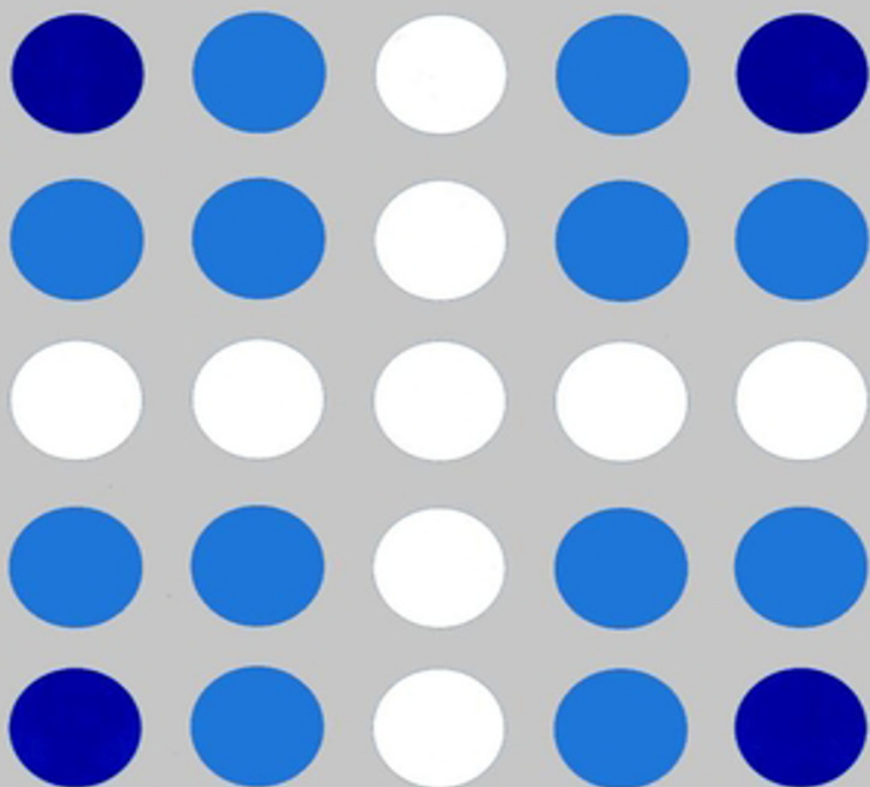


Stuart J. Pocock

CLINICAL TRIALS



A Practical Approach

CLINICAL TRIALS

A Practical Approach

CLINICAL TRIALS

A Practical Approach

STUART J. POCOCK

*Professor of Medical Statistics and Director of Clinical Trials Research Group
London School of Hygiene and Tropical Medicine
University of London*

JOHN WILEY & SONS

Chichester · New York · Brisbane · Toronto · Singapore

Copyright © 1983 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyurope.com or www.wiley.co.uk

Reprinted June 1984, September 1985, June 1986, April 1987, July 1988, June 1989,
May 1990, January 1991, June 1991, December 1991, October 1993, December 1994,
August 1996, November 1996, November 1997, August 1998, May 1999, November 2000
March 2002, April 2004, April 2006, September 2006, May 2007, January 2008,
December 2008, December 2009, November 2010

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770571.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 978 0 471 90155 6 (HB)

Typeset in Times New Roman by Eta Services Ltd, Beccles, Suffolk
Printed and bound in Great Britain by TJ International Ltd, Padstow, Cornwall

*To
Peter Armitage and Marvin Zelen
with gratitude.*

Contents

Preface	xi
1. Introduction: The Rationale of Clinical Trials	1
1.1 Types of clinical trial.	1
1.2 Controlled clinical trials and the scientific method	4
1.3 An example of a clinical trial for primary breast cancer	7
2. The Historical Development of Clinical Trials	14
2.1 Clinical trials before 1950	14
2.2 Clinical trials since 1950	18
2.3 Cancer chemotherapy in the United States	21
2.4 Treatment of acute myocardial infarction	24
2.5 The pharmaceutical industry	26
3. Organization and Planning.	28
3.1 The protocol	28
3.2 Administration, staff and finance	31
3.3 Selection of patients	35
3.4 Treatment schedules	38
3.5 Evaluation of patient response	41
4. The Justification for Randomized Controlled Trials.	50
4.1 Problems with uncontrolled trials	51
4.2 Problems with historical controls	54
4.3 Problems with concurrent non-randomized controls	60
4.4 Is randomization feasible?	63
5. Methods of Randomization.	66
5.1 Patient registration	66
5.2 Preparing the randomization list	73
5.3 Stratified randomization	80
5.4 Unequal randomization	87

6. Blinding and Placebos	90
6.1 The justification for double-blind trials	90
6.2 The conduct of double-blind trials	93
6.3 When is blinding feasible?	97
7. Ethical Issues	100
7.1 Medical progress and individual patient care.	100
7.2 Informed patient consent	105
8. Crossover Trials	110
8.1 Within-patient comparisons	110
8.2 The two-period crossover design	112
8.3 The analysis and interpretation of crossover trials	114
8.4 Multi-period crossover designs	119
9. The Size of a Clinical Trial	123
9.1 Statistical methods for determining trial size	123
9.2 The realistic assessment of trial size	130
9.3 The inadequacy of small trials	133
9.4 Multi-centre trials	134
9.5 The number of treatments and factorial designs	138
10. Monitoring Trial Progress	142
10.1 Reasons for monitoring	142
10.2 Interim analyses	143
10.3 Repeated significance testing: group sequential designs	147
10.4 Continuous sequential designs	155
11. Forms and Data Management	160
11.1 Form design	160
11.2 Data management	166
11.3 The use of computers	168
12. Protocol Deviations	176
12.1 Ineligible patients	176
12.2 Non-compliance and incomplete evaluation	179
12.3 Inclusion of withdrawals in analysis	182
13. Basic Principles of Statistical Analysis	187
13.1 Describing the data	188
13.2 Significance tests	197
13.3 Estimation and confidence limits	206

14. Further Aspects of Data Analysis	211
14.1 Prognostic factors.	211
14.2 The analysis of survival data	221
14.3 Multiplicity of data	228
15. Publication and Interpretation of Findings.	234
15.1 Trial reports and their critical evaluation	234
15.2 An excess of false-positives.	239
15.3 Combining evidence and overall strategy	242
References	248
Index	257

Preface

There is an ever-increasing number of treatment innovations which require proper investigation to see if they are of genuine benefit to patients. The randomized controlled clinical trial has become widely regarded as the principal method for obtaining a reliable evaluation of treatment effect on patients. The purpose of this book is to explain in practical terms the basic principles of clinical trials. Particular emphasis is given to their scientific rationale, including the relevance of statistical methods, though ethical and organizational issues are also discussed in some detail.

My intention has been to present the methodology of clinical trials in a style which is comprehensible to a wide audience. I hope the book proves to be especially useful to clinicians and others who are involved in conducting trials and it would be particularly gratifying if this text encouraged more clinicians to undertake or collaborate in properly designed trials to resolve relevant therapeutic issues.

Pharmaceutical companies have a fundamental role in the organization of trials for drug therapy. I have tried to give a balanced view of their activities in this area and hope that my approach to clinical trials is conducive to maintaining high standards of research in the clinical testing of new drugs. However, I wish to emphasize that randomized controlled trials should also be applied to assessing other (non-drug) aspects of therapy and patient management.

The practice of medicine poses a need to interpret wisely the published findings from clinical trials. Accordingly, the medical profession at large and others concerned with the treatment and management of patients may benefit from an increased understanding of how clinical trials are (and should be) conducted.

The proper use of statistical methods is important at the planning stage of a clinical trial as well as in the analysis and interpretation of results. I also recognize that many clinicians and others without mathematical training experience some difficulty in understanding statistical concepts. Hence, I have used a straightforward non-mathematical approach in describing those statistical issues that I consider of relevance to the practice of clinical trials. In particular,

I would like to think that the basic principles of statistical analysis described in chapter 13 may be of more general interest beyond clinical trials. Indeed, some readers who are unfamiliar with statistical terms may find it instructive to begin with this chapter.

My own experience in teaching undergraduate medical students has encouraged me to believe that the introduction of clinical trials and related statistical ideas is a useful aspect of preclinical education. Accordingly, my approach to such courses is reflected in much of this book.

As a medical statistician I believe that clinical trials require a successful collaboration of clinical, organizational and statistical skills. I feel that my profession needs to strive harder to achieve effective communication of our ideas to non-statistical colleagues and I would be delighted if this book could persuade other statisticians towards a commonsense and less theoretical approach to medical research. In this respect, students of biostatistics may find this book a useful antidote to their more mathematical courses!

Lastly, my policy has been always to introduce each concept via actual examples of clinical trials. In this way, the reader should experience the reality of clinical trials, not as an abstract collection of methods, but as a practical contribution to furthering medical knowledge.

I greatly appreciate the contributions of Sheila Gore and Austin Heady who read the book in draft and made many suggestions for improvement. I am also grateful to Tom Meade and Simon Thompson for their helpful comments on the draft. I am indebted to Peter Armitage for first stimulating the publishers to realize the need for such a book. I wish to express sincere thanks to Yvonne Ayton for typing the manuscript and to other colleagues for their invaluable support. Lastly, this whole project was made easier by the help and encouragement of my wife Faith.

CHAPTER 1

Introduction: The Rationale of Clinical Trials

The evaluation of possible improvements in the treatment of disease has historically been an inefficient and haphazard process. Only in recent years has it become widely recognized that properly conducted clinical trials, which follow the principles of scientific experimentation, provide the only reliable basis for evaluating the efficacy and safety of new treatments. The major objective of this book is therefore to explain the main scientific and statistical issues which are vital to the conduct of effective and meaningful clinical research. In addition, some of the ethical and organizational problems of clinical trials will be discussed. The historical perspective, current status and future strategy for clinical trials provide a contextual framework for these methodological aspects.

In section 1.1, I discuss what constitutes a clinical trial and how clinical trials may usefully be classified. Section 1.2 deals with the underlying rationale for randomized controlled clinical trials and their relation to the scientific method. Section 1.3 goes on to describe one particular example, a clinical trial for primary breast cancer, as an illustration of how adherence to sound scientific principles led to an important advance in treatment.

1.1 TYPES OF CLINICAL TRIAL

Firstly, we need to define exactly what is meant by a 'clinical trial': briefly the term may be applied to any form of *planned experiment* which involves patients and is designed to elucidate the most appropriate treatment of future patients with a given medical condition. Perhaps the essential characteristic of a clinical trial is that one uses results based on a limited *sample* of patients to make inferences about how treatment should be conducted in the *general population* of patients who will require treatment in the future.

Animal studies clearly do not come within this definition and experiments on healthy human volunteers are somewhat borderline in that they provide only indirect evidence of effects on patients. However, such *volunteer studies* (often

termed phase I trials) are an important first step in human exposure to potential new treatments and hence are included in our definition when appropriate.

Field trials of vaccines and primary prevention trials for subjects with presymptomatic conditions (e.g. high serum cholesterol) involve many of the same scientific and ethical issues as in the treatment of patients who are clearly diseased, and hence will also be mentioned when appropriate.

An individual *case study*, whereby one patient's pattern of treatment and response is reported as an interesting occurrence, does not really constitute a clinical trial. Since biological variation is such that patients with the same condition will almost certainly show varied responses to a given treatment, experience in one individual does not adequately enable inferences to be made about the general prospects for treating future patients in the same way. Thus, clinical trials inevitably require *groups of patients*: indeed one of the main problems is to get large enough groups of patients on different treatments to make reliable treatment comparisons.

Another issue concerns *retrospective surveys* which examine the outcomes of past patients treated in a variety of ways. These unplanned observational studies contain serious potential biases (e.g. more intensive treatments given to poorer prognosis patients may appear artificially inferior) so that they can rarely make a convincing contribution to the evaluation of alternative therapies. Hence, except in chapter 4 when considering the inadequacies of non-randomized trials, such studies will not be considered as clinical trials.

It is useful at this early stage to consider various ways of classifying clinical trials. Firstly, there is the *type of treatment*: the great majority of clinical trials are concerned with the evaluation of drug therapy more often than not with pharmaceutical company interest and financial backing. However, clinical trials may also be concerned with other forms of treatment. For instance, surgical procedures, radiotherapy for cancer, different forms of medical advice (e.g. diet and exercise policy after a heart attack) and alternative approaches to patient management (e.g. home or hospital care after inguinal hernia operation) should all be considered as forms of treatment which may be evaluated by clinical trials. Unfortunately, there has generally been inadequate use of well-designed clinical trials to evaluate these other non-pharmaceutical aspects of patient treatment and care, a theme which I shall return to later.

Drug trials within the pharmaceutical industry are often classified into four main phases of experimentation. These four phases are a general guideline as to how the clinical trials research programme for a new treatment in a specific disease might develop, and should not be taken as a hard and fast rule.

Phase I Trials: Clinical Pharmacology and Toxicity

These first experiments in man are primarily concerned with drug safety, not efficacy, and hence are usually performed on human volunteers, often pharmaceutical company employees. The first objective is to determine an acceptable single drug dosage (i.e. how much drug can be given without causing

serious side-effects). Such information is often obtained from dose-escalation experiments, whereby a volunteer is subjected to increasing doses of the drug according to a predetermined schedule. Phase I will also involve studies of drug metabolism and bioavailability and, later, studies of multiple doses will be undertaken to determine appropriate dose schedules for use in phase II. After studies in normal volunteers, the initial trials in patients will also be of the phase I type. Typically, phase I studies might require a total of around 20–80 subjects and patients.

Phase II Trials: Initial Clinical Investigation for Treatment Effect

These are fairly small-scale investigations into the effectiveness and safety of a drug, and require close monitoring of each patient. Phase II trials can sometimes be set up as a screening process to select out those relatively few drugs of genuine potential from the larger number of drugs which are inactive or over-toxic, so that the chosen drugs may proceed to phase III trials. Seldom will phase II go beyond 100–200 patients on a drug.

Phase III Trials: Full-scale Evaluation of Treatment

After a drug is shown to be reasonably effective, it is essential to compare it with the current standard treatment(s) for the same condition in a large trial involving a substantial number of patients. To some people the term ‘clinical trial’ is synonymous with such a full-scale phase III trial, which is the most rigorous and extensive type of scientific clinical investigation of a new treatment. Accordingly, much of this book is devoted to the principles of phase III trials.

Phase IV Trials: Postmarketing Surveillance

After the research programme leading to a drug being approved for marketing, there remain substantial enquiries still to be undertaken as regards monitoring for adverse effects and additional large-scale, long-term studies of morbidity and mortality. Also the term ‘phase IV trials’ is sometimes used to describe promotion exercises aimed at bringing a new drug to the attention of a large number of clinicians, typically in general practice. This latter type of enquiry has limited scientific value and hence should not be considered part of clinical trial research.

This categorization of pharmaceutical company sponsored drug trials is inevitably an oversimplification of the real progress of a drug’s clinical research programme. However, it serves to emphasize that there are important early human studies (phases I/II), with their own particular organizational, ethical and scientific problems, which need to be completed before full-scale phase III trials are undertaken. The Food and Drug Administration (1977) have issued

guidelines for drug development programmes in the United States. The guidelines include recommendations on how phase I–III trials should be structured for drugs in 15 specific disease areas.

It should be remembered that each pharmaceutical company has an equally important *preclinical research* programme, which includes the synthesis of new drugs and animal studies for evaluating drug metabolism and later for testing efficacy and especially potential toxicity of a drug. The scale and scientific quality of these *animal experiments* have increased enormously, following legislation in many countries prompted by the thalidomide disaster. In particular any drug must pass rigorous safety tests in animals before it can be approved for clinical trials.

The phase I–III classification system may also be of general guidance for clinical trials not related to the pharmaceutical industry. For instance, cancer chemotherapy and radiotherapy research programmes, which take up a sizeable portion of the U.S. National Institutes of Health funding, can be conveniently organized in terms of phases I–III. In this context, phase I trials are necessarily on patients, rather than normal volunteers, due to the highly toxic nature of the treatments.

Development of new surgical procedures will also follow broadly similar plans, with phase I considered as basic development of surgical techniques. However, there is a paucity of well-designed phase III trials in surgery.

1.2 CONTROLLED CLINICAL TRIALS AND THE SCIENTIFIC METHOD

I will now concentrate on full-scale (phase III) trials and consider the scientific rationale for their conduct. Of course, the first priority for clinical research is to come up with a good idea for improving treatment. Progress can only be achieved if clinical researchers with insight and imagination can propose therapeutic innovations which appear to have a realistic chance of patient benefit. Naturally, the proponents of any new therapy are liable to be enthusiastic about its potential: preclinical studies and early phase I/II trials may indicate considerable promise. In particular, a pharmaceutical company can be very persuasive about its product before any full-scale trial is undertaken. Unfortunately, many new treatments turn out not to be as effective as was expected: once they are subjected to the rigorous test of a properly designed phase III trial many therapies fail to live up to expectation; see Gilbert *et al.* (1977) for examples in surgery and anaesthesia.

One fundamental rule is that *phase III trials are comparative*. That is, one needs to compare the experience of a group of patients on the new treatment with a *control group* of similar patients receiving a standard treatment. If there is no standard treatment of any real value, then it is often appropriate to have a control group of untreated patients. Also, in order to obtain an unbiased evaluation of the new treatment's value one usually needs to assign each patient

randomly to either new or standard treatment (see chapters 4 and 5 for details). Hence it is now generally accepted that the *randomized controlled trial* is the most reliable method of conducting clinical research.

At this point it is of value to present a few examples of randomized controlled trials to illustrate the use of control groups. Table 1.1 lists the six trials I wish to consider.

The first trial, for bacterial meningitis, represents the straightforward situation where a new treatment (cefuroxime) was compared with a standard treatment (the combination of ampicillin and chloramphenicol) to see if the former was more effective in killing the bacterium.

The anturan trial reflects another common situation where the new treatment (anturan) is to be compared with a placebo (inactive oral tablets that the patients could not distinguish from anturan). Thus, the control group of myocardial infarction patients did not receive any active treatment. The aim was to see if anturan could reduce mortality in the first year after an infarct.

The mild hypertension trial has two active treatments which are to be compared with placebo to see if either can reduce morbidity and mortality from cardiovascular-renal causes.

The trial for advanced colorectal cancer is unusual in having three new treatments to compare with the standard drug 5-fluorouracil (5-FU). Two of the new treatments consisted of 5-FU in combination with other drugs. Most trials have just two treatment groups (new *vs.* standard) and in general one needs to be wary of including more treatments since it becomes more difficult to get sufficient patients per treatment.

The last two trials in Table 1.1 are included as reminders that clinical trials can be used to evaluate aspects of treatment other than drug therapy. The stroke trial is concerned with patient management: can one improve recovery by caring for patients in a special stroke unit rather than in general medical wards?

The breast cancer trial represents an unusual situation in that it set out to compare two treatments (radical mastectomy or simple mastectomy + radiotherapy) each of which is standard practice depending on the hospital. In a sense each treatment is a control for the other. Such trials can be extremely important in resolving long-standing therapeutic controversies which have previously never been tested by a randomized controlled trial.

I now wish to consider how a clinical trial should proceed if the principles of the scientific method are to be followed. Figure 1.1 shows the general sequence of events. From an initial idea about a possible improvement in therapy one needs to produce a more precise definition of trial aims in terms of specific hypotheses regarding treatment efficacy and safety. That is, one must define exactly the type of patient, the treatments to be compared and the methods of evaluating each patient's response to treatment.

The next step is to develop a detailed design for a randomized trial and document one's plan in a study protocol. The design needs to fulfil scientific, ethical and organizational requirements so that the trial itself may be conducted efficiently and according to plan. Two principal issues here are:

Table 1.1. Some examples of randomized controlled trials

Reference	Disease	Treatments (control group in italics)*
Swedish Study Group (1982)	Bacterial meningitis	cefuroxime v. <i>ampicillin + chloramphenicol</i>
Anturane Reinfarction Trial (1980)	Acute myocardial infarction	anturan v. <i>placebo</i>
Medical Research Council Working Party (1977)	Mild hypertension	bendrofluazide v. <i>propranolol v. placebo</i>
Douglass <i>et al.</i> (1978)	Advanced colorectal cancer	methyl-CCNU v. $\begin{matrix} 5-FU \\ + \\ TG \end{matrix}$ v. $\begin{matrix} 5-FU \\ + \\ CTX \end{matrix}$ v. <i>5-FU</i>
Garraway <i>et al.</i> (1980)	Acute stroke	special stroke unit v. <i>general medical ward</i>
Langlands <i>et al.</i> (1980)	Operable breast cancer	<i>radical mastectomy</i> v. <i>simple mastectomy + radiotherapy</i>

* 5-FU = 5-fluorouracil, TG = 6-thioguanine, CTX = cytoxan

- (a) *Size* The trial must recruit enough patients to obtain a reasonably precise estimate of response on each treatment.
- (b) *Avoidance of bias* The selection, ancillary care and evaluation of patients should not differ between treatments, so that the treatment comparison is not affected by factors unrelated to the treatments themselves.

Statistical methods should be applied to the results in order to test the prespecified study hypotheses. In particular, one may use significance tests to assess how strong the evidence is for a genuine difference in response to treatment. Finally, one needs to draw conclusions regarding the treatments' relative merits and publish the results so that other clinicians may apply the findings.

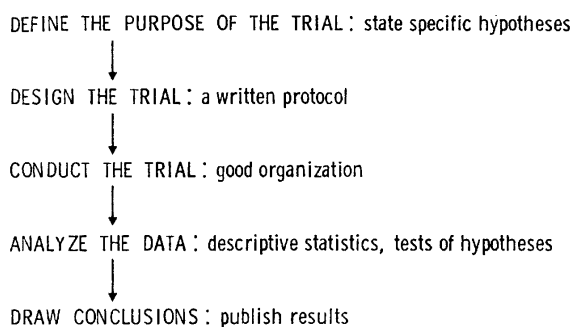


Fig. 1.1. The scientific method as applied to clinical trials

The aim of any clinical trial should be to obtain a truthful answer to a relevant medical issue. This requires that the conclusions be based on an unbiased assessment of objective evidence rather than on a subjective compilation of clinical opinion. Historically, progress in clinical research has been greatly hindered by an inadequate appreciation of the essential methodology for clinical trials. After a brief historical review in chapter 2, the remainder of this book is concerned with a more extensive and practical account of this methodology. As a useful introduction to the main concepts, I now wish to focus on one particular trial for primary breast cancer.

1.3 AN EXAMPLE OF A CLINICAL TRIAL FOR PRIMARY BREAST CANCER

In 1972 a clinical trial was undertaken in the United States to evaluate whether the drug L-Pam (1-phenylalanine mustard) was of value in the treatment of primary breast cancer following a radical mastectomy. Fisher *et al.* (1975) presented the early findings with a subsequent update by Fisher *et al.* (1977). We now consider the development of this trial in the context of the scientific method outlined in figure 1.1.

(1) Purpose of the Trial

Earlier clinical trials for the treatment of patients with advanced (metastatic) breast cancer had shown that L-Pam was one of a number of drugs which could cause temporary shrinkage of tumours and increase survival in some patients. Therefore, it seemed sensible to argue that for patients with primary breast cancer who might still have an undetected small trace of tumour cells present after mastectomy, a drug such as L-Pam could be effective in killing off such cells and hence preventing subsequent disease recurrence. Such a *general concept* is an essential preliminary for a worthwhile clinical trial, but more precise *specific hypotheses* must be defined before a trial can be planned properly. There are four basic issues in this regard: the precise definition of (1) the *patients eligible* for study, (2) the *treatment*, (3) the *end-points* for evaluating each patient's response to treatment, and (4) the need for comparison with a *control group* of patients not receiving the new treatment. In this case these four issues were resolved as follows:

Eligible patients were defined as having had a radical mastectomy for primary breast cancer with histologically confirmed axillary node involvement. Patients were excluded if they had certain complications such as peau d'orange, skin ulceration, etc., or if they were aged over 75, were pregnant or lactating. Thus the trial focussed on those patients who were considered most likely to benefit from L-Pam if indeed it conferred any benefit at all.

Treatment was defined as L-Pam to be given orally at a dose of 0.15 mg/kg body weight for five consecutive days every six weeks, this dose schedule having been well established from studies in advanced breast cancer. Since haematologic toxicity will occur in some patients, *dose modifications* were defined as follows: reduce dose by half if platelet count $< 100\,000$ or white cell count < 4000 , and discontinue drug while platelet count $< 75\,000$ or white cell count < 2500 . For patients without toxicity after three consecutive courses, dosage was increased to 0.20 mg/kg. L-Pam was to be started less than four weeks after the patient's radical mastectomy and continued until treatment failure or for two years, whichever occurred first.

End-points for evaluating treatment were the disease-free interval (i.e. the time from mastectomy until first detection of tumour in local, regional or distant sites), the survival time (i.e. time from mastectomy until death) and also patient toxicity (haematologic and also nausea/vomiting). Disease-free interval would be the main criterion (that is, what percentages of patients were still alive and disease free after one year, two years, etc.), since there would not be many deaths in the first few years of follow-up and toxic effects were reasonably well known from studies in advanced disease.

A *control group* of patients would need to be treated in a standard way: that is, a separate group of patients just as eligible for the study would need to have a radical mastectomy but no subsequent L-Pam. They should then be followed in the same way to allow comparison of the percentages disease free in the

treatment group and control group after one year, after two years, etc. Exactly how such a control group can be arranged is described in the design section to follow.

After the above clarifications, one is in a position to state the *main hypothesis* under study: Does L-Pam (as defined above) prolong the disease-free interval of primary breast cancer patients (as defined above) if given after a radical mastectomy?

Several subsidiary hypotheses concerning patient survival, toxicity and whether any increase in disease free interval is confined to particular subgroups of patients (e.g. premenopausal) are also to be tested if possible.

(2) Design of the Trial

As is necessary for any clinical trial, a *written protocol* was produced which documented all information concerning the purpose, design and conduct of the trial. Just a few of the salient design points will be mentioned here.

It was anticipated that the *number of patients* needed to obtain a clear answer to the main hypothesis would be of the order of several hundred. This required a *multi-centre* trial whereby, in fact, 37 American cancer hospitals agreed to enter patients into the trial. The study was coordinated by the National Surgical Adjuvant Breast Project (NSABP) and funded by the US National Cancer Institute.

The basic design was that each eligible patient was *randomly assigned* to receive either L-Pam or a *placebo* (an inert substance which looked and tasted the same as L-Pam). This randomization was by telephone to a central office in Pittsburgh. Patients were *stratified* by age (under or over 50), nodal status (1–3 or 4+ positive axillary nodes) and institution so that the randomization could be restricted to ensure the two treatment groups of patients would be comparable as regards these three factors. Each patient had a 50/50 chance of being assigned to L-Pam. The precise mechanics of such a stratified randomization will be explained in chapter 5.

The trial was *double-blind* so that neither the patient nor her attending physician nor others concerned with patient care or evaluation knew which treatment she was on, the oral drug or placebo being supplied in anonymous containers. Stratified randomization, the use of placebo and the double-blind restriction were all considered essential to ensure that the comparison of treatment and control groups could not be influenced by any extraneous factors such as the physician's personal judgement or the patient's morale. Such plans to eliminate bias are the key to any successful trial.

Each patient was to have a follow-up examination every six weeks and tests for haematologic toxicity every three weeks. Other blood tests, chest X-rays and bone scans were performed at less frequent but regular intervals. Thus, *end-point evaluation* was performed in the same consistent and objective manner for all patients.

(3) Conduct of the Trial

The first patient was entered into the study in September 1972. *Patient accrual* was terminated in February 1975, by which time 370 patients had been entered from the 37 participating institutions. In each case, *informed patient consent* to take part in the trial was obtained in accordance with standard United States procedure.

In a trial of this size and complexity there were inevitably some *protocol violations*. For instance, five patients were ineligible for the study and 17 patients did not start their treatment according to protocol. These patients were excluded from further study, so that there were 348 patients for analysis, 169 on placebo and 179 on L-Pam.

There were also a few subsequent *patient withdrawals* from the study: reasons included two patients refusing further treatment (placebo, in fact), three patients developing a second cancer unrelated to their primary breast tumour, one myocardial infarction and one renal failure death. It was decided that each of these withdrawals bore no relation to treatment and hence in analysis such patients were handled as if they were lost to follow-up at the time of withdrawal.

For such a large multi-centre trial it was important to have an effective *trial committee* (including a *study chairman*) which would meet periodically to assess progress and make alterations as necessary. For instance, it became evident after a few months that there was some resistance to the initial decision to restrict patient entry to those with four or more positive axillary nodes, so that an early protocol alteration was to allow patients with one or more positive nodes to enter the trial.

In addition, day-to-day running of the trial was handled by the NSABP Headquarters Office in Pittsburgh. Besides monitoring patient entry, such a *central coordinating office* is essential for supervising *data collection* and *processing* prior to statistical analysis. In this case, it was the responsibility of *data managers* to ensure that all *forms* with patient data were received promptly, checked for errors or missing data and computer processed.

(4) Data Analysis

For a trial that takes over two years to recruit sufficient patients and which requires subsequent follow-up of each patient for several years, information about the relative merits of the treatments is accumulated slowly. It is therefore common practice to undertake occasional *interim analyses* of the accumulating results while the trial is in progress. In this particular trial there was considerable pressure to reveal the findings about disease-free survival at an early stage, since it was widely recognized that this trial would provide a major breakthrough in the treatment of primary breast cancer if the results were positive. The study chairman and his trial committee resisted this pressure for premature publication and maintained strict secrecy over their results until there was strong statistical evidence of improved disease-free survival on L-Pam

especially in premenopausal women. Thus, such early findings were first revealed in 1975 but I will now concentrate on the more extensive results published by Fisher *et al.* (1977).

The easiest item to note first as regards disease-free survival is the number of patients on each treatment who had a recurrence of their disease and/or died. However, in such a *follow-up study* this comparison is over-simple since it fails to take into account the different lengths of time patients had been followed for: ranging from 20 months to 48 months in the 1977 analysis. Hence, a statistical technique known as life-table analysis of survival data was used to produce the results in figure 1.2, which shows for each treatment the estimated percentage of patients still alive and disease-free according to the time since mastectomy. This graph shows that 11% of patients on L-Pam had disease recurrence within a year of mastectomy compared with 24% of patients on placebo. After two years' follow-up the estimated percentage recurrence was 24% and 32% on L-Pam and placebo, respectively. Such *descriptive statistics*, clearly displayed in graphical or tabular form, are an important indication as to whether an interesting treatment difference may have arisen.

However, referring back to the main hypothesis before the trial began, one needs a formal *test of hypothesis* to assess whether the apparent improvement in disease-free survival on L-Pam can genuinely be attributed to the drug or could have arisen by chance. Conventionally this is done using a *statistical significance test*, the logic of which is as follows:

- (1) Suppose L-Pam and placebo are really equally effective as regards disease-free survival (this is called the null hypothesis).
- (2) Then, what is the probability P of getting such a big observed difference in disease-free survival as was found in figure 1.2, if the null hypothesis is true.

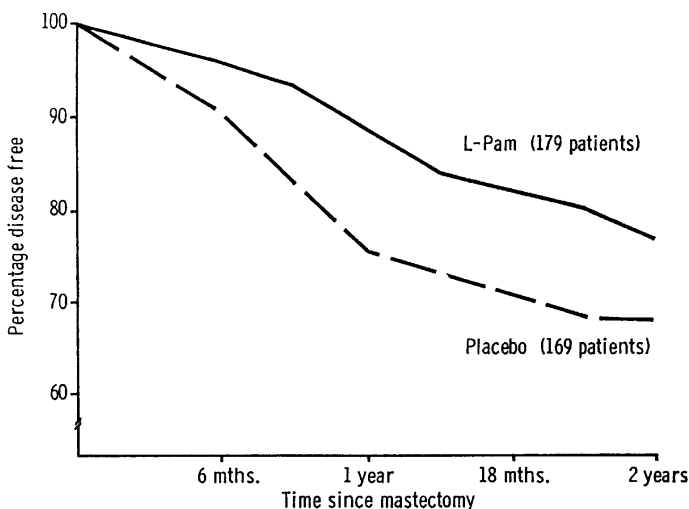


Fig. 1.2. Comparison of disease free survival on L-Pam and on placebo

- (3) The answer is $P = 0.009$, i.e. such a difference is to be expected by chance 9 times in 1000. This was determined by a statistical method called the modified Wilcoxon test, the details of which need not concern us. The standard phraseology is then to declare that the treatment difference in disease-free survival is statistically significant at the 1 % level (i.e. $P < 0.01$ for short).
- (4) This formal procedure enables one to say that there is strong evidence that L-Pam does prolong disease-free survival. However, it should be noted that in any clinical trial one can *never* obtain absolute proof of a treatment difference, but merely assess the extent to which the evidence is indicative of a treatment difference; such is the reality of the scientific method.

In addition to this global comparison of treatments relating to all patients in the trial, it is useful to examine whether the apparent benefit of L-Pam might depend on some *prognostic factors*, i.e. clinical or personal features of a patient as recorded in the initial patient status upon entry into the trial. In this trial it was anticipated that the patient's age, menopausal state and number of positive axillary nodes might influence the effect of L-Pam. As shown in figure 1.3 it turned out that the difference between L-Pam and placebo was more marked in patients under age 50 than in those over age 50. However, one needs to be careful in interpreting such apparent subgroup differences in treatment effect.

Patient survival has also been studied, there being 84 % and 90 % alive after two years on placebo and L-Pam, respectively. This difference is not statistically significant, but this does not indicate that L-Pam has no effect on patient survival. One really needs to follow such patients for up to five years in order to give a clear verdict on patient survival.

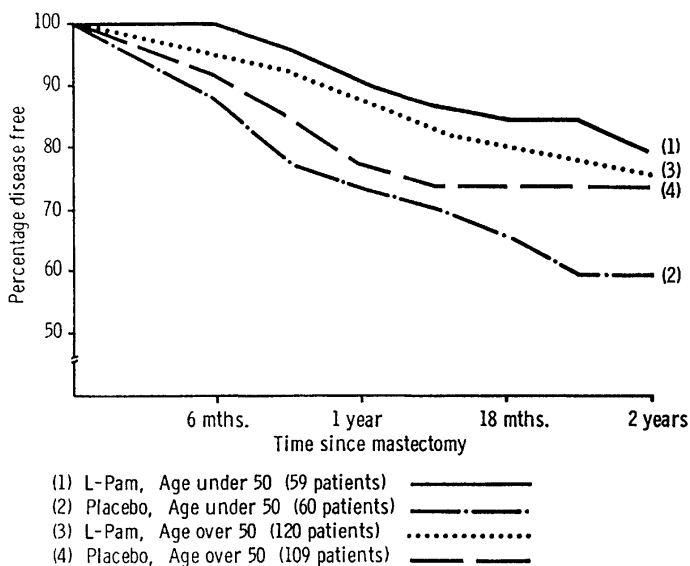


Fig. 1.3. Disease-free survival according to treatment (L-Pam or placebo) and age

Assessment of the *toxic side-effects* of L-Pam is important, since one wants to avoid undue drug toxicity in treating patients who have no observable disease after mastectomy. White cell count and/or platelet counts were lowered in the majority of patients on L-Pam, sufficient to require treatment to be stopped for a while in over a quarter of patients, but no life-threatening cases were reported. Also, 40% of L-Pam patients experienced some degree of nausea and vomiting (so did 11% of placebo patients, an indication that all untoward events cannot automatically be attributed to drug therapy). However, in view of the serious nature of the disease and other potential benefits of L-Pam, such toxicity was generally considered acceptable.

(5) Conclusions from the Trial

The overall assessment of L-Pam treatment focusses on the main hypothesis concerning disease-free interval, with appropriate account being taken of the subsidiary hypotheses concerning survival and toxicity. Thus it appears that L-Pam after mastectomy is a useful supplement to treatment of primary breast cancer with positive axillary nodes, but the benefit is more evident for younger premenopausal women than for older postmenopausal women. However, patient follow-up continues and subsequent survival comparisons will extend the conclusions. The trial organizers felt that the benefits were sufficient to prohibit the use of placebo in their next clinical trial started in 1975 which compares L-Pam with L-Pam + 5-FU. Another trial of three-drug chemotherapy has also now been started. It is interesting to note that the new trials have accrued patients at a much faster rate: that is, it is much easier to get physicians to enter patients on a clinical trial once earlier pioneering trials have shown the general approach to be beneficial. Fisher *et al.* (1981) review subsequent progress in these trials.

The main means of bringing the outcome of a trial to the attention of a general medical audience is to *publish the results* in a medical journal. The introduction, methods, results and conclusions sections of such a paper (the standard layout of scientific articles) correspond to the purpose, design and conduct, analysis and conclusions stages of a trial as outlined in figure 1.1. All of the paper prior to the conclusions will concentrate on objective statements of factual evidence, whereas the conclusions tend to be a more subjective opinion of the authors based on their experienced interpretation of the evidence. However, in any trial, and indeed this trial of L-Pam for primary breast cancer is no exception, the ultimate conclusion rests with other practising physicians whose subsequent experience of L-Pam and similar therapies either in future trials or as part of their regular practice will determine whether such therapy is generally applicable.

I hope the above description of one specific clinical trial has given a sense of reality to the main requirements of clinical trials in general. Of course citing one such example has its limitations since each particular trial has its own unique aspects. Nevertheless, many of the principles described in chapters 3–15 have been encapsulated in this example.

CHAPTER 2

The Historical Development of Clinical Trials

Attempts to evaluate the use of therapeutic procedures can be traced back to prehistoric times, and Bull (1959) provides an extensive account of the historical development of clinical trials up until 30 years ago. However, it is largely in these last 30 years that we have seen the development and general acceptance of properly conducted clinical trials which have conformed to the scientific principles outlined in this book. Furthermore, there has been an enormous continuing expansion in clinical trial activity throughout the 20th century which will probably carry on through the 1980's. A comprehensive historical review of clinical trials would require a book all to itself. Hence only a few of the major highlights in actual trials and conceptual developments will be mentioned here.

Section 2.1 gives a brief account of some interesting landmarks in clinical trials pre-1950, culminating in the pioneering postwar trials by the Medical Research Council. Section 2.2 brings us into the modern era of properly designed clinical trials, focussing on two early randomized trials in polio vaccine and diabetes. Sections 2.3–2.5 deal with three general areas of progress: cancer chemotherapy, post-infarction trials and the pharmaceutical industry.

2.1 CLINICAL TRIALS BEFORE 1950

There are some early landmarks in clinical investigation which anticipate the current methodology. For instance, Lind (1753) planned a comparative trial of *the most promising treatments for scurvy*. 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 them all. Two of these were ordered a quart of cider a day. Two others took twenty-five gutts of elixir vitriol ... Two others took two spoonfuls of vinegar ... Two were put under a course of sea water ... Two others had each two oranges and one lemon given them each day ... Two others took the bigness of a nutmeg. The most sudden and visible good effects were perceived from the use of 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.