showed an improvement with the serum in all but the type III infections which were almost unchanged. Park, Bullowa and Rosenblueth (1928) made a similar trial treating alternate patients with serum. Typing was also used but the authors point out that randomisation of cases within serological types was not feasible since treatment must begin before a typing result is available. The conclusions were very similar to those of Cecil and were subsequently confirmed in other trials.

New assessments of drugs used for chronic cardiac disease were made by Sir Thomas Lewis and his school. Harris (1927) made a study of the long-term results of quinidine treatment of auricular fibrillation. Lewis (1930) gave a detailed clinical description of the effect of amyl nitrite upon four cases of angina pectoris. A complementary study of the same and alternative drugs was made by Wayne and Laplace (1933) upon the functional capacity of patients with angina estimated by the amount of exercise required to precipitate pain. This group of trials had in common an emphasis upon precise diagnosis, and careful clinical observation of relatively few patients. The diseases studied were of chronic and sustained type so that the treated patients themselves could be considered as controls. The careful accessory observations provided not only an estimate of the utility of the remedies but also new information on their mode of action.

The hypothesis, based on animal experiments, that vitamins had a prophylactic value against infection was tested by Green, Pinder, Davis and Mellanby (1931). They studied the incidence of puerperal sepsis in 550 women when supplementary vitamins A and D were given to alternate admissions. The treated and control series were carefully analysed for comparability in respect of parous state, age and the like. The results showed a consistent improvement in those receiving the vitamins in all clinical severities of sepsis. The official standard of puerperal morbidity was felt to be inadequate since in spite of the evident and statistically significant lowered incidence of sepsis, judged by the morbidity standard there was no significant advantage.

The last fifteen years have possibly seen more clinical trials than occurred in the whole of previous medical history. It is evidently not feasible to review even briefly all these recent therapeutic studies. Instead a few examples only will be considered representing the fields of chemotherapy, substances used for their physiological effects, and antibiotics.

Dogmagk's announcement in 1935 of successful treatment by Prontosil of experimental streptococcal infection in animals led to numerous trials of the sulphonamides for human disease. Colebrook with Kenny (1936) and with Purdie (1937) showed that Prontosil and sulphanilamide were effective in puerperal sepsis. They observed a mortality rate of 8% which they compared with the finding of 22% for the previous years at the same hospital where the condition had long been studied. In 1937 Snodgrass and Anderson showed in a controlled trial upon 312 patients that Prontosil shortened the durations of spread, of primary pyrexia and of toxæmia in erysipelas. Control groups received ultra-violet radiation therapy or scarlet fever antitoxin. Their comparability with the group receiving Prontosil was checked in respect of age and state and severity of the disease. The authors noted that mortality in this condition was too small for it to be a useful criterion of comparison. Evans and Gaisford (1938) demonstrated the value of sulphapyridine for lobar pneumonia by a controlled trial in which one hundred alternate patients received the drug while the other hundred had routine non-specific therapy in the wards of colleagues. The comparability of the patients in sex and age distribution were tested. The case mortality of the sulphanilamide series was 8% compared with 22% in the controls. Further comparisons are offered with the results at other hospitals and previously at the same hospital which showed a mortality similar to that of the control series in the trial. Banks (1938, 1939) showed that the mortality from meningococcal meningitis could be reduced spectacularly. He first used a variety of doses combined with serum treatment and in this series 16% died; those receiving larger doses did better, so in a second series adequate amounts of sulphanilamide or sulphapyridine only were given and among 67 patients so treated only one died. The trial provided also valuable information on dosage, route of administration and the resulting levels in blood and cerebro-spinal fluid. No formal controls were arranged, comparisons being implied with the previously uniformly high mortality in this clearly defined disease and with the intermediate results in the first inadequately treated cases. A trial of

sulphonamides by Wagle and others (1941) during an epidemic of plague in Bihar is an interesting example of what can be achieved in difficult circumstances. Successive admissions to an emergency centre were treated either with anti-plague serum or iodine, a common Indian remedy, or sulphonamide. Treatment was commenced before completion of diagnosis by culture, negative cases were subsequently excluded. Other exclusions had to be made because some patients discharged themselves, but enough cases remained to show that sulphonamides reduced the mortality to less than half that of the Iodine treated controls. A similar advantage was also shown when the more severe septicaemia cases were considered alone.

Another field of advance in chemotherapy has been that of the anti-malarial drugs. Methods of study analogous to those of the sulphonamide trials have been used but there have been further special features. Therapeutic infections with malaria in the treatment of general paralysis provided a supply of known cases of standard type on which anti-malarial therapy could be tested. The results of such studies were reported by the League of Nations Malaria Commission (1933,1937). The definite course and symptomatology of the induced disease and the clear criteria of cure made it possible to give precise comparisons of the effects of different drugs and treatment routines. The ease with which the experimental disease could be produced made it possible also to use otherwise normal volunteers for trials either in hospital where detailed responses could be studied or under various conditions of stress simulating active service conditions. The latter method was used in a large scale war-time trial on the chemotherapeutic suppression and prophylaxis of the disease in the S.W. Pacific (Fairley 1945, 1946). The research team included entomologists to supply the infected mosquitoes, pathologists to study the effect of the drugs upon parasites and patients, and clinicians to look after the patients, supervise therapy and study the clinical course of the disease. Infected treated volunteers were compared with infected untreated controls for occurrence of the disease, appearance of parasites in the blood and capacity to infect other volunteers by sub-inoculation. Not only

was the great value of mepacrine and later that of paludrine clearly established but new ancillary information on the life-cycle of the malarial parasite was obtained; further, this and the pre-war trials at last firmly assessed the anti-malarial status of quinine after it had been in clinical use for 300 years.

In addition to the spectacular developments of chemotherapy for infection there has been progress in the use of substances acting upon the disordered physiology of disease. Transfusion of burned patients with plasma is an example. Black (1940) made a careful study of the biochemical and blood volume changes produced by such treatment in a small series of patients. The findings before, during and after the transfusions were compared and further comparisons were made between cases receiving different quantities of plasma for different severities of burning. Subsequent experience has confirmed these observations (Ross 1950). A therapy with a less definite rationale but which presumably acts upon the physiology of the circulation is the treatment of migraine with ergotamine tartrate. Lennox and von Storch (1935) reported a trial upon 120 patients in whom other treatment had failed. Given by various routes the drug relieved consistently the attacks of 109 of these people; this was contrasted with the previous unsuccessful treatment of this chronically recurrent condition in the same patients. O'Sullivan (1936) reported a similar trial; experience over several years had suggested to him the value of ergotamine so he made a trial of it on 97 patients during two years. This showed abrupt termination of attacks in 89 of them and of 1,132 headaches 92% were immediately relieved after the use of the drug. No data are given to indicate how many attacks would have subsided spontaneously nor in how many the psychological effects of the therapy of this largely subjective disease may have played a part. Comparisons are only available with the previous general experience of the disease, the observer's previous experience of the disease when treated with other drugs and the previous course in the same patients. The striking results give strong presumptive evidence of the value of ergotamine but cannot be considered conclusive.

The trials of methionine for infective hepatitis were based on biochemical and histological studies in animals. Wilson, Pollock and Harris (1945) studied 100 cases of this disease, giving alternate patients a dose of methionine in orangeade, while controls received the orangeade only. Treatment was otherwise similar and the two series were checked for comparability in respect of age and stage and severity of disease. Comparisons of the durations of anorexia, jaundice, liver enlargement, liver tenderness, and hospital treatment and of the incidence of relapses showed slight and insignificant advantages among the methionine treated patients. Higgins, O'Brian, Peters, Steward and Witts (1945) made a similar trial upon a smaller series of alternate cases. They followed these patients very thoroughly and checked the quantities of extraneous methionine given in the diet. Their conclusions were similar to those of other workers.

Penicillin, still the outstanding antibiotic, was first subjected to full clinical trial by Abraham, Chain, Fletcher, Gardner, Heatley, Jennings and Florey (1941). The scarcity of the material determined the selection of a few desperately ill patients, several of them children so that the maximum results could be obtained with small quantities. Dramatic improvement resulted though administration was stopped early and some relapses then occurred. Apart from trouble with pyrogenicity due to impurities the low toxicity was established. When more Penicillin became available the trial was extended on similar lines to more cases and systemic dosage was arranged to maintain inhibitory blood levels against the causal organisms (Florey and Florey 1943). The development of methods of detection and estimation of the drug greatly aided the precision of this trial and its conclusions have since been amply confirmed. No formal controls were used but the implied comparison of the clinical course of these patients under treatment with that before treatment and with the known prognosis is sufficiently convincing. Further confirmation was provided by the demonstration of the mode of action of the drug in that bactericidal levels could be shown in body fluids which corresponded to changed bacteriological findings and clinical

improvement. The effects of local application of penicillin in 15 patients with war wounds was investigated by Pulvertaft (1943) who reported a great improvement in bacteriological results in infections with Gram-positive organisms but no effects upon Gram-negative infections. The clinical effects of penicillin in the treatment of such wounds were studied in North Africa and in an extensive trial conducted in the 21 Army Group (Porritt and Mitchell (1945). For this the co-operation of surgeons was enlisted to compare the effects of penicillin with those of the most favoured alternative in the treatment of similar conditions of infection. The period of healing was used as criterion of success. It was soon found that allocation of cases was not impartial since surgeons did not feel justified in withholding penicillin from the more seriously injured. In spite of this adverse loading of the scales the penicillin group did demonstrably better. In retrospect this trial would appear to have been more successful than might have been anticipated with such indirect organisation and multiple observers. Had penicillin been less effective the biased control might have caused an inconclusive result; since the effect was so great perhaps a smaller and more precise trial would have demonstrated it with greater efficiency. While supplies of penicillin were still short a large scale planned trial of its value in sub-acute bacterial endocarditis was begun (Christie 1946, 1948). Fourteen treatment centres were organised, and different treatment schedules allotted to different centres. No controls were arranged, since the universally bad prognosis of untreated cases provided sufficient comparison. All centres at first gave a total of 5 million units, but the length of the course was varied. This showed that a prolonged course gave the highest rate of cure for a given total dose, so in the second period all patients had a long course but the daily dosage was varied. By this method of successive approximation invaluable information was obtained on the treatment of this extremely serious disease. As it happened, the optimal dosage does not seem to have been 'straddled' in either of the sections of the trial and work has continued on yet higher doses for longer periods. A less severe condition in which the relatively slight effect of penicillin makes precise controls imperative is that of finger pulp infections. Various series have been published in

which the results with use of penicillin have been compared with those of preceding or subsequent cases. Conflicting conclusions were drawn but a carefully controlled trial clarified the problem (Harrison, Topley and Lennard-Jones 1948). Patients were first admitted to the experimental series and were then randomly allocated to the different treatment groups. Detailed clinical progress was recorded and a considerable advantage could be demonstrated in patients receiving systemic penicillin. As frequently happens in such studies other improvements in treatment were developed during the trial but each such improvement was applied equally to 'treated' and control cases. Had concurrent controls not been used the effect of these changes would have been confused with that of the treatment under trial. There would seem to be many opportunities for trials of this type where several factors in the therapy such as incision, rest and chemotherapy are important but where the optimal balance cannot easily be determined.

In addition to those mentioned a host of other therapeutic substances and procedures are now undergoing clinical trial. Like those in the past each such trial presents its own problems. Some of these problems will be considered in the following section.

PRINCIPLES

The Basis of Comparison

In his study of the comparative roles of observation and experiment in medical progress Trotter (1930) pointed out that in ‘the accumulation of verifiable knowledge’ the difficulties in reliable observation of sequences, and hence of cause and effect, can sometimes be by-passed by experiments. From this point of view the clinical trial is a refinement of haphazard observation whereby the uncertainties of doubtfully comparable cases, treatments and results are minimised by experimental planning.

All therapeutic trials aim to assess the results obtained with the therapeutic agent by comparing them with some standard. Even if it is not explicit, some such standard is always implied and the choice of its most appropriate form is fundamental to the efficiency of the trial. In the words of Claude Bernard (1865), ‘En thérapeutique surtout la nécessité de l’expérience comparative a toujours frappé les médecins doués de l’esprit scientifique. On ne peut juger de l’influence d’un remède sur la marche et la termination d’une maladie, si préalablement on ne connaît la marche et la termination naturelles de cette maladie’. Bernard quotes Pinel saying to his students, ‘This year we will observe diseases without treating them....’. Relevant to this also are the studies of the limits of the normal such as are being made by the Department of Social Medicine at Oxford (Ryle 1947). For many clinical trials knowledge of what is epidemiologically ‘normal’ is also needed. In 1835 Graves pointed this out in connection with the various epidemics of scarlet fever in Dublin at the beginning of the last century. ‘The long continuance of the period during which the character of scarlet fever was....so mild....led many to believe that the fatality of the former epidemic was chiefly if not altogether due to the erroneous method of cure resorted to.... It was argued that, had the cases which proved fatal in 1801-2 been treated by copious depletion in their very commencement, the fatal debility would never have set in.... The experience derived from the present (1834-36) epidemic has completely refuted this reasoning and has proved that, in spite of our boasted

improvements, we have not been more successful in 1834-5 than were our predecessors in 1801-2.' (Creighton 1894). These and other difficulties in providing reliable comparisons are overcome in the design of a trial by comparing only 'like with like', that is, treated cases with cases otherwise similar but untreated. Since human material is inevitably variable this, in practice, involves knowing what forms of 'unlikeness' effect the comparison and therefore must be 'controlled'.

Various bases of comparison have been used in different trials. The different types are appropriate to different circumstances and will be considered in turn along with their specific advantages and limitations.

General clinical experience

The general clinical experience of the disease may be compared with the results following treatment. One presumes that this was the method of the primitive early trials as it is of many modern ones. The results obtained are often inconclusive owing to the lack of precision in the standard and the danger of other factors than the therapy being responsible for the result. Nevertheless this method has frequently served in the past to demonstrate simple advances when absence of interfering factors has combined with well marked results so as to make greater precision unnecessary. The great field for this method is in trials on diseases with a virtually certain fatal outcome when untreated; for such, the standard of comparison is precise. A recent example is the trial of streptomycin for tuberculosis meningitis (Streptomycin in Tuberculosis Trials Committee 1948). In such a disease, provided that indubitable diagnosis can be assured, any recoveries after treatment are so exceptional that they may safely be attributed to the therapy. In these circumstances there is no objection to comparison being made with general clinical experience.

Previous special experience

Previous special experience of the disease may be taken as the basis of comparison. National or local statistics, the records of the same hospital or cases treated previously by the same workers may be the standard. This method offers a more precise comparison than does 'general clinical experience' but is itself subject to important sources of error. These arise from questions of comparability of the treated and 'control' cases. The virulence of the disease or, in the case of prophylactic trials, the exposure to risk may vary in the two periods so that perhaps improved results are attributed to treatment when untreated patients would have shown a similar improvement. This applies particularly to infectious diseases; Graves' experience of scarlet fever has already been quoted; further examples are diphtheria and pulmonary tuberculosis and in both these the assessment of the value of prophylaxis and therapy has been notoriously difficult.

For some diseases variability of virulence is a negligible objection but there still remain other important difficulties. Unless the same care and interest has been previously shown in the condition studied it is unlikely that previous experience will provide a valid comparison. Criteria and accuracy of diagnosis often change, features of treatment other than the one studied may vary and criteria of success may change. It is easy to see that qualitative judgments of favourable results may be subject to personal bias or change of standards. Some quantitative measures may also be misleading especially when retrospective statistics are used.

Treatment times, for instance, may be influenced by the pressure for beds in hospital, by the doctor's interest in the disease or by social conditions making it difficult for patients to have prolonged treatment. An example of this kind of danger is shown in a trial of penicillin for sepsis of the hand (Webster 1947); figures are given both for 'control cases' from past records and for concurrent 'controls'. The mean period of disability of the former group was 26.5 days and if this had been accepted as the basis of comparison an improvement with penicillin would have been deduced since the cases so treated averaged 16.7 days disability. On the other hand the concurrent controls averaged 13.7 days, even

less than the penicillin cases. In this trial the concurrent controls were not randomly selected so that conclusions on the value of penicillin are ambiguous.

The method has its greatest value when good records of previous experience are available, when the effect of treatment is well marked, and when possibilities of changes of virulence, of diagnostic criteria, of other treatment, and of criteria of success are reduced to a minimum. Examples of this situation are provided by the early trials of sulphonamides for puerperal sepsis (Colebrook and Kenny 1936, Colebrook and Purdie 1937) and meningococcal meningitis (Banks 1938, 1939). Both these conditions had already been subjected to intensive and large scale study at the same hospitals, great care was taken to maintain criteria of diagnosis and success, and the results showed marked improvement upon previous experience of the same disease. Less convincing are examples of statistical studies of prophylaxis, such as of smallpox in Germany by vaccination (Jochmann 1913) and of tetanus in the British Army by anti-tetanic serum in 1914-18 (Andrews 1922) for here changes of effective exposure and of virulence may have contributed to the observed improvements.

The same patient

Experience of the disease in the same patient can sometimes be used as the basis of comparison. The condition of the patient before the treatment is compared with that after and when the effects of treatment have subsided it may be possible to observe a further control period and perhaps then repeat the observations. Since it is never possible to be certain what would have happened had the therapy not been given it may be difficult to eliminate a 'post hoc propter hoc' fallacy from such comparisons. This method is therefore most valuable in trials upon chronic diseases with minimal secular variation, and of treatments with short action such that the initial state can be regained after the test and so greater confidence be placed in the control. If it is feasible to repeat the test on the same patient the reproducibility of the effects can be checked. The studies of the value of amyl nitrite in angina pectoris (Lewis 1930, Wayne and Laplace

1933) are good examples of the effective use of this method in circumstances which fulfil these desiderata.

Wright (1941) in pointing out the advantages of this type of control to which he gave the name 'se-ipsic' or 'idio-proteric', stated that statistical methods had not yet been devised for its analysis. This is no longer true, for a number of recent experiments (e.g., Fourman 1950) have combined this method of control with Flasher's analysis of variance technique.

A danger of this method is that in the detailed study of a few patients the possibility of individual variations may be overlooked, the reproducibility of an effect in one person giving a perhaps erroneous impression of a universally applicable response. If this is realised it should usually be simple to make the necessary checking observations on a more representative sample.

As mentioned by Gaddum (1948), a special type of 'se-ipsic' control is possible in studies of skin conditions. Here different areas upon the same patient can be compared concurrently. This was done in a trial of sulphonamides and other local applications for impetigo (Sheehan and Fergusson 1943). Twenty-five observations upon ten different treatments were obtained and the resulting rates of healing compared. A smaller untreated group showed the marked tendency to spontaneous cure in this condition but the advantages of an Alibour paste with sulphathiazole were demonstrated. A study of the effects of local applications for burns used the same technique (Cannon and Cope 1943). Tanning and other types of treatment were compared upon different areas of the donor sites of skin grafts. Tanning and dye methods were shown to delay healing as compared with the control application of boric acid ointment. The use of donor sites rather than burns themselves was determined by the need to have similar severities of lesion for each comparison; where this can be arranged the method would seem to be highly efficient within its rather limited field.

Concurrent controls outside the trial

Concurrent experience of the disease in patients other than those treated by the method under trial provides a basis of comparison free from some of the objections so far considered in that secular changes can be eliminated. The technique is used in several forms; these can be grouped according to the method of selection of control and treated cases. A simple arrangement is for all patients at one centre to receive the new therapy and the controls to be provided by another centre. This method was used in the early French trials of anti-diphtheritic serum. All diphtheria patients at the Hospital des Enfants Malades were given the serum and results were compared with those at the Hospital Trousseau during the same period (Roux, Martin and Chaillou 1894). Such a method had evident administrative advantages but great care must be taken that the cases in the two centres are similar in all material respects. This can seldom be ensured unless both series are under unified supervision and this can often be better achieved within one hospital or by having 'treated' and 'control' cases at each hospital in the trial. An advantage of the cruder method is that it may allow enthusiasts for a particular treatment to employ it and hence get the best possible results and perhaps to have these results compared with those of enthusiasts for an alternative treatment. This has applied to the trials of Stockholm and Paris technique of radium treatment for carcinoma of the cervix uteri and to their comparison with the Wertheim operation. Considerable care has been taken in these trials to maintain comparable standards of severity so that as far as possible 'like' can be compared with 'like'. Another advantage is that the segregation of the two series may obviate a moral difficulty of using treatments which are not believed by the doctor to be best for the patients. Thus if there is a difference of medical opinion on the treatment under test then it may be possible to compare the results of supporters and opponents without interfering with the treatment the patients would have received had there been no trial.

Comparisons between centres are particularly suitable for trials of treatments which require the co-operation of many persons not directly concerned in the

research. Williams and Miles (1949) point out that it may be the only feasible method for trials of dressing techniques which are to be used by nurses; having perhaps persuaded the staff of a factory surgery to adopt a new procedure one can hardly ask them to treat control cases by the old methods. Instead, the best available comparison is with cases treated at another centre as far as possible comparable, but not using the new methods (Clayton-Cooper and Williams 1945).

Controlled series

The form of trial which usually best fulfils the logical requirements of comparability is that in which cases are so allocated to treated and control series that any otherwise uncontrolled variation which may effect the results is randomly distributed in the two series, and where the treatment of the two series only differs in respect of the item under trial. When these aims are achieved none of the previous objections apply; the difficulty lies in their achievement. The method will demonstrate slight improvements in results such as are missed by comparisons with previous experience or with concurrent results at other centres, its field of potential application is therefore wider than that of any other technique. Some details of this method are considered again later.

The question may arise whether it is justifiable to have 'untreated' controls. This difficulty is often more apparent than real. Firstly the question is not important when the disease is a mild one nor in the common situation where the expected improvement is a matter of a few days of treatment time; in such circumstances the greater certainty of a well-planned trial is much to be preferred to dubious results and possibly repeated inconclusive trials by less efficient methods. Secondly a new and possibly potent therapy is usually only available in small amounts when first introduced. However effective it may be there is not sufficient to treat all cases of the appropriate disease. In these circumstances it is clearly best that the cases which can be treated shall be used to yield the maximum of information by being compared carefully with cases not so treated.

Thirdly, apart from this ethical duty to advance medical knowledge, it must be remembered that many proposed new treatments are themselves dangerous and it is unjustifiable to use such treatments unless the maximum of information on their possible deleterious effects can also be extracted. In less unexplored fields of therapy it may be possible to anticipate injury to treated or control cases by planning a safety clause whereby patients whose condition warrants it are given other treatment. The war-time trials of mepacrine and paludrine for malaria illustrate this, a known effective anti-malarial course being provided for failed cases (Fairley 1945, 1946). As a compromise the control patients may be treated by the most favoured alternative agent. This was done in the 21 Army Group penicillin trial (Porritt and Mitchell 1945) in which the clinicians were allowed to choose their preferred alternate therapy for equivalent cases to those receiving penicillin.

Criteria for inclusion of cases

A trial may include a wide variety of cases or it may be limited to a carefully defined group. Sir Almroth Wright referred to this aspect of experimental design as 'claustration', meaning the degree of shutting in of the field of observation (1937). Early trials of drugs have often been over a wide field, sometimes because the likely usefulness was unknown and sometimes because of inadequate diagnostic discrimination between different conditions. Withering's study of digitalis exemplifies this; the drug had been proposed as a cure for dropsy and Withering, having no source of knowledge on the different cases of dropsy tried it for a large range of patients. Being a careful observer he was able to identify the kind of case most benefited. More recently, in keeping with greater knowledge of aetiology and pathology, it has become usual to make the wider investigations in the laboratory by in vitro, or, animal experiments so that the likely action of the drug in man is already established. Suitable patients can then be selected in a 'claustrated' field, so giving maximum precision of results with the minimum effort. Though this is doubtless the most useful general method for 'screening' new therapeutic agents the discrepancy between

laboratory and clinical effects may be so great as to lead to serious error. Something of this sort happened in the development of the sulphonamides. Compounds of this type were synthesised in 1908 and were tested as antibacterial agents in vitro in 1913 but their clinical possibilities were overlooked until Dogmagk and his collaborators made their famous studies upon mouse and human infections twenty years later (Long and Bliss 1939).

Criteria of Success

Another group of criteria important in the structure of a trial are those accepted as evidence of success of the therapy. Not only does the final interpretation of the results depend on these criteria but their correct choice can greatly improve the efficiency of a trial in terms of return for work done. Too often in the planning of trials this aspect is overlooked and effort is therefore wasted in making observations which are later found to be unnecessary or useless.

The choice of an appropriate criterion may be helped by detailed knowledge of the pathology and course of the disease. Examples of this are the use of reticulocyte counts in the assessment of anaemia remedies and of Wassermann reactions in trials of treatments for syphilis. Such short-cuts must not be accepted too readily. A warning of this is given in the recent trials upon Addisonian anaemia where folic acid which corrects the blood picture has failed to prevent the more insidious nervous lesions.

The chosen criteria whether qualitative or quantitative must be precisely defined. If the qualitative criterion is one such as death or survival the definition should give no trouble, but should it be eradication of infection or freedom from symptoms then constant and precise standards of observation and recording must be used throughout.

Criteria for inclusion and for subdivision into groups as well as of success are subject to error on the part of the observer. This can be reduced by using

wherever possible objective rather than subjective assessments. There usually remain features which must be matters of clinical judgment and precision and consistency can here be improved by careful definition. In their recent study of this problem in the taking of clinical histories, Cochrane, Chapman and Oldham (1950) distinguish between the variation of estimates made by the same observer on different occasions ('Intra-observer error') and differences from the estimates of other observers ('Inter-observer error'). They confirm the value of a prearranged grading of symptoms and signs and recommend parallel judgments made independently by several observers (Cochrane 1950). Similar conclusions were reached in a study of estimates of nutritional status (Bean 1948). An aspect of 'observer error' which can be particularly troublesome in clinical trials is the progressive change of judgement which is liable to occur imperceptibly in the course of the trial. The careful study of a series of similar patients inevitably increases the clinician's understanding of the conditions and hence his judgment. A safeguard is to arrange that both treated and control series are similarly affected by any such trends of judgement; this can usually be achieved by randomisation of sequence but though the two series may thus be made comparable a trend may still reduce the precision of the comparisons.

The criteria of success or failure of the trial as a whole must be distinguished from those relating to the results in a single patient. Both may be stated in similar terms such as mortality, length of treatment or disability, but in practice the worth of a treatment must be judged on the degree of improvement in comparison with the cost in effort, risk of deleterious effects and the like. Greenwood (1935) drew a similar distinction between the 'effectiveness' of a prophylaxis which is directly measured by the degree of protection achieved and the 'advantage' which relates this protection to practical cost and return for effort. The statistical method of Sequential Analysis, to be mentioned again later, emphasises this aspect by requiring a decision by the investigator on the degree of improvement in results which shall be considered as sufficient for a judgement of overall success. For instance it would not be sensible to

recommend an expensive in-patient treatment for a minor complaint on the grounds of a five per cent. more rapid cure, but it might be correct were the improvement one of fifty per cent.

The role of Statistics

Both enumeration of cases and mathematical methods for estimating significance are commonly called statistics. Most modern clinical trials use both these aspects of the subject. The value of enumeration was ably defended by Louis (1834) both as a discipline to prevent weighting of clinical impressions by exceptional cases and as a method of averaging the inevitable variability of clinical material. Louis himself made good use of the mean duration of illness as a method of comparing alternative treatments for pneumonia (1835). Long comparative series were already used in the eighteenth century as in the studies of inoculation and vaccination for smallpox but the employment of statistical methods for tests of significance is a development of the twentieth century. Karl Pearson (1857-1936) first defined the Standard Deviation in 1894 and introduced the X^2 test in 1900 (Cajeri 1929). The method of comparing the means of small samples was developed by 'Student' in 1908 as the 't' test and further improvements to these tests were later made by Fisher.

The aim of the statistical tests of significance is to estimate the probability that the numerical results obtained were the result of 'chance' and therefore not attributable to the action of the remedy under trial. In this way a check is provided upon possible over-estimation by the experimenter. The other value of such tests is that they can indicate the point in a trial at which a problem has been solved at a definite level of probability. This makes for maximum efficiency in time and material and these techniques should be obligatory in trials involving the possibility of human suffering or damage.

The grounds for the objections sometimes raised against the use of statistics lie not in the methods themselves but in the danger of too great reliance being

placed upon them so that they are used as a substitute for careful observation and sound reasoning. No elaboration of enumerative statistics can correct inadequacies of recording nor any test of significance overcome failure to allow for the appropriate comparisons in the design of the trial. What statistics can do is to provide precise methods of analysis where otherwise personal impression and guess work are only available. In this way they act more often as a further critical discipline than as a short cut to results.

A practical difficulty which may give trouble in the study of long series was pointed out by Wright (1937). Where, as in the assessment of a partially effective prophylaxis of a rare disease, very large numbers of persons must be observed to demonstrate any effect it becomes progressively more difficult with such large numbers to ensure the accuracy of dosage, the constancy of the agent and the reliability of observation and recording. If, as is postulated, no other technique of trial is possible the only solution would appear to be a more elaborate staff of observers so that sufficiently greater constancy is obtained. Short of this the value of the prophylaxis must remain in doubt, the criticism being of the standard of observation and not of statistical method.

A more fundamental difficulty lies in the implications of the classifications used in statistical analysis. Any grouping involves some merging of individual detail in the group. The results of statistical analysis are usually calculated for groups, and deal with the probable behaviour of such groups in the absence of detailed information on the causation of individual results; but the conclusions of therapeutic trials are usually applied to individual patients. There is thus a danger that inappropriate treatment may be given to exceptional patients. This difficulty again is not primarily statistical but is common to all generalisations in medicine. Its dangers can be offset by making clear the limitations of interpretation of the results of a trial and as far as possible using simple clinical classifications in the statistical analysis. A more common source of error in practice is the reverse one of claiming that clinical impressions establish certain patients as requiring

exceptional treatment while making no attempt to obtain statistical evidence showing this to be true.

A technical criticism of the use of statistics concerns the mathematical assumption made in some of the tests of significance. The commonly used tests for the differences of means assume that the values compared are normally distributed. This is frequently not true of the observations from which the means are derived and the methods might therefore appear to be invalid. Fortunately, considerable departures from normality in the data make little difference to the distribution of their means so that methods such as the 't' test are still applicable (Student 1908). As a safeguard it is advisable to test experimental distributions for their approximate normality since alternate methods of analysis such as are considered later may be more suitable.

Statistical methods can be useful for estimating the numbers of observations which must be made, in practice usually the number of patients who must be studied, before a trial can be considered conclusive. In general the number needed depends on the variability of the material, the degree of improvement observed, and the consistency with which the improvement is produced. More specifically, the precision of the estimate of a mean varies as the square root of the number of observations and inversely as the scatter, measured as the standard deviation, of the observations. When two means are compared, the value of 't' and with it the significance, varies directly with the observed difference between them. Thus a marked difference, such as might result from a very effective therapy, which is consistently produced upon consistent material will demand few observations to reach the usual standards of experimental probability for excluding a chance occurrence. On the other hand, a therapy causing a slight and irregular improvement upon variable material will require extensive trial before it can be proved to be of value. Stated in this form statistical methods confirm the judgments of common sense, the advantage being that statistics can make these judgements more precise.

Certain methods of design and analysis help to reduce the loss of efficiency due to variability. This can sometimes be achieved by pairing treated and untreated cases and calculating not the means of the treated and control series but the mean of the differences between pairs. The estimate of error is then based not on the scatter of all the values but on the scatter of the differences. If the pairs have some important feature in common such as being observations on the same person before and after treatment, then this technique will reduce the effect of variation between individuals. On the other hand should there be no considerable correspondence within pairs the method may be ineffective since the degrees of freedom for the estimate or error are reduced by half. If pairing is used it must be upon logical a priori principles and not, for instance, by ranking order of results, an error discussed by Fisher (1935).

We will now consider some statistical methods which are not in common use for therapeutic trials but which may have valuable applications in the future.

The greatest development of statistical experimentation in recent years has been in agricultural trials. For these, very effective methods have been introduced by Fisher (1935) which combine the statistical technique of the Analysis of Variance with a design of trial specially adapted to this analysis. An advanced form of this is the factorial Latin Square experiment in which a series of treatments can be tested simultaneously upon several varieties of plant with full correction for soil variability. The results can be analysed to give comparisons for any pair of treatments or varieties so that one experiment can be substituted efficiently for the long programme of study necessary to obtain similar results by the classical method of paired series. There are considerable difficulties in applying this method to clinical trials. For instance, it is seldom possible to obtain an even distribution of patients in severity categories which might correspond to 'varieties' in the agricultural model; similarly alternative medical treatments cannot commonly be combined in the manner of fertilizers. A Latin

Square design was successfully used in a recent trial of d-amphetamine and thyroid for obesity (Edwards and Swyer 1940). Here there was a supply of comparable patients who were given four combinations of treatments and placebos in successive months. The allocation of patients to different treatments and the sequence of treatments for each patient were randomised as a Latin square. A fully balanced statistical analysis showed with great efficiency the improved results with amphetamine and the ineffectiveness of thyroid under the conditions of the test. The Latin Square design was also used by Mackworth and Winson (1947) for allocation of treatments in a psycho-pharmacological trial of amphetamine.

Apart from these special designs, Analysis of Variance may be a useful alternative to 't' testing in the more usual types of trial. In principle the variation between treated and control groups is compared with the variation when interfering factors such as a type of patient or order of admission have been excluded. The method both measures and eliminates these factors and the number of them that can be so handled depends on the design of the trial. Factors within the power of the observer to distribute among the cases can be readily analysed. In this way such variables as place of treatment, or dosage and duration of therapy can be separately studied. Other differences which are intrinsic in the patients such as age or location and severity of disease are more difficult since it cannot usually be foreseen how many will fall into pre-arranged groups and analysis of unequal groups is much less efficient than of equal ones. These are the same difficulties which also attend the Latin Square design. A different kind of difficulty has arisen in agricultural experiments when plots of ground have been too small for all treatments to be tried on each. The method of Balanced Incomplete Blocks solves this problem and would also be applicable to the comparison of multiple therapies for skin disease, each area being analogous to a plot. Another similar manoeuvre is possible in factorial experiments whereby economy is effected in the numbers of observations at the price of 'confounding' and eliminating together differences between certain groups of observations and

interactions between different factors. These modifications are not particularly appropriate to the problems of therapeutic trials but something similar might adapt the technique of analysis of variance better to clinical material. Should this be possible an enormous saving of effort and increase of information could be achieved in many trials.

Where the results of therapy do not lend themselves to accurate quantitative measurement it may nevertheless be possible to rank them in order of merit. Thus in a series the first two patients may have been treated by the method under trial, the third best by the control method, the fourth as the first two and so on. The probability that a given result of this type would occur by chance can be calculated by methods described by Fisher and Yates (1948) and by Whitfield (1947). This technique, though not yet widely used, would seem very suitable for many therapeutic trials where criteria of success are qualitative or depend on the assessment of several different features whose combination cannot be given a quantitative value but which can be placed in order of merit by clinical judgement.

Another recent statistical technique which may be useful for clinical trials is that of Sequential Analysis. This was used in war-time research on equipment testing (Statistical Research Group 1945 and 1947). It offers a method of evaluating the results of a trial case by case so that it is immediately known when the evidence shows at a pre-arranged level of probability that the tested procedure is better or worse than a predetermined standard. The technique is applicable to quantal rather than quantitative data and to situations where a rigid standard of excellence of result can be prescribed. It might thus be very useful in animal assay experiments where the quantal response could be death due to a given trauma and the therapy under test be intended to prevent death in say fifty percent. of cases. Clinical trials with a sample criterion of success such as presence or absence of infecting organisms could be similarly analysed though the aim is usually to show whether one therapy is better than another rather than

to compare it with a fixed standard. Nevertheless there are trials for which Sequential Analysis is appropriate and further extension of the method may easily make it more widely applicable.

It has been mentioned that a possible criticism of some statistical tests is that they may imply unjustifiable assumptions as to the distributions of the value analysed. Though the Binomial and Normal distributions are the ones most commonly employed it is remarkable how few biological populations really fit these curves. Measurements of stature and body-weight closely approximate to the Normal distribution (Yule and Kendall 1937) and they might be used in clinical trials but the more usual measures such as duration of illness are commonly skew. Gaddum (1945) pointed out that many of these biological distributions are fitted better by a log-normal curve. This implies that if the time units are treated as a geometrical series, the interval between two and four days being equated to that between four and eight, the distribution becomes symmetrical about the most frequently occurring value. The mode thus approximates to the median. The log-normal distribution does not appear to have been used in therapeutic trials through several examples of clinical data have been found to fit it, such as the treatment times of burns and the healing times of finger-pulp infections (Bull 1949). Another method of analysis with similar characteristics is that using the logarithmic decrement. This gives a satisfactory fit to many data of treatment or disability periods and was employed by Whitfield (1946) in his study of colliery accidents. It has useful properties in that the rate of the persons studied, for instance the number of cases healing in a period, can be given as a fixed proportion of the number at risk; another practical advantage is that serial studies can be made without waiting for the completion of long term cases (Whitfield 1950). This method has not yet been used for clinical trials but it might have useful application to large scale studies of types of treatment based on hospital statistics. In both the log-normal and the log-decrement distributions the most convenient average is the median rather than the mean. This can be used as a representative measure for comparisons but it is

less useful than the mean for certain calculations; for instance of the hospital accommodation which might be needed for treating a certain number of patients.

A statistical technique which has been developed for experiments on biological assay but which also has potential value for clinical trials is Probit Analysis (Finney 1947). This is a method for fitting accurately a straight line to data of sigmoid distribution. The percentage mortality of animals commonly gives a linear Probit relationship when plotted against dosage of a toxic drug expressed in logarithmic units. An accurate estimate of the 'LD50' can then be made and the slope of the Probit line is a measure of the scatter of susceptibility in the population studied. The method has been used in an assay of hormones in human beings whereby the effects of different synthetic oestrogens were compared in terms of the amount of drug necessary to induce withdrawal bleeding in fifty per cent. of patients (Bishop, Kennedy and Wynn Williams 1948). An application of the technique to mortality due to burns has also been made, using the age of the patient and the surface area burnt as equivalent to dosage of noxious agent (Bull and Squire 1949). This has been suggested as the most efficient way yet available for comparing the results of treatment of burns at different centres. Probit Analysis appears to be most promising for clinical trials which have a quantal end point such as death or withdrawal bleeding combined with a quantitative measure of severity of lesion or dosage of therapy. An alternative to Probit Analysis, applicable to similar data, is the method of Angular transformation (Fisher and Yates 1948). This technique has no simple biological interpretation and its status is that of a convenient mathematical manoeuvre, though it has practical advantages over Probit Analysis in ease of calculation. Even in their present forms these methods have many useful applications but as with other techniques mentioned, greater adaptation to clinical data might increase their scope.

Estimation of Optimum dosage

The problem of establishing optimum dosage frequently arises, particularly in trials of chemotherapy. This dosage is such that therapeutic effect is as great as possible consistent with low toxicity, convenience and economy. It may need definition as quantity per day, route and frequency of administration, and total length of course. The ideal dose may be related to the patient's weight or age and may allow for early loading of body fluids and for later reduction to offset cumulation. The full details of such optimal dosage schedules have seldom been thoroughly established. Arsenical treatment of syphilis, sulphonamides for pneumonia, gonorrhoea and meningitis, and certain anti-malarial drugs are among the best documented but the optimum dosage of most antibiotics is still uncertain.

A therapeutic dose, preferably close to the optimum, is required in clinical trials and one aim of a trial may be to estimate what this dose should be. Different patients may be treated with different dosage levels and the results compared between levels and against control cases. Where there are objections to 'untreated' controls it may be sufficient to compare sub-optimally treated groups with others receiving higher dosage though more cases may be needed to demonstrate the less marked difference in results.

A method applicable to certain types of therapy is that of building up the dosage upon individual patients until the required response is obtained. This minimum effective dose offers a basis for further experiments to establish the optimum. The technique corresponds to the 'Cat method' of pharmacological assay for digitalis and provides a serial 'se-ipsic' type of control. A similar procedure is often employed in the clinical use of drugs having a physiological type of effect with a recognisable end point. Examples are anaesthetics, analgesics, anticoagulants, drugs affecting cardiac rhythm, emetics and purgatives; the method would be appropriate to clinical trials of treatments producing such effects. When an optimum has been reached it may need confirmation with omission of preliminary sub-threshold doses so that summation effects are

excluded, and, as with other tests using se-ipsic controls, sufficient patients should be studied to check the general applicability of the results.

Where different levels of dosage are studied it may be possible to show a mathematical relationship between quantity of therapy and its effect. The statistical techniques of regression analysis and curve fitting may assist in fixing the optimum and may also suggest a particular mode of action of the therapy. This should be interpreted with great caution since, as Clark (1933) pointed out, even the more precise data of this kind from the laboratory can often be equally well fitted by a variety of different curves, each implying a different mode of action.

Most dosage schedules, as mentioned above, have several variables such as quantity per day, frequency and the like, and each variable will have an optimum value which will sometimes be dependent upon other variables. There is the further possibility of several optima as when a five a day course of, say, penicillin in twice daily doses of 500,000 units gives equal results to a three day course of 100,000 units three-hourly. In these circumstances convenience and economy would determine which schedule was preferable.

One method of estimating optima of a series of variables is to fix all the variables but one at reasonable levels on a priori considerations, then to test the remaining item at different values until the best results are obtained. Maintaining this at the newly established optimum, another variable is selected and tested at different levels to find its best value; and so for all the others. Finer adjustments can be made by further rounds of tests. The principles of this method are outlined by the Statistical Research Group of Columbia University (1947) as applied to engineering problems and are similar to those used in the trials of penicillin for bacterial endocarditis (Christie 1946, 1948). There seems to be room for improvement of this technique to make the tests more economical and to provide for estimates of precision. These might be better supplied by a factorial design as

described earlier in which different variables are studied concurrently though such a method does not yet appear to have been used in clinical trials.

The Technique of the Controlled Trial

Having chosen an appropriate form of comparison, criteria of inclusion and success and the method of analysis for a therapeutic trial, there remain certain details of planning and execution which will now be considered.

The Agent

The therapeutic agent must be studied in detail. The preliminary information required for immunising substances has been described by Topley (1933) and for chemotherapeutic drugs by Marshall (1947). In general, the maximum available information is needed upon the expected therapeutic action, any toxic or damaging effects of the therapy, and methods of standardisation, of administration and dosage. Knowledge of mode of action, and in the case of drugs, pharmacology, metabolic products, chemical structure and properties, and methods of estimation can also be invaluable. Taking a locally applied antibiotic as an example it is necessary to know the bactericidal spectrum in vitro with and without serum, with special attention to resistant organisms, the effect of the antibiotic upon animal infections and any corresponding human evidence, its toxicity to animals in therapeutic and greater dosage, its local toxicity to living tissues, leucocytes and fibroblasts and any antigenic or sensitisation effects. The study of toxicity must include long term and possible delayed effects. It should be possible to standardise the agent so that preparations of known strength and purity can be used. The dosage should be devised in relation to effective experimental concentrations and to toxicity, and it may be desirable to test several dosage levels. Frequency of application must be arranged so that effective concentrations are maintained in the lesion. The choice of an appropriate vehicle may be crucial for such a trial; it should combine satisfactory properties of easy application, good degree of penetration and persistence with

freedom from toxic effects and from chemical interference with the action of the agent.

The disease

A similar close study is required of the disease to be treated. It is necessary to find the most suitable criteria of diagnosis, of admission to the trial and of success as previously discussed, and to establish what groups are to be distinguished in respect of type, severity and location. Examples of such groups are those derived from the bacterial types in pneumonia, the depth of lesion in a burn, or the anatomical site in fractures. If these groups are likely to differ in response to the therapy or have a different prognosis then it is usually better to test each group separately against its own control. Should the grouping seem unnecessary on analysis of results it can at that stage be abandoned with no loss of precision. The opposite occurrence of the finding of variation which has not been controlled may defeat the trial completely or render it much less efficient.

The patients

As with the disease it is necessary to consider in what way the available patients may vary and so affect the interpretation of the trial. In some conditions such as pneumonia or burning injuries the age of the patients affects the prognosis. In others there may be a sex or nutritional effect; for all these factors appropriate groups must be arranged so that such variation is not confused in the results with an effect of treatment. If ample clinical material is available it may be more convenient to exclude some of the rarer categories from the trial so as to concentrate upon obtaining a clear result for a common well-defined group. To do this entails the dangers of Wright's 'hyper-claustration' as discussed previously.

With some conditions it may be possible to test the treatment upon artificially induced disease. This can be of great advantage as has been shown in trials upon Malaria. A considerable moral responsibility may be incurred and in general,

volunteers only should be used who have had the dangers of the procedure fully explained. The difficulties in applying this to prisoners has been discussed by Ivy (1948). Such trials must be conducted to yield maximum results with the minimum of human suffering, and medical staff, including where feasible those in charge of the trial, should also be subjected to the procedure. A complete travesty of these principles seems to be presented by some of the experiments performed in Nazi Germany (e.g., Ding 1943).

The control therapy

The treatment given to the control cases must differ from that of the other patients only in the absence of the active agent. This does not mean that the controls should merely have no special therapy. Apart from any psychological effects an appreciable amount of accessory treatment may be associated with the giving of a specific therapy. It is therefore usually best to devise a dummy treatment for the control cases that will require the same procedures. Preferably, neither patient nor research worker should know which contains the active agent. Examples are inactive coloured tablets for oral therapy (e.g., Askew 1949), saline for injection (e.g., Ferguson, Davey and Topley 1927), or the vehicle only of an ointment tested for local application (e.g., Lowbury and Topley 1950).

Randomisation

The method of randomisation must be adapted to the practical convenience of the particular trial. For many purposes alternate cases are suitable. Bradford Hill (1937) describes the details and scope of this system. Its chief dangers occur when the decision to include patients in the trial is a matter of clinical judgment. Knowledge of what treatment the patient is due to receive may then effect this judgement so that allocation is not truly random. Sometimes it is convenient to use alternate days' admissions. This is simple for subsidiary staff to operate, but in trials upon minor lesions in a closed community as for instance within a factory the patients themselves may learn the code and present themselves selectively on the days of the preferred treatment. The use of a random series is

free from these sources of error. This may be pre-arranged by coin-tossing or card sorting, a list of random numbers may be employed, or the randomisation may be made by one of these methods after the case had been admitted to the trial. It is useful to ensure beforehand that equal numbers of cases shall fall in the various groups of treated and controls. This can be done by limiting the numbers of cards or counters to not more than the proposed cases in the trial and by not returning the counters after selection. Such a method has been used in studies by the British Empire Cancer Campaign and has been recommended for American trials on poliomyelitis (Therapeutic Trials Committee 1949).

Investigation facilities

Having decided upon the special features which should be studied in the trial, the facilities available for making the required observations should be reviewed and the scope and tempo of the trial adjusted to their capacities. Technical facilities will be needed for bacteriological and biochemical studies. In trials of chemotherapy these may include special methods for measuring sensitivity to the agent, for detecting its presence and for estimating its concentration in body fluids. Many such techniques require practice before reliable results are obtained. Clinical examinations and investigations such as X-rays may need special facilities for their consistent performance. Another aspect which may be crucial for the success of the trial is the organisation of adequate follow-up of the patients. This is frequently one of the most difficult administrative problems, and failure to solve it has detracted from the value of many otherwise good trials, from eighteenth century tests of vaccination to the present day. Any loss in follow up should be at random and, in general, study of a small sample which can be completely followed is preferable to a larger trial with high wastage.

Organisation of staff

A therapeutic trial is a scientific experiment. For its success there must be constant vigilance and co-operation by all concerned. This can only be expected if all responsible persons are informed of the purpose and value of the

experiment and are agreed on its method of execution. Where several centres are taking part, special liaison arrangements are required. Within one centre frequent meetings are valuable to discuss difficulties and to maintain interest. Once a trial has become routine there is a natural tendency for standards to be relaxed, but intelligent co-operation can overcome this and so efficiency be maintained.

The Pilot Trial

When the outline of the trial has been determined it is often useful to test the plan by applying it initially to a few cases only. This will show its feasibility and may indicate important omissions, errors of emphasis, or perhaps some unnecessary over-elaboration. The details of treatment may require clearer definition, the system of recording may need revision, or criteria of success can perhaps be simplified; it may be possible also to use the pilot trial to fill in gaps of the preliminary information. The order of difference to be expected between the treated and control series is often soon evident as well as the frequency of cases, so that a more precise estimate can then be made of the probable extent and duration of the trial. If the pilot confirms the soundness of the design it should be possible to include its results with those of the main series. Should important changes be required it is better to start afresh rather than force the remainder of the observations into an unsatisfactory mould.

Whether or not a formal pilot trial is used it is often valuable to make an interim review of results early in the study. If it is planned to assist the final analysis of results by transfer of the data to punch-cards or the like then this early review is an opportunity to test the coding methods. The records should be treated as for the complete analysis, inefficiencies detected at this stage can be corrected before much harm is done. The findings will check further the likely duration and conclusiveness of the trial and will show whether the necessary discipline in the management of the cases is being maintained and will indicate the adequacy of the methods of randomisation. Where the research is being done by a group it may be inadvisable to announce the interim results in details since this may

cause prejudice in subsequent observations. How much should be made known will be a matter of judgment; reassurance that progress is satisfactory will often suffice.

DISCUSSION

The clinical trial may be considered as an application of experimental scientific method to the validation of therapeutic procedures. The ‘trial and error’ investigations which presumably have been used for simple problems throughout the history of medicine contain the germ of this experimental approach but are inadequate to solve more complex problems. The need for the development of clinical trials to deal with these has been frequently obscured by undue confidence in tradition and a priori theory. This is still the position in many primitive cultures (Ackerknecht 1944) where even trial and error methods may be discounted. Among such peoples Rivers (1924) demonstrated a distinction between the orthodox ‘professional’ therapy of exorcism and ritual based logically on beliefs in possession or sorcery and the ‘domestic’ therapy for minor complaints. In many part of the world this includes poulticing, blood letting, massage, bathing and counter-irritation, all unrelated to local aetiological theories and much closer to scientific therapy. A similar development of practical empirical therapeutics divorced from orthodox theory has been noted in mediaeval and renaissance time, and the folk medicine from which the eighteenth century workers drew their material probably had similar origins.

The strength of traditional beliefs can be judged from examples of even leaders in this field of clinical trials who have failed to rely upon their own findings. Having demonstrated the effect of fruit in the cure of scurvy, Lind recommended change of air as the first requirement of treatment (1753); similarly after Louis had shown the futility of bleeding in a variety of conditions he still continued to advise it for the haemoptysis and pleurisy of phthisis (1843). The emergence of the modern therapeutic trial is thus an aspect of the growth of reliance upon impartial observation as the foundation of scientific medicine.

Study of the principles of trials gives an indication of the circumstances in which they can be successfully and easily performed, and conversely the sources of difficulty. The therapy should be simple and measurable so that it can be applied

similarly on all occasions. The condition treated should be clearly definable and either of regular severity in different patients or of measurable severity so that similar cases can be grouped together. This can sometimes be achieved by experimental induction of the disease. A satisfactory comparison must be available between treated cases and cases otherwise similar but not receiving the treatment under test. This may be provided by a known uniform prognosis in untreated cases, by use of alternating control periods in chronic diseases, or by randomisation of different cases into treated and untreated series. The criteria of success and failure should be clear and should be recognisable early. Preferably they should be chosen so that the control cases show a high proportion of 'failure' since the trial may otherwise be prolonged or inconclusive. The trial will further increase in efficiency the more regular is the clinical material and the more constant and marked the effect of the therapy. A controlled trial offers an enormous increase of precision over other methods but it is still imperative that the soundness of the comparison be critically examined. Results must be carefully interpreted so that conclusions drawn are within the limits of the logical structure of the trial and due regard to special experimental conditions should be paid when the findings are applied clinically.

The varying complexity of trials appropriate to different therapeutic problems has been indicated. Historically a difference between Medicine and Surgery has been noted; whereas Surgery made great progress in ancient times the development of medical therapy was delayed until recent years. Part of this discrepancy can be attributed to the differences in the respective clinical trials demanded. Much of Surgery deals with lesions of simple aetiology and diagnosis, and the application of simple mechanical principles often produces immediate spectacular results. Suturing of wounds and reduction of dislocations and fractures give evidence of success at once and leave little opportunity for elaborate theorising. The fundamentals of such treatment had already been worked out in ancient Greece and Egypt. By contrast few medical conditions could be understood until ancillary scientific methods has been developed. The

aspects of medical diseases which could be appreciated such as fever, dropsy, jaundice or wasting were of such varied aetiology that any but the simplest treatment by rest or diet would have inconstant effects and even if active agents were available the results would seldom be immediately apparent. It is thus not surprising that superstition and scholastic theorising thrived and few therapeutic advances were made.

It seems possible that the success of simple surgical trials in the past has led workers to believe that similar methods remain applicable to the complex procedures of modern Surgery. In his criticism of the validation of gastro-jejunoscopy, Ryle (1948) pointed out that trial and error methods have been repeatedly misleading and that, in the absence of controls, adequate follow-up of patients and statistical analysis such an operation cannot be fairly assessed. The effects of extensive surgery can be quite as complex as those of an elaborate medical therapy; not only must immediate operative mortality and the achievement of anatomical success be considered but also late effects of altered function and possibly secondary disease. Such a situation demands the same care in the planning of a satisfactory clinical trial as does the detailed assessment of a new and potentially dangerous antibiotic.

The unintentional 'trials' of wound treatment made by Paré and Hunter were incidental to the usual routine of practice. In a comparable way some modern trials have been incidental to other studies. The observation of Anning (1947) on the effects of Calciferol upon chilblains provide an example. The occurrence of chilblains upon patients receiving this treatment for other complaints was compared with their occurrence on patients not so treated. No difference was found between the groups. Studies on the pathogenicity of certain organisms can also be profitably combined with tests of anti-bacterial therapy; this is being done in current trials upon Polymyxin and Aureomycin (Lowbury and Topley 1950). This integration of therapeutic trials with other aspects of medicine is further shown in the diagnostic advances made in Withering's study of digitalis

and the Vienna studies of rickets and in the new information on the life cycle of plasmodium vivax which resulted from Fairley's trials of mepacrine and paludrine. A thorough trial inevitably involves a close study of patients, often a large number of similar ones; it is therefore not surprising that other discoveries may be made even though they are not anticipated.

It is of interest to review the historical circumstances of the main therapeutic advances. Until the modern era little is known of the methods of investigation though the results were certainly of limited scope. Modern science dates from the sixteenth and seventeenth centuries, and such men as Bacon and Boyle recognised the need for the application of science to medicine but there was little practical progress until the eighteenth century. Therapeutic problems were then insistently presented by the growth of towns and the voyages of exploration and trade. Scurvy and smallpox were two such problems and were early subjects of clinical trials. The building of hospitals to serve the new populations provided facilities for the study of series of similar cases and the scientific approach to medicine received fresh impetus from the successes of applied science in engineering and chemistry. This spirit was developed and spread by the new provincial scientific societies and by the wealth of personal correspondence between doctors, scientists and industrial leaders. It is noteworthy that the therapeutic agents used in the eighteenth century trials were seldom of scientific origin. Inoculation, vaccination, scurvy remedies and foxglove were all derived from folk medicine. The new feature was the application of simple scientific methods to the investigation of their effects.

In the nineteenth century the same trends continued with the addition of a new source of remedies derived from other rapidly developing sciences, with greater understanding of the aetiology of disease and with improvements of technique in clinical and statistical investigation. Chemical and pharmacological research yielded the alkaloids and other new drugs, among them anaesthetics. The corresponding clinical trials were of simple type in accordance with the

physiological type of action of these substances. Later the advances of Bacteriology led to the enormous extension of Surgery by the use of antiseptic and aseptic methods; in Medicine almost as great developments of therapy followed on the studies of immunity. These advances were seldom amenable to simple 'trial and error' testing and deliberately planned trials were increasingly used.

The twentieth century has seen a further intensification of therapeutic problems, not least by two World Wars, and an increase of facilities for their study in hospital and laboratory accommodation and in research organisations. As well as the continued production of new remedies from bacteriology, there have been added contributions from Organic Chemistry such as the arsenicals and later the sulphonamides. More recently still, again from Bacteriology, the antibiotics have been introduced and some of them have already been synthesised by the chemists. There is increasing appreciation of the importance of adequate controls and in other fields new statistical methods have been developed, many of which still wait to be fully applied to clinical trials.

The stage is therefore set for great therapeutic advances in which clinical trials must play an important part. This is not to say that all features of therapy should be subjected to trial; some are too minor for the effort and expenses to be justifiable, others too complex to be feasible. A reasonable balance would seem to be the establishment of crucial principles by careful clinical trials followed by practical application of the conclusions, modified where necessary in the light of scientific and clinical knowledge. At a later stage an assessment of results may be possible which will check whether the modifications are justifiable. Further exploratory tests can be made as new treatments become available or inconsistencies appear in results, and in due course a new full scale trial will be required. In this way therapeutics should advance by a series of steps, careful trials providing at each stage a firm basis for exploratory hypotheses and experiments which can assist in determining the next advance.

SUMMARY

1. We are ignorant of the methods of trial used in the earliest therapeutic advances. The chief successes were in Surgery rather than Medicine and the type of advance suggests a simple empirical approach.
2. Greek and Roman authors outlined the theory of scientific therapeutics, but, in application, a priori assumptions predominated.
3. While little further advance was made in medieval Europe the Arabian physicians extended the knowledge of drugs.
4. The need for critical scientific approach to therapeutics was expressed by several writers in the thirteenth, sixteenth and seventeenth centuries. Paré's early trial on war wounds occurred by accident.
5. In the eighteenth century a combination of problems, personalities and appropriate milieu resulted in classical trials upon scurvy, dropsy and smallpox. The agents used were chiefly of folk origin.
6. Nineteenth century trials extended to substances derived from Chemistry and Bacteriology. Their planning was improved and their interpretation was facilitated by increased understanding of modes of action.
7. The number and scope of clinical trials has increased enormously in the present century. New therapeutic substances and procedures have been derived from all branches of science. The need for careful planning and adequate controls, and the value of statistical method have been increasingly realised.
8. The basis of comparison is crucial to the design of a trial. It may be of several types each with advantages, limitations and appropriate fields of application.
9. The choice of criteria of inclusion of cases, and of success, and the methods of statistical analysis also affect the efficiency.
10. The practical technique of applying these principles to a controlled trial is outlined.
11. The scope of clinical trials, the conditions of their development in the past and their possible success in the future are discussed.

REFERENCES

- Abraham, E. P., Chain, E., Fletcher, C. M., Gardner, A.D., Heatley, N. G., Jennings, M. A. and Florey, H. W. (1941) *Lancet* ii, 177.
- Ackerknecht, E. H. (1944) *Bull. Inst. Hist. Med.* 19, 467.
- Allbutt, T. C. (1901) *Science and Mediaeval Thought*, London.
- Andrews, F. W. (1922) (in) *History of the Great War, Medical Services, Surgery of the War*, London, vol. I, p. 151.
- Anning, S. T. (1947) *Lancet* ii, 794.
- Asher, R. (1948) *Lancet* ii, 771.
- Bacon, F. (1623) *De augmentis scientiarum* (in) *Works* (1818) London, New ed. Vol. vi, p. 153.
- Bacon, R. (c. 1268) *De erroribus medicorum*, trans. Withington, E. T. (1924) (in) *Essays on the History of Medicine presented to Karl Sudhoff*, London, p. 139.
- Banks, H. S. (1938) *Lancet* ii, 7.
- Banks, H. S. (1939) *Lancet* ii, 921.
- Banting, F. G., and Best, C. H. (1922) *J. Lab. Clin. Med.* 7, 251.
- Banting, F.G., Best, C. H., Collip, J. B., Campbell, W. R., and Fletcher, A. A. (1922) *Canad. Med. Ass. J.* 12, 141.
- Bartlett, E. (1844) *An Essay on the Philosophy of Medical Science*, Philadelphia.
- Bayliss, W. M. (1919) *Med. Res. Comm. Spec. Rep. Ser. No. 25*, London, p.11.
- Bean, W. B. (1948) *J. appl. Physiol.* I, 458.
- Behring, E., Boer, O., and Kossel, H. (1893) *Dtsch. Med. Wschr.* 19, 389 & 415.
- Bennett, J. H. (1865) *The Restorative Treatment of Pneumonia*, Edinburgh.
- Berkeley, G. (1744) *Siris; A chain of Philosophical Reflexions and Inquiries concerning the virtues of Tar Water*, London.
- Bernard, C. (1865) *Introduction a l'étude de la Médecine Expérimentale*, Paris, p. 340.
- Bishop, P. M. F., Kennedy, G. C. and Wynn Williams, G. (1948) *Lancet* ii, 764.
- Black, D. A. K. (1940) *Brit. Med. J.* ii, 693.
- Boyle, R. (1663) *The Usefulness of Natural Philosophy*, Oxford (and *Works*, 1744, ed. Birch, vol. I, p. 505.)
- Bradley, H. (1733) *Physical and Philosophical Remarks on Dr. Dover's late pamphlet*, London.
- Breasted, J. H. (1930) *The Edwin Smith Surgical Papyrus*, Chicago.
- British Museum (1930) *Introductory Guide to the Egyptian Collections*, London, pp. 71, 98.

- Brown, H. (1948) *Isis*, 39, 15.
- Bryan, C. P. (1930) *The Papyrus Ebers*, London.
- Bull, J. P. and Squire, J. R. (1949) *Ann. Surg.* 130, 160.
- Bull, J. P. (1949) Unpublished.
- Cannon, B. and Cope, C. (1943) *Ann. Surg.* 117, 85.
- Cajori, F. (1929) *A History of Mathematics*, New York, p. 381.
- Cecil, R. L. (1928) *Arch. Intern. Med.* 41, 295.
- Celsus, A. C. *De re medicina*, trans. Spencer, W. G. (1935) London.
- Chick, H., Dalyell, E.J., Hume, E. M., Mackay, H. M. M., Smith, H. H. and Wimberger, H. (1922) (in) *Med. Res. Coun. Spec. Rep. Ser. No. 77*, London, p. 19.
- Chick, H. and Dalyell, E. J. (1921) *Brit. Med. J.* ii. 1064.
- Christie, R. V. (1946) *Lancet* i, 369.
- Christie, R. V. (1948) *Brit. Med. J.* i, 1.
- Clark, A. J. (1933) *The mode of action of drugs on cells*, London, p. 3.
- Clark-Kennedy, A. H. (1929) *Stephen Hales*, Cambridge, p. 215.
- Clayton-Cooper, B., and Williams, R. E. O. (1945) *Brit. J. industry. Med.*, 2, 146.
- Cobbett, W. (1800) *The Rush Light*, New York.
- Cochrane, A. L. (1950) *Proceedings of the Medical Research Council's Conference on Industrial and Service medical research (to be published)*.
- Cochrane, A. L., Chapman, P. J., and Oldham, P. D. (1950) *To be published*.
- Colebrook, L. and Kenny, M. (1936) *Lancet* I, 1279, 1319.
- Colebrook, L. and Purdie, A. W. (1937) *Lancet* ii, 1237, 1291.
- Cook, J. (1776) *Philos. Trans.* 66, 402.
- Creighton, C. (1894) *A History of Epidemics in Britain*, Cambridge, vol. ii, p. 725.
- Cumston, C. G. (1926) *An Introduction to the History of Medicine*, London, P. 372.
- Dampier-Whetham, W. C. D. (1930) *A History of Science*, Cambridge, 2nd ed., p. 106.
- Dawson, W. R. (1930) *The Beginnings; Egypt and Assyria*, New York.
- Ding, E. (1943) *Z. Hyg. Infektkr.* 124, 670.
- Dogmagk, G. (1935) *Dtsch. Med. Wschr.* 61, 250.
- Dover, T. (1733) *The Ancient Physician's Legacy to his Country*, London.
- Drummond, J. C. and Wilbraham, A. (1939) *The Englishman's Food*, London.

Drummond, H., and Taylor, E. S. (1919) (in) *Med. Res. Comm. Spec. Rep. Ser. No. 25*, London, p. 135.

Duncum, B. H. (1947) *The Development of Inhalation Anaesthesia*, London.

Edwards, D. A. W. and Swyer, G. I. M. (1950) *Clin. Sci.* 9, 115.

Evans, G. M. and Gaisford, W. F. (1938) *Lancet*, ii, 14.

Fairley, N. H. (1945) *Trans. R. Soc. trop. Med. Hyg.* 38, 311.

Fairley, N. H. (1946) *Trans. R. Soc. trop. Med. Hyg.* 40, 105.

Ferguson, F. R., Davey, A. F. C., and Topley, W. W. C. (1927) *J. Hyg. Camb.* 26, 98.

Fibiger, J. (1898) *Hospitalstidende* 4:6, 309, 337.

Finney, D. J. (1947) *Probit Analysis*, London.

Fisher, R. A. (1935) *The Design of Experiments*, London.

Fisher, R. A., and Yates, F. (1948) *Statistical tables for Biological, Agricultural and Medical Research*, 3rd ed., London, pp. 8, 21.

Florey, M. E., and Florey, H. W. (1943) *Lancet* I, 387.

Fourman, L. P. R. (1950) To be published.

Franklin, B. (1757) Letter to John Pringle (in) *Benjamin Franklin's Experiments*, ed. Cohen, I. B. (1941) Harvard.

Gaddum, J. H. (1945) *Nature*, 156, 463.

Gaddum, J. H. (1948) *Pharmacology*, London, 3rd ed., p. 463.

Garrison, F. H. (1922) *An Introduction to the History of Medicine*, Philadelphia, 3rd ed.

Gibson, R. J. H. (1919) *Outlines of the History of Botany*, London, p. 14.

Gordon J. (1942) *Ann. med. Hist.*, 4, 219.

Green, H. N., Pinder, D., Davis, G., and Mellanby, E. (1931) *Brit. Med. J.*, ii, 595.

Greenwood, M. (1935) *Epidemics and Crowd Diseases*, London, p. 83.

Greenwood, M. and Yule, G. U. (1915) *Proc. Roy. Soc. Med., Sec. epidem. and state Med.*, 8, 113.

Gunther, R. T. (1933) *The Greek Herbal of Dioscorides*, Oxford.

Haggard, H. W. (1932) *The Lame, The Halt and The Blind*, London, p. 35.

Haggis, A. W. (1941) *Bull. Inst. Hist. Med.* 10, 417.

Hales, S. (1740) *An Account of some Experiments and Observations on Mrs. Stephen's Medicines for dissolving the Stone*, London.

Hales, S. (1747) *An Account of some Experiments and Observations on Tar Water*, London.

- Harington, J. (1607) *The School of Salernum; Regimen sanitatis salernitarum* (Reprinted (1920) New York).
- Harris, K. E. (1927) *Heart*, 14, 283.
- Harrison, S. H., Topley, E., and Lenmard-Jones, J., (1949) *Lancet* I, 425.
- Harvey, W. (1628) *De motu cordis*, Frankfurt, (New ed. Keynes, G. (1928) London).
- Harvey, W. F. and Acton, H. W. (1922) *Indian J. med. Res.* 9, 852.
- Haygarth, J. (1801) *Of the Imagination as a cause and as a cure of disorders of the body*, Bath, New ed. p. 3.
- Heberden, W. (1745) *Antithriaka, An Essay on Mithridatium and Theriaca*, London.
- Herodotus *Clio* 197, trans. Cary, H. (1904) London, p. 86.
- Higgins, G., O'Brien, J. R. P., Peters, R. A., Stewart, A., and Witts, L. J. (1945) *Brit. med. J.* i, 401.
- Hill, A. B. (1937) *Principles of Medical Statistics*, London, p. 5.
- Hippocrates *Regimen in Acute Diseases* (in) *Works* trans. Jones, W. H. S. (1923) London, vol. ii, p. 63.
- Hippocrates *Precepts* (in) *Works* trans. Jones, W. H. S. (1923) London, vol. i, p. 313.
- Holmes, O. W. (1860) *Currents and Countercurrents in Medical Science* (in) *Works* (1891) vol. ix, p. 185.
- Hunter, J. (1794) *A Treatise on the Blood, Inflammation and Gun-Shot Wounds*, London, p. 531.
- Ivy, A. C. (1948) *Science* 108, 1.
- Jastrow, M. (1917) *Ann. med. Hist.* I, 231.
- Jenner, E. (1798) *An Inquiry into the Causes and Effects of the Variolae Vaccinae*, London.
- Jochmann, G. (1913) *Pocken and Vaccinationslehre*, Leipsig, p. 257.
- Laplace, P-S. (1812) *Théorie Analytiques des Probabilités*, Paris.
- LaWall, C. H. (1927) *Four Thousand years of Pharmacy*, Philadelphia.
- League of Nations Malaria Commission (1933) *Bull. health Organ., League of Nations*, 2, 181.
- League of Nations Malaria Commission (1937) *Bull. health Organ., League of Nations*, 6, 895.
- Lennox, W. G., and Storch, T.J. C. von (1935) *J. Amer. med. Ass.* 105, 169.
- Lewis, T. (1930) *Heart* 15, 305.
- Lilley, S. (1948) *Men, Machines and History*, London.

- Lind, J. (1753) *A Treatise of the Scurvy*, Edinburgh.
- Lister, J. (1870) *Lancet* I, 4, 40.
- Long, P. H. and Bliss, E. A. (1939) *The Clinical and Experimental Use of Sulfanilamide, Sulfapyridine and Allied Compounds*, New York, p. 3.
- Louis, P. C. A. (1834) *Essay on Clinical Instruction*, trans. Martin, London, pp. 26, 27, 28.
- Louis, P. C. A. (1835) *Recherches sur les effets de la saignée*, Paris.
- Louis, P. C. A. (1843) *Researches on Phthisis*, 2nd ed., trans. Walshe, W. H. (1844) London, pp. 551, 552.
- Lowbury, E. J. L. and Topley, E. (1950) To be published.
- MacCurdy, E. (1938) *The Notebooks of Leonardo da Vinci*, London.
- Mackenzie, J. (1908) *Diseases of the Heart*, London, pp. 269, 334.
- Mackworth, N. H. and Winson, F. H. (1947) *Royal Naval Personnel Res. Comm. R.N.P. 47/401*.
- Magendie, F. (1822) *Formulaire pour la Préparation et l'Emploi de plusieurs Nouveaux Médicaments*, Paris, 3rd. Ed., Preface, p. vi.
- Maitland, C. (1722) *Account of inoculating the Small Pox*, London.
- Marshall, E. K. (1947) *Medicine*, Baltimore, 26, 155.
- Mettler, C. C. (1947) *History of medicine*, Toronto, p. 190.
- Minot, G. R. and Murphy, W. P. (1926) *J. Amer. med. Ass.* 87, 470.
- Moore, J. E. Stanton, M. H., Robinson, H. M., and Eagle, H. (1936) *Amer. J. Syph.* 20, 503.
- Murray, G. M. (1891) *Brit. med. J.* ii, 796.
- Neuberger, M. (1910) *History of Medicine*, trans. Playfair, E., London vol. i, p. 389.
- Osler, W. (1900) *Lancet* ii, 1115.
- Osler, W. (1908) *An Alabama Student, and other biographical essays*, Oxford.
- O'Sullivan, M. E. (1936) *J. Amer. med. Ass.* 107, 1208.
- Packard, F. R. (1925) *The Life and Times of Ambroise Paré*, New York, 2nd ed., pp. 27, 163.
- Park, W. H., Bullowa, J. G. M., and Rosenbluth, M. B. (1928) *J. Amer. med. Ass.* 91, 1503.
- Parker, G (1920) *The Early History of Surgery in Great Britain*, London.
- Payne, J. F. (1900) *Thomas Sydenham*, London, p. 152.
- Pearson, G. (1798) *An Inquiry concerning the History of Cowpox*, London.
- Pearson, H. (1930) *Doctor Darwin*, London.

- Pearson, J. (1800) *Observations on the Effects of various Articles of the Materia Medica in the cure of Lues Venerea*, illustrated with cases, London, Introd. p. 1.
- Pearson, K (1894) *Philos. Trans. Ser. A.* 185, 171.
- Pearson, K. (1900) *Phil. Mag.* 50, 157.
- Porritt, A. E., and Mitchell, G. A. G. (1945) *An investigation into the prophylaxis and treatment of wound Infection (in) Penicillin Therapy and a control in 21 Army Group*, London. p. 7.
- Pulvertaft, R. J. V. (1943) *Lancet* ii, 341.
- Ranke, H. (1933) *Bull. Inst. Hist. Med.* 1, 237.
- Richter, J. P. (1939) *Leonardo da Vinci*, Oxford, vol., ii, p. 239.
- Rivers, W. H. R. (1924) *Medicine, Magic and Religion*, London, p. 81.
- Roddie, L. H. (1936) *Ann. Med. Hist.* 8, 93 and 185.
- Rogers, L. (1912) *Brit. Med. J. I*, 1424.
- Ross, W. P. D. 1950 to be published.
- Roux, E., Martin, L., and Chaillou, A., (1894) *Ann. Inst. Pasteur*, 8, 640.
- Ryle, J. A. (1947) *Lancet* i, 1.
- Ryle, J. A. (1948) *Changing Disciplines*, Oxford, p. 104.
- Scott, H. H. (1939) *A History of Tropical Medicine*, London, p. 223.
- Sheehan, H. L. and Ferguson, A. G. (1943) *Lancet* I, 547.
- Sherrington, C. S. (1946) *The Endeavour of Jean Pernel*, London.
- Sholly, A. I. von, and Park, W. H. (1921) *J. Immunol.* 6, 103.
- Shryock, R. H. (1948) *The Development of Modern Medicine*, London, p. 24.
- Singler, C. (1928) *From Magic to Science*, London, p. 59.
- Snodgrass, W. R. and Anderson, T. (1937) *Brit. med. J.* ii, 101.
- Snow, J. (1858) *on Chloroform and other Anaesthetics*, London.
- Statistical Research Group (1945) *Sequential Analysis in Inspection and Experimentation*, New York.
- Statistical Research Group (1947) *Selected Techniques of Statistical Analysis*, New York, p. 363.
- Streptomycin in Tuberculosis Trials Committee (1948) *Lancet* i, 582.
- “Student” (1908) *Biometrika*, 6,1.
- Sutton, H. G. (1865) *Guy’s Hosp. Rep.* ii, 392.
- Takaki, K. (1906) *Lancet* i, 1369.
- Therapeutic Trials Committee of Pharmacy and Chemistry (1949) *J. Amer. med. Ass.* 140, 534.

- 'Times' (1881) London, June 3, p. 5.
- Topley, W. W. C. (1933) An outline of Immunity, London, p. 328.
- Trotter, W. (1930) Brit. med. J. ii, 129.
- Vallery-Radot, R. (1922) Pasteur, 1822-1895, Paris.
- Wagle, P. M., Sokhey, S. H., Dikshit B. B. and Ganapathy, K. (1941) Ind. Med. Gaz. 76, 29.
- Waterhouse, B. (1800) A Prospect of Exterminating the Small-Pox, Boston.
- Wayne, E. J. and Laplace, L. B. (1933) Clin. Sci. 1,103.
- Webster, A. L. (1947) Lancet ii, 610.
- Wechselmann, W. 1910 Berl. klin. Wochr. 47, 1261.
- Whitfield, J. W. (1946) Accidents to Coal Miners, Reports to Medical Research Council M.R.C. 47/226, A.P.U. 36 and M.R.C. 47/383 A.P.U. 43.
- Whitfield, J. H. (1947) Biometrika 34, 292.
- Whitfield, J. W. (1950) Brit. med. Bull. (in the press).
- Williams, R. E. O. and Miles, A. A. (1949) Med. Res. Coun. Spec. Rep. Ser. No. 266, p. 53.
- Wilson, C., Pollock, M. R., and Harris, A. D. (1945) Brit. med. J. i, 399.
- Wiseman, R. (1676) Severall Chirurgicall Treatises, London, p. 123.
- Withering, W. (1785) An Account of the Foxglove and some of its Medical Uses with Practical Remarks on Dropsy and other diseases, Birmingham.
- Withington, E. T. (1854) Medical History from the Earliest Times, London.
- Woodville, W. (1799) Reports of a Series of Inoculations for the Variolae Vacinae or Cowpox, London.
- Wright, A. E. (1900) Brit. med. J. i, 122.
- Wright, A. E. (1937) On Induction in General (in) Studies on Immunisation 2nd Series (1944) London, p. 229.
- Wright, A. E. (1941) Prolegomena to the Logic which searches for Truth, London.
- Yule, G. U. and Kendall, M. G. (1937) An Introduction to the Theory of Statistics 11th ed., London, p. 186.