Statistical Thinking for Non-Statisticians in Drug Regulation

Richard Kay

Consultant in Statistics for the Pharmaceutical Industry
Great Longstone
Derbyshire, UK

John Wiley & Sons, Ltd

Statistical Thinking for Non-Statisticians in Drug Regulation

Statistical Thinking for Non-Statisticians in Drug Regulation

Richard Kay

Consultant in Statistics for the Pharmaceutical Industry
Great Longstone
Derbyshire, UK

John Wiley & Sons, Ltd

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.wiley.com

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, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

Library of Congress Cataloging in Publication Data

Kay, R. (Richard), 1949–
Statistical thinking for non-statisticians in drug regulation / Richard Kay.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-470-31971-0 (cloth : alk. paper)
1. Clinical trials—Statistical methods. 2. Drugs—Testing—Statistical methods. 3. Drug approval—Statistical methods. 4. Pharmaceutical industry—Statistical methods. I. Title. [DNLM: 1. Clinical Trials—methods. 2. Statistics. 3. Drug Approval. 4. Drug Industry. QV 771 K23s 2007]
R853.C55K39 2007
615.5′8′0724—dc22

2007022438

British Library Cataloguing in Publication Data

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

ISBN 978-0-470-31971-0

Typeset in 10.5/13pt Minion by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd., Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

To Jan, Matt, Sally and Becci

Contents

Sections marked with an asterisk refer to some more challenging sections of the book.

Preface

This book is primarily concerned with clinical trials planned and conducted within the pharmaceutical industry. Much of the methodology presented is in fact applicable on a broader basis and can be used in observational studies and in clinical trials outside of the pharmaceutical sector; nonetheless the primary context is clinical trials and pharmaceuticals. The development is aimed at non-statisticians and will be suitable for physicians, investigators, clinical research scientists, medical writers, regulatory personnel, statistical programmers, senior data managers and those working in quality assurance. Statisticians moving from other areas of application outside of pharmaceuticals may also find the book useful in that it places the methods that they are familiar with, in context in their new environment. There is substantial coverage of regulatory aspects of drug registration that impact on statistical issues. Those of us working within the pharmaceutical industry recognise the importance of being familiar with the rules and regulations that govern our activities and statistics is a key aspect of this.

The aim of the book is not to turn non-statisticians into statisticians. I do not want you to go away from this book and 'do' statistics. It is the job of the statistician to provide statistical input to the development plan, to individual protocols, to write the statistical analysis plan, to analyse the data and to work with medical writing in producing the clinical report; also to support the company in its interactions with regulators on statistical issues.

The aims of the book are really three-fold. Firstly, to aid communication between statisticians and non-statisticians, secondly, to help in the critical review of reports and publications and finally, to enable the more effective use of statistical arguments within the regulatory process. We will take each of these points in turn.

In many situations the interaction between a statistician and a non-statistician is not a particularly successful one. The statistician uses terms, for example, power, odds ratio, p-value, full analysis set, hazard ratio, non-inferiority, type II error, geometric mean, last observation carried forward and so on, of which the non-statistician has a vague understanding, but maybe not a good enough understanding to be able to get an awful lot out of such interactions. Of course it is always the job of a statistician to educate and every opportunity should be taken for imparting knowledge about statistics, but in a specific context there may not be time for that. Hopefully this book will explain, in ways that are understandable,

just what these terms mean and provide some insight into their interpretation and the context in which they are used. There is also a lot of confusion between what on the surface appear to be the same or similar things; significance level and p-value, equivalence and non-inferiority, odds ratio and relative risk, relative risk and hazard ratio (by the way this is a minefield!) and meta-analysis and pooling to name just a few. This book will clarify these important distinctions.

It is unfortunately the case that many publications, including some leading journals, contain mistakes with regard to statistics. Things have improved over the years with the standardisation of the ways in which publications are put together and reviewed. For example the CONSORT statement (see Section 16.5) has led to a distinct improvement in the quality of reporting. Nonetheless mistakes do slip through, in terms of poor design, incorrect analysis, incomplete reporting and inappropriate interpretation – hopefully not all at once! It is important therefore when reading an article that the non-statistical reader is able to make a judgement regarding the quality of the statistics and to notice any obvious flaws that may undermine the conclusions that have been drawn. Ideally the non-statistician should involve their statistical colleagues in evaluating their concerns but keeping a keen eye on statistical arguments within the publication may help to alert the non-statistician to a potential problem. The same applies to presentations at conferences, posters, advertising materials and so on.

Finally the basis of many concerns raised by regulators, when they are reviewing a proposed development plan or assessing an application for regulatory approval, is statistical. It is important that non-statisticians are able to work with their statistical colleagues in correcting mistakes, changing aspects of the design, responding to questions about the data to hopefully overcome those concerns.

In writing this book I have made the assumption that the reader is familiar with general aspects of the drug development process. I have assumed knowledge of the phase I to phase IV framework, of placebos, control groups, and double-dummy together with other fundamental elements of the nuts and bolts of clinical trials. I have assumed however no knowledge of statistics! This may or may not be the correct assumption in individual cases, but it is the common denominator that we must start from, and also it is actually not a bad thing to refresh on the basics. The book starts with some basic issues in trial design in Chapter 1 and I guess most people picking up this book will be familiar with many of the topics covered there. But don't be tempted to skip this chapter; there are still certain issues, raised in this first chapter, that will be new and important for understanding arguments put forward in subsequent chapters. Chapter 2 looks at sampling and inferential statistics. In this chapter we look at the interplay between the population and the sample, basic thoughts on measuring average and variability and then explore the process of sampling leading to the concept of the standard error as a way of capturing precision/reliability of the sampling process. The construction and interpretation of confidence intervals is covered in

Chapter 3 together with testing hypotheses and the (dreaded!) p-value. Common statistical tests for various data types are developed in Chapter 4 which also covers different ways of measuring treatment effect for binary data, such as the odds ratio and relative risk.

Many clinical trials that we conduct are multi-centre and Chapter 5 looks at how we extend our simple statistical comparisons to this more complex structure. These ideas lead naturally to the topics in Chapter 6 which include the concepts of adjusted analyses, and more generally, analysis of covariance which allows adjustment for many baseline factors, not just centre. Chapters 2 to 6 follow a logical development sequence in which the basic building blocks are initially put in place and then used to deal with more and more complex data structures. Chapter 7 moves a little away from this development path and covers the important topic of Intention-to-Treat and aspects of conforming with that principle through the definition of different analysis sets and dealing with missing data. In Chapter 8, we cover the very important design topics of power and the sample size calculation which then leads naturally to a discussion about the distinction between statistical significance and clinical importance in Chapter 9.

The regulatory authorities, in my experience, tend to dig their heels in on certain issues and one such issue is multiplicity. This topic, which has many facets, is discussed in detail in Chapter 10. Non-parametric and related methods are covered in Chapter 11. In Chapter 12 we develop the concepts behind the establishment of equivalence and non-inferiority. This is an area where many mistakes are made in applications, and in many cases these slip through into published articles. It is a source of great concern to many statisticians that there is widespread misunderstanding of how to deal with equivalence and non-inferiority. I hope that this chapter helps to develop a better understanding of the methods and the issues. If you have survived so far, then Chapter 13 covers the analysis of survival data. When an endpoint is time to some event, for example death, the data are inevitably subject to what we call censoring and it is this aspect of so-called survival data that has led to the development of a completely separate set of statistical methods. Chapter 11 builds on the earlier discussion on multiplicity to cover one particular manifestation of that, the interim analysis. This chapter also looks at the management of these interim looks at the data through data monitoring committees. Meta-analysis and its role in clinical development is covered in Chapter 15 and the book finishes with a general Chapter 16 on the role of statistics and statisticians in terms of the various aspects of design and analysis and statistical thinking more generally.

It should be clear from the last few paragraphs that the book is organised in a logical way; it is a book for learning rather than a reference book for dipping into. The development in later chapters will build on the development in earlier chapters. I strongly recommend, therefore, that you start on page 1 and work through. I have tried to keep the discussion away from formal mathematics. There are

formulas in the book but I have only included these where I think this will enhance understanding; there are no formulas for formulas sake! There are some sections that are more challenging than others and I have marked with an asterisk those sections that can be safely sidestepped on a first (or even a second) run through the book.

The world of statistics is ever changing. New methods are being developed by theoreticians within university departments and ultimately some of these will find their way into mainstream methods for design and statistical analysis within our industry. The regulatory environment is ever changing as regulators respond to increasing demands for new and more effective medicines. This book in one sense represents a snapshot in time in terms of what statistical methods are employed within the pharmaceutical industry and also in relation to current regulatory requirements. Two statistical topics that are not included in this book are Bayesian Methods and Adaptive (Flexible) Designs (although some brief mention is made of this latter topic in section 14.5.2). Both areas are receiving considerable attention at the moment and I am sure that within a fairly short period of time there will be much to say about them in terms of the methodological thinking, examples of their application and possibly with regard to their regulatory acceptance but for the moment they are excluded from our discussions.

The book has largely come out of courses that I have been running under the general heading of Statistical Thinking for Non-Statisticians for a number of years. There have been several people who have contributed from time to time and I would like to thank them for their input and support; Werner Wierich, Mike Bradburn and in particular Ann Gibb who gave these courses with me over a period of several years and enhanced my understanding through lively discussion and asking many challenging questions. I would also like to thank Simon Gillis who contributed to Chapter 16 with his much deeper knowledge of the processes that go on within a pharmaceutical company in relation to the analysis and reporting of a clinical trial.

Richard Kay
Great Longstone
January 2007

Abbreviations

AIDAC	Anti-Infective Drugs Advisory Committee
ANCOVA	analysis of covariance
ANOVA	analysis of variance
AE	adverse event
ARR	absolute relative risk
AUC	area under the time concentration curve
BMD	bone mineral density
CDER	Center for Drug Evaluation and Research
CFC	Chlorofluorocarbon
CHMP	Committee for Medical Products for Human Use
CI	confidence interval
CPMP	Committee for Proprietary Medicinal Products
C_{MAX}	maximum concentration
CMH	Cochran-Mantel-Haenszel
CNS	central nervous system
CRF	Case Report Form
CR	complete response
crd	clinically relevant difference
DMC	Data Monitoring Committee
DSMB	Data and Safety Monitoring Board
DSMC	Data and Safety Monitoring Committee
ECG	Electrocardiogram
ECOG	Eastern Cooperative Oncology Group
EMEA	European Medicines Evaluation Agency
FDA	Food and Drug Administration
FEV_1	forced expiratory volume in one second
GP	General Practitioner
HAMA	Hamilton Anxiety Scale
HAMD	Hamilton Depression Scale

HER2	human epidermal growth factor receptor-2
HIV	human immunodeficiency virus
HR	Hazard Ratio
ICH	International Committee on Harmonisation
ITT	Intention-to-Treat
IVRS	Interactive Voice Response System
KM	Kaplan-Meier
LDH	lactate hehydrogenase
LOCF	last observation carried forward
MedDRA	*Med*ical *D*ictionary for *R*egulatory *A*ctivities
MH	Mantel-Haenszel
MI	myocardial infarction
NNH	number needed to harm
NNT	number needed to treat
NS	not statistically significant
OR	odds ratio
PD	progressive disease
PEF	peak expiratory flow
PHN	post-hepatic neuralgia
PR	partial response
RECIST	Response Evaluation Criteria in Solid Tumours
RR	relative risk
RRR	relative risk reduction
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	stable disease
sd	standard deviation
se	standard error
VAS	visual analogue scale
WHO	World Health Organisation

1

Basic ideas in clinical trial design

1.1 Historical perspective

As many of us who are involved in clinical trials will know, the randomised, controlled trial is a relatively new invention. As pointed out by Pocock (1983) and others, very few clinical trials of the kind we now regularly see were conducted prior to 1950. It took a number of high profile successes plus the failure of alternative methodologies to convince researchers of their value.

Example 1.1: The Salk Polio Vaccine trial

One of the largest trials ever conducted took place in the US in 1954 and concerned the evaluation of the Salk Polio Vaccine. The trial has been reported extensively by Meier (1978) and is used by Pocock (1983) in his discussion of the historical development of clinical trials.

Within the project there were essentially two trials and these clearly illustrated the effectiveness of the randomised controlled design.

Trial 1: Original design; observed control
1.08 million children from selected schools were included in this first trial. The second graders in those schools were offered the vaccine while the first and third graders would serve as the control group. Parents of the second graders were approached for their consent and it was noted that the consenting parents tended to have higher incomes. Also, this design was not blinded so that both parents and investigators knew which children had received the vaccine and which had not.

Statistical Thinking for Non-Statisticians in Drug Regulation Richard Kay
 ISBN 978-0-470-31971-0

Example 1.1: (Continued)

Trial 2: Alternative design; randomised control
A further 0.75 million children in other selected schools in grades one to three were to be included in this second trial. All parents were approached for their consent and those children where consent was given were randomised to receive either the vaccine or a placebo injection. The trial was double-blind with parents, children and investigators unaware of who had received the vaccine and who had not.

The results from the randomised control trial were conclusive. The incidence of paralytic polio for example was 0.057 per cent in the placebo group compared to 0.016 per cent in the active group and there were four deaths in the placebo group compared to none in the active group. The results from the observed control trial, however, were less convincing with a smaller observed difference (0.046 per cent versus 0.017 per cent). In addition, in the cases where consent could not be obtained, the incidence of paralytic polio was 0.036 per cent in the randomised trial and 0.037 per cent in the observed control trial, event rates considerably lower than those amongst placebo patients and in the untreated controls respectively. This has no impact on the conclusions from the randomised trial, which is robust against this absence of consent; the randomised part is still comparing like with like. In the observed control part however the fact that the 'no consent' (grade 2) children have a lower incidence that those children (grades 1 and 3) who were never offered the vaccine potentially causes some confusion in a non-randomised comparison; does it mean that grade 2 children naturally have lower incidence than those in grades 1 and 3? Whatever the explanation, the presence of this uncertainty reduced confidence in other aspects of the observed control trial.

The randomised part of the Salk Polio Vaccine trial has all the hallmarks of modern day trials; randomisation, control group, blinding and it was experiences of these kinds that helped convince researchers that only under these conditions can clear, scientifically valid conclusions be drawn.

1.2 Control groups

We invariably evaluate our treatments by making comparisons; active compared to control. It is very difficult to make absolute statements about specific treatments and conclusions regarding the efficacy and safety of a new treatment are made relative to an existing treatment or placebo.

ICH E10 (2001): 'Note for Guidance on Choice of Control Group in Clinical Trials'

'Control groups have one major purpose: to allow discrimination of patient outcomes (for example, changes in symptoms, signs, or other morbidity) caused by the test treatment from outcomes caused by other factors, such as the natural progression of the disease, observer or patient expectations, or other treatment.'

Control groups can take a variety of forms, here are just a few examples of trials with alternative types of control group:

- Active versus placebo
- Active A versus active B (versus active C)
- Placebo versus dose level 1 versus dose level 2 versus dose level 3 (dose-finding)
- Active A + active B versus active A + placebo (add-on)

The choice will depend on the objectives of the trial.

Open trials with no control group can nonetheless be useful in an exploratory, maybe early phase setting, but it is unlikely that such trials will be able to provide confirmatory, robust evidence regarding the performance of the new treatment.

Similarly, external or historical controls (groups of subjects external to the study either in a different setting or previously treated) cannot provide definitive evidence. Byar (1980) provides an extensive discussion on these issues.

1.3 Placebos and blinding

It is important to have blinding of both the subject and the investigator wherever possible to avoid unconscious bias creeping in, either in terms of the way a subject reacts psychologically to a treatment or in relation to the way the investigator influences or records subject outcome.

ICH E9 (1998): 'Note for Guidance on Statistical Principles for Clinical Trials'

'Blinding or masking is intended to limit the occurrence of conscious or unconscious bias in the conduct and interpretation of a clinical trial arising from the influence which the knowledge of treatment may have on the recruitment and allocation of subjects, their subsequent care, the attitudes of subjects to the treatments, the assessment of the end-points, the handling of withdrawals, the exclusion of data from analysis, and so on.'

Ideally the trial should be *double-blind* with both the subject and the investigator being blind to the specific treatment allocation. If this is not possible for the investigator, for example, then the next best thing is to have an independent evaluation of outcome, both for efficacy and for safety. A *single-blind* trial arises when either the subject or investigator, but not both, is blind.

An absence of blinding can seriously undermine the validity of an endpoint in the eyes of regulators and the scientific community more generally, especially when the evaluation of that endpoint has an element of subjectivity. In situations where blinding is not possible it is essential to use hard, unambiguous endpoints.

The use of placebos and blinding go hand in hand. The existence of placebos enable trials to be blinded and account for the placebo effect; the change in a patient's condition that is due to the act of being treated, but is not caused by the active component of that treatment.

1.4 Randomisation

Randomisation is clearly a key element in the design of our clinical trials. There are two reasons why we randomise subjects to the treatment groups:

- To avoid any bias in the allocation of the patients to the treatment groups
- To ensure the validity of the statistical test comparisons

Randomisation lists are produced in a variety of ways and we will discuss several methods later. Once the list is produced the next patient entering the trial receives the next allocation within the randomisation scheme. In practice this process is managed by 'packaging' the treatments according to the pre-defined randomisation list.

There are a number of different possibilities when producing randomisation lists:

- Unrestricted randomisation
- Block randomisation
- Unequal randomisation
- Stratified randomisation
- Central randomisation
- Dynamic allocation and minimisation
- Cluster randomisation

1.4.1 Unrestricted randomisation

Unrestricted (or simple) randomisation is simply a random list of, for example, As and Bs. In a moderately large trial, with say $n = 200$ subjects, such a process will likely produce approximately equal group sizes. There is no guarantee however that this will automatically happen and in small trials, in particular, this can cause problems.

1.4.2 Block randomisation

To ensure balance in terms of numbers of subjects, we usually undertake *block randomisation* where a randomisation list is constructed by randomly choosing from the list of potential blocks. For example, there are six ways of allocating two As and two Bs in a 'block' of size four:

AABB, ABAB, ABBA, BAAB, BABA, BBAA

and we choose at random from this set of six blocks to produce our randomisation list, for example:

ABBA BAAB ABAB ABBA, …

Clearly if we recruit a multiple of four patients into the trial we will have perfect balance, and approximate balance (which is usually good enough) for any sample size.

In large trials it could be argued that block randomisation is unnecessary. In one sense this is true, overall balance will be achieved by chance with an unrestricted randomisation list. However, it is usually the case that large trials will be multicentre trials and not only is it important to have balance overall it is also important to have balance within each centre. In practice therefore we would allocate several blocks to each centre, for example five blocks of size four if we are planning to recruit 20 patients from each centre. This will ensure balance within each centre and also overall.

How do we choose block size? There is no magic formula but more often than not the block size is equal to two times the number of treatments.

What are the issues with block size?

ICH E9 (1998): 'Note for Guidance on Statistical Principles for Clinical Trials'

'Care must be taken to choose block lengths which are sufficiently short to limit possible imbalance, but which are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length . . . '

Shorter block lengths are better at producing balance. With two treatments a block length of four is better at producing balance than a block length of 12. The block length of four gives perfect balance if there is a multiple of four patients entering, whereas with a block length of 12, perfect balance is only going to be achieved if there are a multiple of 12 patients in the study. The problem, however, with the shorter block lengths is that this is an easy code to crack and inadvertent unblinding can occur. For example suppose a block length of four was being used in a placebo controlled trial and also assume that experience of the active drug suggests that many patients receiving that drug will suffer nausea. Suppose the trial begins and the first two patients suffer nausea. The investigator is likely to conclude that both these patients have been randomised to active and that therefore the next two allocations are to placebo. This knowledge could influence his willingness to enter certain patients into the next two positions in the randomisation list, causing bias in the mix of patients randomised into the two treatment groups. Note the comment in the ICH guideline regarding keeping the investigator (and others) blind to the block length. While in principle this comment is sound, the drug is often delivered to a site according to the chosen block length, making it difficult to conceal information on block size. If the issue of inadvertent unblinding is going to cause problems then more sophisticated methodologies can be used, such as having the block length itself varying; perhaps randomly chosen from two, four or six.

1.4.3 Unequal randomisation

All other things being equal, having equal numbers of subjects in the two treatment groups provides the maximum amount of information (the greatest power) with regard to the relative efficacy of the treatments. There may, however, be issues that override statistical efficiency:

- It may be necessary to place more patients on active compared to placebo in order to obtain the required safety information.
- In a three group trial with active A, active B and placebo(P), it may make sense to have a 2:2:1 randomisation to give more power for the A versus B comparison as that difference is likely to be smaller then the A versus P and B versus P differences.

Unequal randomisation is sometimes needed as a result of these considerations. To achieve this, the randomisation list will be designed for the second example above with double the number of A and B allocations compared to placebo.

For unequal randomisation we would choose the block size accordingly. For a 2:1 randomisation to A or P we could randomly choose from the blocks:

AAP, APA, PAA

1.4.4 Stratified randomisation

Block randomisation therefore forces the required balance in terms of the numbers of patients in the treatment groups, but things can still go wrong. For example, let's suppose in an oncology study with time to death as the primary endpoint that we can measure baseline risk (say in terms of the size of the primary tumour) and classify patients as either high risk (H) or low risk (L) and further suppose that the groups turn out as follows:

A: HHLHLHHHHLLHHHLHHLHHH (H=15, L=6)
B: LLHHLHHLLHLHLHLHHLLHLL (H=10, L=12)

Note that there are 15 patients (71 per cent) high risk and six (29 per cent) low risk patients in treatment group A compared to a split of 10 (45 per cent) high risk and 12 (55 per cent) low risk patients in treatment group B.

Now suppose that the mean survival times are observed to be 21.5 months in A and 27.8 months in group B. What conclusions can we draw? It is very difficult; the difference we have seen could be due to treatment differences or could be caused by the imbalance in terms of differential risk across the groups, or a mixture of the two. Statisticians talk in terms of *confounding* (just a fancy way of saying 'mixed up') between the treatment effect and the effect of baseline risk. This situation is very difficult to unravel and we avoid it by *stratified randomisation* to ensure that the 'case mix' in the treatment groups is comparable.

This simply means that we produce separate randomisation lists for the high risk and the low risk patients, the strata in this case. For example the following lists (which are block size four in each case):

H: ABBAAABBABABABABBBAAABBAABABBBAA
L: BAABBABAAABBBAABABABBBAABBAABAAB

will ensure firstly that we end up with balance in terms of group sizes but also secondly that both the high and low risk patients will be equally split across those groups, that is balance in terms of the mix of patients.

Having separate randomisation lists for the different centres in a multi-centre trial to ensure 'equal' numbers of patients in the treatment groups within each centre is using 'centre' as a stratification factor; this will ensure that we do not end up with treatment being confounded with centre.

ICH E9 (1998): 'Note for Guidance on Statistical Principles for Clinical Trials'

'It is advisable to have a separate random scheme for each centre, i.e. to stratify by centre or to allocate several whole blocks to each centre. Stratification by important prognostic factors measured at baseline (e.g. severity of disease, age, sex, etc.) may sometimes be valuable in order to promote balanced allocation within strata . . . '

Where the requirement is to have balance in terms of several factors, a stratified randomisation scheme using all combinations of these factors to define the strata would ensure balance. For example if balance is required for sex and age, then a scheme with four strata:

- Males, < 50 years
- Females, < 50 years
- Males, ≥ 50 years
- Females, ≥ 50 years

will achieve the required balance.

1.4.5 Central randomisation

In *central randomisation* the randomisation process is controlled and managed from a centralised point of contact. Each investigator makes a telephone call through an Interactive Voice Response System (IVRS) to this centralised point when they have identified a patient to be entered into the study and is given the next allocation, taken from the appropriate randomisation list. Blind can be preserved by simply specifying the number of the (pre-numbered) pack to be used to treat the particular patient; the computerised system keeps a record of which packs have been used already and which packs contain which treatment. Central randomisation has a number of practical advantages:

- It can provide a check that the patient about to be entered satisfies certain inclusion/exclusion criteria thus reducing the number of protocol violations.
- It provides up-to-date information on all aspects of recruitment.
- It allows more efficient distribution and stock control of medication.
- It provides some protection against biased allocation of patients to treatment groups in trials where the investigator is not blind; the investigator knowing the next allocation could (perhaps subconsciously) select

patients to include or not include based on that knowledge; with central randomisation the patient is identified and information given to the system before the next allocation is revealed to them.

- It gives an effective way of managing multi-centre trials.
- It allows the implementation of more complex allocation schemes such as minimisation and dynamic allocation.

Earlier we discussed the use of stratified randomisation in multi-centre trials and where the centres are large this is appropriate. With small centres however, for example in GP trials, this does not make sense and a stratified randomisation with 'region' defining the strata may be more appropriate. Central randomisation would be essential to manage such a scheme.

Stratified randomisation with more than a small number of strata would be difficult to manage at the site level and the use of central randomisation is then almost mandatory.

1.4.6 Dynamic allocation and minimisation

ICH E9 (1998): 'Note for Guidance on Statistical Principles for Clinical Trials'

'Dynamic allocation is an alternative procedure in which the allocation of treatment to a subject is influenced by the current balance of allocated treatments and, in a stratified trial, by the stratum to which the subject belongs and the balance within that stratum. Deterministic dynamic allocation procedures should be avoided and an appropriate element of randomisation should be incorporated for each treatment allocation.'

Dynamic allocation moves away from having a pre-specified randomisation list and the allocation of patients evolves as the trial proceeds. The method looks at the current balance, in terms of the mix of patients and a number of pre-specified factors, and allocates the next patient in an optimum way to help redress any imbalances that exist at that time.

For example, suppose we require balance in terms of sex and age (≥ 65 versus < 65) and part way through the trial we see a mix of patients as in Table 1.1.

Table 1.1 Current mix of patients

	A	B
Total	25	25
Male	12/25	10/25
Age ≥ 65	7/25	8/25