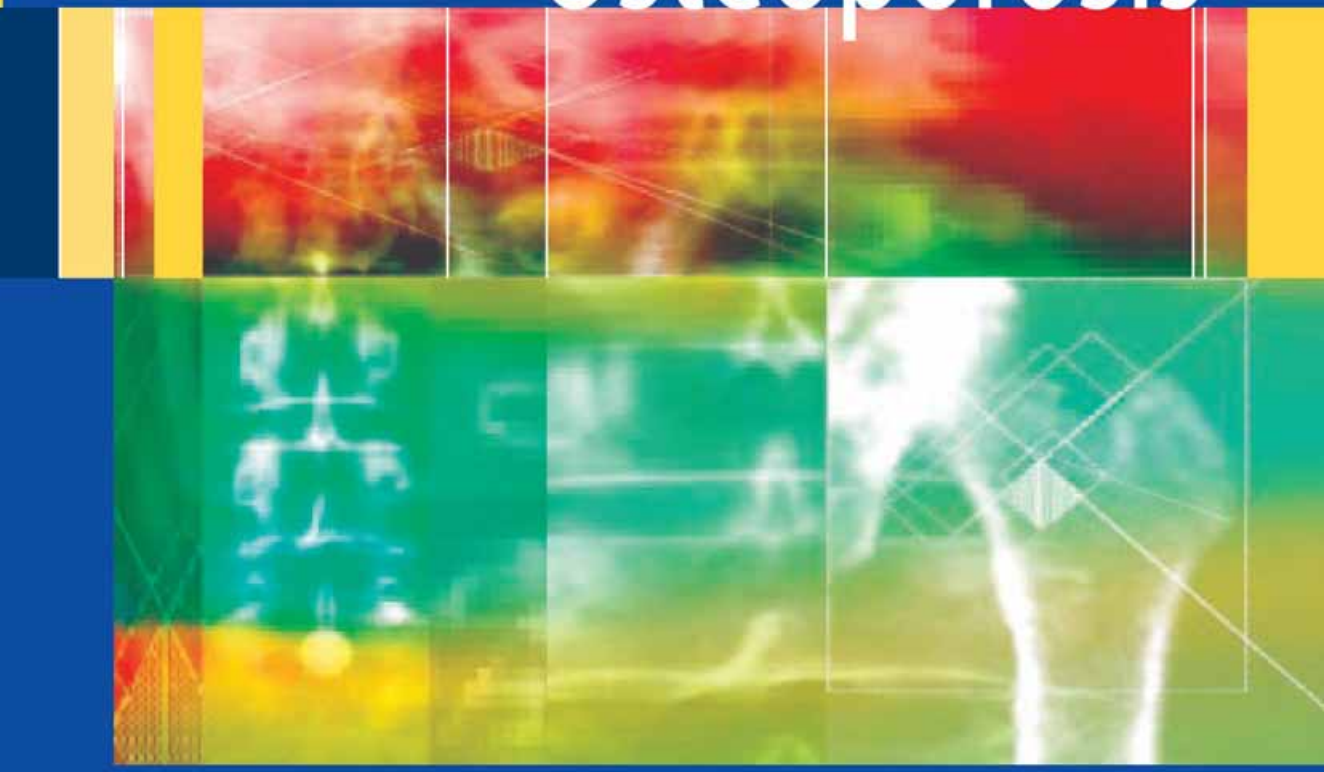
Derek Pearson
Colin G. Miller *Editors*

Clinical Trials in Osteoporosis



Second Edition

Springer

Clinical Trials in Osteoporosis

Second Edition

Clinical Trials in Osteoporosis

Second Edition

Edited by

Derek Pearson

and

Colin G. Miller

Springer

Derek Pearson, BSc, PhD
Joint Service Manager
Medical Physics & Clinical
Engineering Department
Nottingham University Hospitals
NHS Trust
Nottingham
UK

Colin G. Miller, PhD, FICR
Sr. Vice President
Medical Affairs
Bio-Imaging Technologies Inc.
Newtown, PA
USA

British Library Cataloguing in Publication Data
Clinical trials in osteoporosis - 2nd ed.
1. Osteoporosis - Chemotherapy - Research - Handbooks, manuals, etc. 2. Drugs - Testing - Handbooks, manuals, etc.
I. Pearson, Derek, 1954- II. Miller, Colin G., 1960-
616.7′16

Library of Congress Control Number: 2006931121

ISBN-10: 1-84628-389-2 2nd edition
ISBN-13: 978-1-84628-389-5 2nd edition
ISBN-10: 1-85233-229-8 1st edition

e-ISBN-10: 1-84628-587-9 2nd edition
e-ISBN-13: 978-1-84628-587-5 2nd edition

Printed on acid-free paper

First published 2002
Second edition 2007

9 8 7 6 5 4 3 2 1

Springer Science+Business Media
springer.com

Foreword to the First Edition

Before the mid-1990s, very few people were aware of the disease of osteoporosis, not only in the general population, but also in the medical profession. In the past 10 years there have been many advances and developments in the understanding of this crippling disease, to the point that this therapeutic field has lost its "fledgling" status and has grown up rapidly. This is well demonstrated by the number of effective therapies that are available, the new methodologies for the diagnosis and monitoring of the disease, and the fact that the importance of osteoporosis, as a disease, is now recognised and is taught as an integral part of the syllabus in most medical schools.

With the maturation of this therapeutic area, it has now become a challenge for physicians, scientists, and technologists to gain an easy understanding of the intricacies of running studies or clinical trials in osteoporosis. This book provides an excellent introduction and handbook for those wishing to pursue research in this field. It not only provides an overview of the field of osteoporosis, the measurement methodologies available, and current therapies, but also covers all the necessary regulations and good clinical practice requirements that are both specific to the disease state and generic across all clinical trials. Furthermore, an interesting slant has been taken in providing "mock trial data". The reader is then taken through the analysis of the study and, rather than having to do this theoretically, can work with the data provided. It will then be very straightforward for the reader to apply the calculations derived to their own data.

This book has been written in an easy-to-read style, and novices to the field of osteoporosis and/or clinical trials will be guided through the whole process. It is unusual on two major accounts. First, it has been written to appeal to those working in the pharmaceutical industry, in addition to those at the trial site: principal investigators, study site coordinators, and bone densitometry technologists. This has been achieved by the inclusion of several guest chapters, so the whole clinical trial arena has been covered. Second, this book has been written with both European, American and Canadian markets in mind. The principal authors, although both British, live on opposite sides of the Atlantic, and so the book has a very comprehensive feel to it. This unusual meld of editors and authors, from both industry and academia, has provided a unique opportunity for the development of this book, encompassing the many facets of clinical trials in osteoporosis. The addition of several notable guest authors has increased the depth of this book.

As a physician who has been involved in the field of osteoporosis for many years, I believe that this book enters the marketplace in a timely manner. Because the field has matured, this reference work is needed to help the researcher obtain the principles of the disease and clinical trial environment in a rapid and convenient

manner. It answers most of the basic questions, and many of the more complex ones. Because it covers the clinical trial programme from start-up to data analysis and publication, it will become a very useful and widely used handbook. Although some will read it from beginning to end, it also lends itself to being dipped into at the appropriate points in a trial lifecycle, without having to be onerous on time, which is rarely available in great quantities to anyone involved in clinical trials.

Ignac Fogelman
Guys Hospital,
London, UK

Contents

Section 4 Background

Contributors

Dr Dorothy Adams
Jackson Foundation, Madison, WI, USA

Dr Jonathan R. Bayly
Faculty of Education, Health and Sciences, University of Derby, Derby, UK

Dr Penny Blackwell
Clinical Chemistry, Nottingham University Hospitals NHS Trust, Nottingham, UK

Mrs Susan A. Earnshaw
South Canterbury, New Zealand

Professor Ignac Fogelman
Department of Nuclear Medicine, Guys Hospital, London, UK

Dr Ian M. Godber
Biochemistry Department, Wishaw General Hospital, Wishaw, UK

Professor David J. Hosking
Nottingham University Hospitals NHS Trust, Nottingham, UK

Dr Nigel Lawson
Clinical Chemistry, Derby Hospital NHS Foundation Trust, Derby City General Hospital, Derby, UK

Dr Colin G. Miller
Senior Vice President, Medical Affairs, Bio-Imaging Technologies Inc., Newtown, PA, USA

Dr Derek Pearson
Medical Physics and Clinical Engineering, Nottingham University Hospitals NHS Trust, Nottingham, UK

Contributor Biographies

Dorothy Adams is an experienced Research Administrator, having been involved in the running of numerous osteoporosis trials. She is an Executive Director of the Jackson Foundation, a clinical research organization based in Madison, WI, USA.

Jonathan R. Bayly is an associate lecturer at the University of Derby, Derby, UK, in the faculty of Education, Health and Sciences, where he works on the Radiography, Informatics, and Osteoporosis programme (http://www.derby.ac.uk/osteoporosis/index.htm). He is also the lead clinician for the Gloucestershire Primary and Community Care Audit Group, Gloucestershire, UK (http://www.glospccag.nhs.uk), supporting clinical governance in the three Primary Care Trusts within the county. He was a general practitioner for 20 years in Stroud, UK. He is a member of the Royal College of General Practitioners, the American Society for Bone and Mineral Research, and the British Geriatrics Society and is on the scientific advisory group of the UK National Osteoporosis Society.

Penny Blackwell graduated from Birmingham University, Birmingham, UK, in biochemistry and trained in clinical biochemistry at Nottingham City Hospital, Nottingham, UK. She obtained her PhD in bone biochemistry in 2000, and is now a third-year medical student. She plays the French horn particularly well.

Susan A. Earnshaw was employed as a Superintendent Radiographer responsible for the bone densitometry service at Nottingham City Hospital, Nottingham, UK. She worked in this field for more than nine years, being involved with clinical subjects and more than 40 bone densitometry research trials. During this time, the development of the capability of dual-energy X-ray absorptiometry (DXA) machines to perform morphometric X-ray analysis (MXA) led to her pursuing and gaining an MPhil in spinal morphometry. She had strong links with the University of Derby (Derby, UK) distance-learning programme, writing about the practical aspects of DXA for the osteoporosis modules, and was also a module tutor. She is now living with her family in New Zealand, raising sheep!

Ian M. Godber is a graduate of the Universities of St Andrews and Dundee, Scotland, UK. Ian trained as a Clinical Biochemist at Ninewells Hospital in Dundee, during which time he completed an MSc at the University of Surrey, Guildford, UK. He was a Senior Clinical Biochemist at Nottingham City Hospital, Nottingham, UK, in January 2000, where his main responsibilities included coordinating the biochemical analyses involved in a number of research projects investigating the

biochemical markers of bone turnover and calcium homeostasis. He is currently Principal of Biochemist at Wishaw General Hospital in Wishaw, Scotland, and also edits the Association for Clinical Biochemistry website and the public information website for laboratory tests, *Lab Tests Online UK*.

David J. Hosking is a Consultant Physician at Nottingham City Hospital, Nottingham, UK. After postgraduate training in Birmingham, UK, he went to the University of Leiden, Leiden, the Netherlands as an Medical Research Council Travelling Fellow working in the Department of Endocrinology. His research interests are in the fields of osteoporosis, Paget's disease, and vitamin D metabolism. He has published more than 100 papers in this field and is co-author of a textbook entitled *Management of Metabolic Bone Disease*.

Nigel Lawson graduated from the Biochemistry Department at the University of Sheffield, Sheffield, UK, in 1976, and obtained his PhD on the control of lipid synthesis from the University of Nottingham, Nottingham, UK, in 1980. After undertaking some more research in lipids, he worked in the Clinical Chemistry Departments of the East Birmingham Hospital and the Birmingham Children's Hospital, plus the Regional Immunology Department, in Birmingham, UK. Nigel moved to the Clinical Chemistry Department at Nottingham City Hospital in 1991 to take up the post of Consultant Clinical Scientist and was made a Fellow of the Royal College of Pathologists in 1997. He is now working at the Derby Hospitals NHS Foundation Trust, Derby, UK.

Colin G. Miller graduated in 1983 from the University of Sheffield, Sheffield, UK, with a degree in physiology and zoology. He worked for four years as a Research Assistant at the Bio-Engineering Research Unit, Doncaster University, Doncaster, UK, and then as a Clinical Research Associate at Syntex Research, Department of Clinical Pharmacology and Therapeutics Investigation, Maidenhead, UK. During this time, he obtained a PhD from Hull University, Hull, UK, on the measurement of broadband ultrasonic attenuation for assessing hip fracture in the elderly. Since then, he has worked as a Clinical Research Coordinator at Norwich Eaton Limited (a Procter & Gamble Company) and as Head of the Physical Measurements Team, Europe, Procter & Gamble Pharmaceuticals, Egham, Middlesex, UK. Between 1991 and 1993, he worked on a part-time basis at Guys Hospital, London, UK, in the Department of Nuclear Medicine, working for Prof. Ignac Fogelman. From 1994, he was Director of Clinical Services at Bona Fide Ltd, Madison, Wisconsin, USA (a wholly owned subsidiary of the Lunar Corporation, Madison, Wisconsin, USA), taking up his current post as Vice President, Business Development, Bio-Imaging Technologies Inc., Newtown, Pennsylvania, USA in 1999, where he is now as Senior Vice President of Medical Affairs.

Derek Pearson graduated in Physics from Nottingham University, Nottingham, UK, in 1976. He obtained a PhD in medical physics from the University of Surrey,

Guildford, UK and then spent three years as a postdoctoral fellow at the University of Leeds, Leeds, UK, working in the body composition group. He moved back to Nottingham in 1982 to work as a Senior Grade Physicist in nuclear medicine. He has been Clinical Director of Medical Physics at Nottingham City Hospital NHS Trust (now part of Nottingham University Hospitals NHS Trust, Nottingham UK) since 1989. He has a broad range of interests in imaging and radiation physics and has published more than 40 papers, most recently specializing in the measurement of bone density and bone imaging.

1

Introduction

Colin G. Miller and Derek Pearson

1.1. Why a Book about Clinical Trials in Osteoporosis?

There have been many books published about the design, conduct, and analysis of clinical trials. Why are osteoporosis trials a special case that deserve a book of their own? There are three main reasons. First, most diseases have a well-understood definition and aetiology. Osteoporosis is a disease that is understood by those working within the subspecialty, but currently there is no definition that is agreeable to both medical and scientific communities and its aetiology is poorly understood. It is within this framework that the pharmaceutical industry is trying to develop new treatments for the so-called "silent epidemic".

In layman's terms, the disease of osteoporosis is defined as "brittle bones occurring in the elderly that could lead to fractures." The classical definition was "a bony fracture caused by minimal trauma owing to a loss in bone mineral." A published consensus definition states that osteoporosis is "a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fractures."[1] The National Institutes of Health (NIH) Consensus Conference Statement on Osteoporosis Prevention, Diagnosis, and Therapy states that "osteoporosis is a skeletal disorder characterized by compromised bone strength predisposing to an increased risk of fracture."[2] The World Health Organization (WHO) operationally defines osteoporosis as "bone density 2.5 standard deviations (SDs) below the mean for young white adult women at lumbar spine, femoral neck, or forearm".[3] It is now recommended that the diagnostic use of this definition is restricted to bone density of the femur.[4] Although it is not clear how to apply this in men and children, it is recommended that the same diagnostic thresholds can be used in men.[4,5]

The NIH statement recognises that "bone strength reflects the integration of two main features: bone density and bone quality. Currently, there is no accurate measure of overall bone strength. Bone mineral density (BMD) is frequently used as a proxy measure and accounts for approximately 70% of bone strength." Thus, osteoporosis has become a disease that is characterized by measurement of BMD.

The endpoint of many clinical trials is BMD, either used as a primary endpoint in its own right or used as a surrogate marker for fracture risk.

Regulatory authorities tend to consider osteoporosis in terms of fracture when it comes to licensing new treatments for the management of the disease, and, increasingly, BMD for the prevention of osteoporosis. It is, therefore, imperative that the researcher understands which definition of the disease they are using and what the endpoint or hypothesis they are trying to evaluate is before they embark on a research programme.

Second, because osteoporosis is a disease that is diagnosed using a measurement of BMD and is monitored over many years using such measurements, there are a range of technical issues to ensure the quality and consistency of BMD measurements that must be considered. Several of these relate to the choice of equipment, standardization, and quality control before a trial begins, in addition to technical issues that must be considered throughout the life of the study.

Third, osteoporosis trials are often long-term trials carried out in normal, asymptomatic women, in whom proven drugs for the treatment and prevention of osteoporosis are already licensed. This is particularly true of clinical trials in women who are close to the menopause. This presents ethical issues because the latest version of the Declaration of Helsinki (Finland), produced in Edinburgh (UK) in 2000, specifically states that placebo control in the presence of a proven treatment is unethical.[6] This conflicts with the requirements of the US Food and Drug Administration (FDA), which still requires placebo control for licensing purposes. These women are also unlikely to gain any direct benefit from a short-term trial, which raises other ethical issues. Postmenopausal women (aged 55 to 65 years) are unlikely to have any long-term reduction in fracture risk if the fracture does not occur until they are aged 80 years. Any protective effect of treatment will have worn off. What happens at the end of the study? Will treatment still be available to subjects if a proven treatment effect is demonstrated?

In summary, the definition of osteoporosis is not universally agreed, it is a disease defined by a measurement of BMD and often clinical trials are carried out in normal, asymptomatic women. For researchers entering into this therapeutic area, it seems to be initially confusing and technically challenging. On this basis, osteoporosis clinical trials deserve a book that provides an introduction to the novice and clearly explains the design and implementation of these trials. This is not designed to be an in-depth book for the expert, but nowhere else is this overview currently available in an easy-to-find manner.

1.2. How this Book Works

The aim of this book is to lead the researcher through all the stages of a clinical trial. Section 1 covers study design (Chapter 2) and the pretrial phase, including ethical considerations specific to osteoporosis trials (Chapter 3) and the standardization and pretrial quality control required to ensure consistent measurement of BMD (Chapter 4).

Section 2 looks at the day-to-day running of the trial. Chapter 5 gives a rundown of the current regulatory framework, the organization of the trial by the sponsor, and the requirements for audit. The administrative organization of the trial is covered in Chapter 6. The endpoints in most osteoporosis trials rely on BMD measurements and morphometric measurements of vertebral height to detect vertebral fractures morphometric X-ray analysis (MXA). Good subject positioning, an understanding of the limitations of dual-energy X-ray absorptiometry (DXA) measurements, and review of DXA results are vital. These are covered in Chapter 7. At any point in the trial, participating centres could find themselves the subject of audit by the sponsor or regulatory authorities. Quality control of the equipment and biochemical markers of bone turnover is covered in Chapter 8. Several useful tools for monitoring quality control data are discussed.

Section 3 covers data analysis and presentation. Chapter 9 is a guide to writing a paper for a peer-reviewed journal to a standard that will ensure that readers will gain full benefit from your trial and the results will be easily included into subsequent metaanalysis. Sample data are given in Appendix B to enable the reader to check the worked examples in the text. The sample data are from a small placebo-controlled, double-blind study looking at the use of calcitonin in postmenopausal women. The primary endpoint is BMD, which is measured using DXA at the lumbar spine. Secondary endpoints include measuring ultrasound transmission through the heel (ultrasonometry or quantitative ultrasound [QUS]). Data for BMD and QUS are given in the appendix. Other aspects of this trial are considered at other points in the book (e.g. study design in Chapter 2 and ethical considerations and written, informed consent in Chapter 3).

Section 4 gives background information on current and future therapies for osteoporosis (Chapter 10) and considers what the ideal treatment for osteoporosis might be (Chapter 11). Help in selecting the best equipment for measuring BMD is given in Chapter 12, whereas Chapter 13 reviews the wide range of biochemical markers of bone turnover. The final chapter (Chapter 14) looks to the future to consider where clinical trials in osteoporosis might be leading.

We have attempted to use standard terminology throughout the book. This includes the sponsor (usually a pharmaceutical company that is funding the research) and the clinical research organization (CRO), which is a company responsible for administering the trial, quality assurance of the data, analysing the data, and producing the final report. In addition, sponsors often appoint an independent company to act as a quality assurance (QA) centre to review and analyse BMD results. A clinical research associate (CRA) is the representative of the sponsor or CRO who liaises between the CRO and each site participating in the trial. The investigator is the researcher at the local site who has responsibility for recruiting research subjects and running the study locally. We have chosen to refer to those taking part in osteoporosis clinical trials as "subjects" rather than "patients" for the reason that many of them are not ill, but are normal women and men. The drug under investigation is usually under development by the sponsor and is known throughout the book as the "new molecular entity" (NME).

Most chapters contain many references to other source material. These enable the reader to learn about any of the issues covered in the book in greater depth. As always, there are many abbreviations in the book, but a full glossary is given in Appendix A.

References

1. Christiansen, C. (1993). Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med* **94:**646–50.
2. National Institutes of Health Consensus Development Conference Statement: Osteoporosis Prevention, Diagnosis, and Therapy March 27–29, 2000 (http://consensus.nih.gov/2000/2000Osteoporosis111html.htm accessed 13/11/06).
3. Kanis, J.A. (1994). Assessment of Fracture Risk and its Application to Screening for Postmenopausal Osteoporosis. Report of a WHO Study Group. Geneva: World Health Organization.
4. Kanis, J.A. and Gluer, C.C. (2000). An update on the diagnosis and assessment of osteoporosis with densitometry. Committee of Scientific Advisors, International Osteoporosis Foundation. *Osteoporos Int* **11:**192–202.
5. Selby, P.L., Davies, M. and Adams, J.E. (2000). Do men and women fracture bones at similar bone densities? *Osteoporos Int* **11:**153–7.
6. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland, 1964, and revised by the 52nd World Medical Assembly, Edinburgh, Scotland, 2000 (http://www.wma.net/e/policy/b3.htm accessed 13/11/06).

Section 1

Study Design and the Pretrial Phase

2

Study Design and Endpoints

Derek Pearson and Colin G. Miller

2.1. Introduction

It is the design of your study that will give it credence in the wider scientific community. If it is clear from the dissemination of the study that there has not been a clear hypothesis, endpoints are inappropriate, and the study design is inadequate, it is unlikely that the outcome of the study will be used to contribute to the process of regulatory approval, or form part of a metaanalysis. It is important to get the design right from the start. The time taken will be very worthwhile. The purpose of this chapter is to give an overview of the following:

1. Phase I to IV studies and different trial types
2. Endpoints
3. Study design, size, and power
4. The effect of antiresorptive therapies versus true anabolic therapies on trial design
5. Inclusion and exclusion criteria
6. Study protocols and documentation
7. Sampling and randomization
8. Design of data collection.

The chapter will also discuss the advantages of true clinical and health-status outcomes compared with the more common use of surrogate outcomes.

2.2. Types of Clinical Trial

There are four main types of drug trial (phase I to IV studies; Table 2.1). This book is most relevant to phase II and phase III trials, in which the effectiveness of the drug or new molecular entity (NME) will be the primary endpoint. Trials can be either cross-sectional or longitudinal. Cross-sectional trials are usually used to investigate prevalence (i.e. the number of osteoporotic individuals in the population at a given

Table 2.1. Definition of clinical trial types.

Phase	Description	Number of subjects
I	Safety, toxicity, and pharmacokinetic studies. Normally carried out in normal volunteers. Phase I trials will include dose-ranging studies, in which the dose of the drug is increased to test tolerability and set the dose for phase II studies	80–100
II	Safety and efficacy. These are usually not randomized studies, but are to see if the dose of the drug from the phase I trial is effective and well tolerated in the subject population that the drug is designed to benefit	100–200 overall 20 in an individual study
III	Usually RCTs to demonstrate effectiveness compared with a placebo control or active comparator	Several hundred to several thousand
IV	Postregulatory approval and marketing. Monitoring of long-term safety and efficacy in a wider subject population. Postmarketing comparisons with other products	Several thousand

time) or incidence (i.e. the number of new cases of osteoporosis within a population at a given time). For example, a study of ultrasonometry in fracture cases and controls could indicate the different prevalence of osteoporosis in the fracture population or the incidence of fracture in treated versus nontreated populations. They are open to bias because, in the former example, the fracture cases and controls might be different in other ways. If the control population is younger, the difference in the prevalence of osteoporosis could be because of age. Other factors, such as body weight, smoking habits, or previous drug history, might be confounding the results. Because of this, the majority of clinical trials in osteoporosis are longitudinal.

There are a number of different types of longitudinal study, as follows:

Cohort studies. In such studies, a group of subjects is selected because of risk or exposure to a factor being studied. Phase IV postmarketing trials in osteoporosis are a type of cohort study. Subjects who are at risk of osteoporosis or have established osteoporosis are treated with a licensed, effective treatment and followed up to measure long-term side effects and see if any gain in bone mineral density (BMD) is maintained over a long period.

Case-controlled studies. In case-controlled studies, a group of subjects with osteoporosis, for example, are identified as "cases". A control group who do not have the disease, but are similar to the cases in other factors, is selected. For example, the controls might be matched for age, menopausal status, and weight. Such a study might follow the change in BMD over a number of years in controls (who might be expected to lose BMD) and cases (who might be expected to gain bone because of treatment). Clearly, in evaluating new treatments, such studies could well introduce bias because of poor matching between cases and controls. In such studies, it is important that the investigator evaluating the endpoints is blinded to the subject group.

Randomized, controlled trials (RCTs). The majority of osteoporosis trials will fall into this category. Subjects are randomly assigned to the treatment or control

(placebo or active comparator) group and monitored over a number of years. The length of follow-up will depend on regulatory requirements (Chapter 5) and the endpoint chosen. Investigators and subjects will normally be blinded to the treatment (double-blinding), although in some trials with intravenous delivery of bisphosphonates, for example, it might only be possible to blind investigators to the treatment (single-blinding).

Cross-over trials. A cross-over trial is a trial in which subjects are randomly assigned to the treatment or control group and then, after a fixed time followed by a period of washout for the active drug, subjects are swapped from the treatment to control group, or vice versa. Such trials are not often used in osteoporosis because of the length of time treatments take to have a measurable outcome and the length of washout period that would be required.

Factorial designs. These are randomized trials in which more than one treatment is tested alone and in combination against each other and placebo. Although the majority of osteoporosis trials are the traditional RCTs, in future combination therapies that combine antiresorptive and true anabolic agents might be more common, leading to an increase in the use of factorial designs.

Equivalence or noninferiority trials. When current therapies come off patent protection, there are a number of generic products that are produced that require equivalence testing. For bisphosphonates, which are considered drugs as opposed to biologics, standard pharmacokinetic studies can be performed. For products such as recombinant parathyroid hormone (PTH), noninferiority or superiority trials will have to be conducted.

There is often confusion between the use of the terms "efficacy" and "effectiveness" when considering clinical trial design. A trial investigating efficacy aims to determine whether the NME works in those who receive it. By implication, these trials have to be carefully controlled to ensure compliance with treatment and work best if the investigator has control over the administration of the NME, for example by intravenous administration. Phase II trials are often designed to investigate the safety and efficacy of an NME. Effectiveness trials determine whether the NME works in subjects who are offered it. This is much more similar to real life, in which subjects often do not comply with treatment. Less control over the administration of the NME is required, but compliance and subject drop-out could indicate the acceptability of the treatment in the wider subject population.

The US Food and Drug Administration (FDA) distinguish between clinical trials for the prevention of osteoporosis and those for the treatment of established osteoporosis. Their expectation is that effectiveness has to be demonstrated in subjects with established osteoporosis, in addition to data from pre-clinical studies in animal models, before investigators can begin trials in the prevention of osteoporosis.[1]

The first stage in the design of the clinical trial is to agree the main purpose of the study and choose a trial design that will answer the research hypothesis. The objective of the sample study included with this book was to determine the effectiveness of calcitonin in preventing bone loss in postmenopausal women.

The trial was designed, therefore, as a randomized, double-blind, placebo-controlled trial.

2.3. Endpoints

The choice of endpoint for a trial is of crucial importance because it determines the value of the trial in the long term. If the wrong endpoint is chosen, the trial could be excluded from subsequent metaanalysis or it could be less than ideal for inclusion in a submission for regulatory approval. The objectivity of the outcome measure and the blinding of those assessing outcome to the treatment are important in selection of trials for inclusion in metaanalysis.[2] Endpoints should be as follows:

1. Relevant to the disease and the population under investigation. Fracture, for example, is relevant to the study of osteoporosis because it is fracture that is the cause of the mortality and morbidity associated with the disease. It would not be relevant as a measure of outcome in a perimenopausal population in an osteoporosis prevention trial. The rate of fracture would be too low, requiring a vast number of subjects in the study to demonstrate a significant treatment effect.
2. Acceptable to the scientific community and regulatory authorities. New types of equipment become available all the time and are offered to centres by manufacturers for evaluation. It would be wrong, however, not to use a commonly accepted method of measuring BMD as a primary endpoint. The use of a novel technique would rule a study out of future metaanalysis and make it unlikely to be recognised by the regulatory authorities. It is perfectly acceptable to use new techniques as secondary endpoints, in order to validate their use in clinical trials.
3. Responsive to change. It is well known that BMD at the spine will demonstrate a more rapid response to treatment than BMD at the femur, forearm, or heel. Thus, BMD at the spine is often chosen as a primary endpoint to reduce the length of a study or demonstrate maximum treatment effect. Change in BMD does not correlate well to change in fracture risk, because small changes in BMD can result in large changes in fracture incidence. For example, in established vertebral osteoporosis, treatment with alendronate resulted in a 50% reduction in subsequent vertebral fracture, but only a small percentage increase in BMD.[3] If using ultrasonometry as an endpoint, however, it would not be appropriate to use the velocity of sound (VOS) or speed of sound (SOS) as a primary outcome measure. This is because the range of VOS and SOS *in vivo* is small and the responsiveness to change in BMD is poor. For example, the 4-year treatment effect of hormone-replacement therapy (HRT) is only 1.1% for SOS compared with 6.4% for broadband ultrasound attenuation (BUA).[4] Questionnaires for assessment of health status must have adequate powers of discrimination within the subject population studied. It might, therefore, be important to use questionnaires validated in your subject population to ensure that they will be sensitive to the expected changes in health status. This could well be difficult in osteoporosis prevention studies, in which there might be no apparent benefit from treatment. The rate of change is

also important in choosing the endpoint and design of a study. It must be possible to demonstrate the change expected in the outcome measure within the timescale of the study.

4. Reproducible. The outcome measure must give the same result if the subject is measured more than once on the same visit. The reproducibility of dual-energy X-ray absorptiometry (DXA) and quantitative ultrasound (QUS) is discussed at length elsewhere (Chapter 4), but the same applies to other outcome measures chosen, for example health-status questionnaires or vertebral morphometry.

5. Widely available, particularly when considering multicentre trials. Although crosscalibration between DXA instruments is possible, sponsors could well want to standardize the make and model of DXA and the version of analysis software. The more widely available equipment is, the easier it is to establish a multicentre trial and the more likely that the measurement will be acceptable to the scientific community and regulatory authorities.

6. Reliable and acceptable to subjects. In long-term clinical trials, subject drop-out is a significant problem. Using an endpoint that is a simple assessment, does not require multiple visits to hospital, and is acceptable to subjects in terms of comfort and convenience will reduce the problems of drop-out from the study. Much of this will depend on the skills and training of those administering the outcome measure and sponsors are advised to invest in staff training to help ensure subject compliance. The reliability of the equipment used to assess the outcome will reduce the number of cancelled appointments and improve compliance.

There are two types of endpoint commonly used in clinical trials in osteoporosis.[5] The first category is the true clinical outcome, which includes fracture or the assessment of health status. The second category is the surrogate outcome, such as BMD, morphometric X-ray analysis (MXA), or biochemical markers. If possible, measures of true clinical outcome should be used. Intermediate and surrogate endpoints should be avoided because they can be misleading. This is a significant issue in osteoporosis research, because the majority of studies use BMD as a surrogate for fracture risk. Fracture (a true clinical endpoint) is only required as an endpoint by the FDA in phase III studies of nonoestrogen-based NMEs used for the treatment of osteoporosis;[1] the draft European guidelines[6] also require fracture information to be available for licensing NMEs for the treatment of osteoporosis. Measurement of BMD by DXA is now so well accepted as an appropriate surrogate endpoint that it is approved as an endpoint for phase III studies of the prevention of osteoporosis and the treatment of osteoporosis using oestrogen-based NMEs. The risk of relying on surrogate endpoints is that unproven technology is introduced for widespread clinical use outside the setting of the clinical trial without adequate validation. An example of this is, ultrasonometry, where BUA is reduced in a fracture population, correlates (poorly) to BMD at central skeletal sites (spine and femur), and demonstrates a response to treatment, all within the setting of clinical trials in which it is used as a surrogate outcome for bone quality. In a population, ultrasonometry might perform as well as femoral DXA in predicting hip fracture risk, but its use in an individual for diagnosis and monitoring of osteoporosis is far from clear.

Thus, the main endpoints used within osteoporosis trials[4,6] are as follows:

1. Fracture. The endpoint should be based on the incidence of new fractures and not the worsening of previous fractures. Both prevalent and incident vertebral fractures should be recorded. Fractures must be defined using clear criteria[7–9] because there is an ongoing debate about the definition of vertebral fracture. Serial radiographs and morphometric assessments of the spine should be used. It is important that the number of subjects who experience a new fracture is the outcome measure, rather than the total number of new fractures. The latter could introduce bias into the study if a small number of subjects experience a large number of fractures.

2. BMD. The method of choice is DXA at the site of osteoporotic fracture at the spine or hip. Peripheral measurements at this stage do not meet all the desirable characteristics for an endpoint.

3. Health-outcome questionnaires that review symptoms, activities of daily living, involvement in exercise, and emotional and social status. There are some questionnaires that are specific to osteoporosis,[10–13] e.g. the Osteoporosis Quality of Life Questionnaire[10] and the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO),[13] and others that are general but might have application in the field of osteoporosis, such as the SF-36 Health Status Questionnaire[14] and the European Quality of Life Questionnaire.[15] The questionnaires have been validated to different degrees and might require further validation in RCTs.

4. Biochemical markers of bone turnover are an appropriate endpoint for phase II dose-finding trials.[6] However, because of the unresolved issues of assay standardization and a poor coefficient of variation associated with biochemical markers (Chapter 13), they are not indicated for use as a primary endpoint in trials for the treatment of primary osteoporosis.

5. Height and stature.[5] Height can be used as a surrogate for vertebral fracture because loss of height is a feature in vertebral osteoporosis. It is of most use in trials involving older women. Kyphosis has also been used within a community-based study, but not within RCTs.

2.4. Study Design, Size, and Power

Having determined the type of study to be carried out and chosen the endpoint, it is important to determine objectively the size and power of the study. The advantage of osteoporosis trials is that the outcome is often a proportion (e.g. fracture rate or results from a quality-of-life questionnaire) or quantitative (e.g. BMD or biochemical markers). There is also plenty of information in the literature about the variation of quantitative measures, fracture rate, and expected treatment effects that enables the researcher to easily make assumptions about the outcome of a clinical trial. There is less information about the reproducibility and validation of quality-of-life information, making design of such studies somewhat more difficult.

To calculate the number of subjects required in the study, the *type I* and *type II* errors must be set. The *type I* error is the risk that the trial will give a false-positive result, that is that a treatment effect will be demonstrated when the treatments used are equally effective. This is given the symbol α and is usually set at 5%, that is there is a 95% probability that the treatment effect is real. The *type II* error is the risk that the trial will give a false-negative result, that is no treatment effect is demonstrated when there is a real difference between treatments or the treatment and placebo. This is given the symbol β and is often set to 10% or 20%. The value $1-\beta$ is known as the power of the study and will usually be 90% or 80%. To reduce the risk of false-positive results (i.e. reduce the risk of *type I* error), α can be reduced to 2% or 1%, but this leads to bigger trials. The *type II* error can also be reduced, but this also increases the number of subjects that must be entered into the study. The danger of a small trial is there is the possibility, with only a few withdrawals, that a large proportion of the subjects do not complete the trial. This means that the risk of a false-negative result is high, that is the *type II* error in a small trial is high.

2.4.1. Studies of Fracture Risk

In a study of vertebral fracture, the incidence of new vertebral fracture is about 20% over 3 years in osteoporotic women who have a mean age 64 years and at least one prevalent nontraumatic vertebral fracture.[16] With bisphosphonate treatment, a reduction of 50% in fracture rate could be reasonably expected.[3] In this example, 20% of the placebo group and 10% of the treatment group will be expected to fracture over the duration of the study:

$$p_1 = 20\%$$
$$p_2 = 10\%.$$

The number of subjects required in each group can be calculated as follows:

$$n = \frac{p_1 \times (100 - p_1) + p_2 \times (100 - p_2)}{(p_2 - p_1)^2} \times 10.5 = 263, \qquad (2.1)$$

assuming (a *type I* error) $\alpha = 5\%$ and (a *type II* error) $\beta = 10\%$. The value 10.5 depends on the values of α and β chosen and is tabulated.[17] Reducing the power of the trial from 90% to 80% (i.e. $\beta = 20\%$) changes the value of 10.5 to 7.9 and the number of subjects in each group to 198, but the probability of a false-negative result from the trial is increased.

2.4.2. Trials using Quantitative Endpoints

It is sensible to use *a priori* knowledge about quantitative endpoints in the design of clinical trials. For example, as a rule of thumb, in any group of subjects, the

mean BMD will be between 0.8 g/cm^2 and 1.0 g/cm^2, with a standard deviation (SD) of 0.1 g/cm^2. How much will treatment with calcitonin affect the mean BMD? A study of calcitonin showed a gain of 7.8% in the treatment group (200 IU of nasal salmon calcitonin) compared with placebo.[18,19] Assuming the baseline BMD in the treatment group is 0.85 ± 0.1 g/cm^2, the difference between the treatment and placebo groups at the end of the study is 0.067 g/cm^2. The number of subjects required in each group is as follows:

$$n = \frac{2\sigma^2}{(\mu_1 - \mu_2)^2} \times 10.5, \tag{2.2}$$

where σ is the SD and $(\mu_1 - \mu_2)$ is the expected difference in the mean BMD between treatment and control groups at the end of the trial. Substituting the figures from the example above, 47 subjects would be required in each group to demonstrate a significant treatment effect, assuming (a *type I* error) $\alpha = 5\%$ and (a *type II* error) $\beta = 10\%$. The value 10.5 is derived from tables, as before. To allow for drop-out, it would be wise to recruit approximately 55 subjects in each group to ensure that an adequate number complete the study. In our sample study, to simplify the example calculations, results are given for 30 subjects in each group (Appendix B).

2.4.3. Equivalence Trials

Because there are now established drugs for both the prevention of postmenopausal osteoporosis and the treatment of established osteoporosis, it might be ethical to consider the use of an equivalence trial using an active comparator rather than a placebo control. There are many difficulties with the design of equivalence trials because many of the problems that occur in the running of clinical trials tend to bias the trial away from demonstrating a significant treatment effect and towards equivalence.[20] This includes subject withdrawal, missing data, and subjects who are lost to follow-up. The sponsor must ensure that the effects of these are minimized when running the trial.

In order to design an equivalence trial, the size of a clinically relevant equivalence margin must be agreed in advance. This might be, for example, the change in BMD that might be expected if there was no clinically significant difference between the NME and the active comparator. This can be based on the confidence limits of the treatment effect compared with placebo using the active comparator or other similar drugs. For example, suppose an active treatment shows a mean increase in BMD of 6% with 95% confidence limits of approximately ±2%. This could be used as the basis of the clinically relevant equivalence margin. If it is now considered as the active comparator in a new trial, the confidence limits of the mean increase in BMD for the NME would have to lie completely within these confidence limits for equivalence to be demonstrated (i.e. between 4% and 8%). For a noninferiority trial, in which the aim of the trial was to demonstrate that the NME

was no worse than the active treatment, the confidence limits of the mean increase in BMD would have to lie above the lower limit of the clinically relevant equivalence margin (i.e. above 4%). This is illustrated in Figure 2.1.

In order to calculate the study numbers in an equivalence trial, the clinically relevant equivalence interval has to be agreed. In this case, let us assume that this is 3%. The sample size[21] is calculated using the following formula:

$$n = \frac{2\sigma^2 (Z_\alpha + Z_\beta)^2}{d^2}, \tag{2.3}$$

where d is the clinically relevant equivalence margin and Z_α and Z_β are the standardized normal deviates for the *type I* and *type II* errors, respectively. These are tabulated in statistics books, but for $\alpha = 5\%$ and $\beta = 20\%$, values commonly chosen for this type of study, they are 1.64 and 1.28, respectively. If the SD of BMD is 0.1 g/cm^2 with a mean of 0.9 g/cm^2, as above, the number of subjects required in each group to demonstrate equivalence is as follows:

$$n = \frac{2 \times 0.1^2 \times (1.64 + 1.28)^2}{(0.03 \times 0.9)^2} = 234.$$

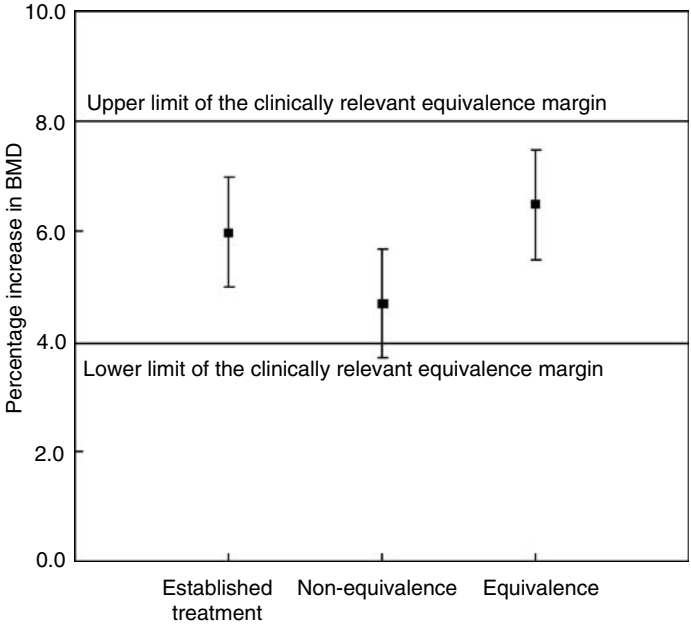


Figure 2.1. The upper and lower clinically relevant equivalence margins for the example in the text. A new drug, for which the confidence limits on the treatment effect overlapped the clinically relevant equivalence margin, would not demonstrate equivalence. Where the confidence limits lie completely inside the relevant equivalence margins, equivalence is demonstrated.

If considering equivalence trials in fracture studies, the equation is as follows:

$$n = \frac{(Z_\alpha + Z_\beta)^2\left[p_1 \times (100 - p_1) + p_2 \times (100 - p_2)\right]}{(p_2 - p_1 - d)^2}. \tag{2.4}$$

In this case, p_1 and p_2 are equal to 20%. Let us assume that 5% would be a clinically relevant equivalence interval; then, Equation 2.4 is as follows:

$$n = \frac{2 \times (Z_\alpha + Z_\beta)^2 \times p \times (100 - p)}{d^2} \tag{2.5}$$

$$n = \frac{2 \times (1.64 + 1.28)^2 \times 20 \times 80}{5^2} = 1091.$$

The number of subjects required to demonstrate equivalence is higher than that required to demonstrate a significant difference from placebo. This is why sponsors are often reluctant to move away from placebo-controlled trials, even if there is a proven treatment. The costs of such trials are much higher, but they might be the only ethical option in long-term trials for the prevention or treatment of osteoporosis.

2.4.4. Other Factors Affecting Study Numbers

The number of subjects required in each group calculated above does not take into account other factors that affect study numbers. The number of subjects recruited to the study is affected by the following:

1. The number of subjects eligible for the trial locally
2. The number of subjects that can be recruited to the trial
3. The entry rate
4. The drop-out rate and noncompliance with treatment increasing the risk of a false-negative result.

Fortunately, the numbers of subjects eligible and that can be recruited to osteoporosis trials are usually high, and recruitment is not a problem. Osteoporosis has a high profile in the minds of the public and there is a high level of public interest and information available through newspapers and magazines. The risk is more that subjects will be recruited to multiple osteoporosis trials and overresearched because there is interest in osteoporosis from a number of medical specialties. If the trial is large, the issue of entry rate into the trial is significant. The recruitment period should ideally be as short as possible compared with the length of the trial, and this depends on the enthusiasm of local investigators to commit to trial. The aim should be to keep the recruitment period to within 2 years.

Drop-out is another significant problem in trials. Adjust the study numbers to ensure that the number completing the trial is adequate by reviewing the literature

for other trials using similar study drugs. Drop-out can occur for a number of reasons, including side effects and inconvenience. Compliance with HRT is a problem, with between 48% and 62% subjects reportedly stopping HRT after 12 months,[22,23] although this level of compliance has been reported after up to 5 years of treatment.[24] Compliance with bisphosphonate treatment is higher,[25] with up to 89% of subjects treated still taking the drug after 3 years. Fortunately, the study population recruited in osteoporosis is well motivated, and BMD measurements improve compliance.[26]

2.5. The Effect of Antiresorptive Therapies Versus True Anabolic Therapies on Trial Design

NMEs are being developed that have a true anabolic (i.e. bone-forming) effect rather than acting as an antiresorptive therapy, for example PTH. The basic fundamentals of study design are the same for both types of compound.

There is an argument that the anabolic compounds will affect the DXA results differently. They have been shown to increase BMD, measures of biomechanical strength, histomorphometry, and cortical thickness.[27–29] A true anabolic compound, it is argued, will increase the bone area, with the effect that the gain in BMD (which is areal density measured in grammes per square centimetre) is reduced. However, this is likely to be an artefact related to edge detection in the DXA software because bone formation is likely to be on the inner edge of the cortex rather than on the outer surface.

Histomorphometric and biomechanical changes have been demonstrated after treatment with PTH,[28,29] which raises the issue of changes in bone quality, in addition to density. This could affect the choice of endpoint, with techniques that measure aspects of bone quality (e.g. ultrasonometry) or discriminate between trabecular and cortical bone (e.g. quantitative computed tomography) becoming the monitoring tools of choice for these compounds.

2.6. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria govern the selection of subjects for the study. They must be ethical, avoid introducing bias into the study, and provide a population that is representative of the subject group in which the NME is intended for clinical use, without introducing too many confounding factors. Subjects must meet all the inclusion criteria to enter the study, but will be excluded from the study if they meet any one of the exclusion criteria. Typical inclusion and exclusion criteria are given in Table 2.2 for a clinical trial investigating the use of a bisphosphonate to treat established osteoporosis. The inclusion criteria, therefore, define the target population for the study as postmenopausal osteoporotic women. The exclusion criteria include avoiding those subjects who might have an adverse reaction to a bisphosphonate. These include those who have preexisting gastrointestinal problems

Table 2.2. Typical inclusion and exclusion criteria for a clinical trial of a bisphosphonate for the treatment of osteoporosis.

Inclusion criteria:
Women aged 55–75 years at baseline
Postmenopausal for at least 2 years
Osteoporotic at lumbar spine or total hip using WHO criteria. NHANES normal reference data
Written informed consent
Exclusion criteria:
Peptic ulcer disease
Dyspepsia
Abnormal renal function
Major medical problems that would preclude participating in a trial lasting 2 years
Severe malabsorption syndrome
Uncontrolled hypertension
Myocardial infarction within 6 months
Unstable angina
Disturbed thyroid or parathyroid function
Prior treatment with HRT or bisphosphonate within the last 6 months

because of the known cases of oesophagitis associated with bisphosphonates. Those with renal insufficiency are also excluded because of contraindications for the use of bisphosphonates. The exclusion of those who have had prior treatment for osteoporosis could introduce study bias because it will tend to overestimate the treatment effect in the study population compared with wider routine clinical use. The danger of having too many exclusions in phase III studies is that the study population is far too unrepresentative of the clinical situation, making the results of the study invalid when applying them to an individual subject within the clinic.

2.7. Study Protocol and Associated Documentation

A good, clear study protocol is a must for a properly conducted clinical trial. The key elements of a good protocol are given in Table 2.3. Reviewing well-reported trials in the literature and considering the requirements for dissemination[30] will help in the preparation of a good protocol.

The protocol should be considered as controlled documentation and should have a unique study number, version number, and date. Other controlling information should include the contact details of all those involved in the trial [e.g. the sponsor and clinical research organization (CRO)] and signatures of those persons authorized to approve the protocol on behalf of the sponsor. The International Committee for Harmonization (ICH) guidelines for good clinical practice (GCP)[31] contain a checklist that is helpful in the preparation of a good protocol. The guidelines also include guidance on other study documentation that is essential if the trial is going to be used for an application for regulatory approval for an NME, but might not be required if it is a trial run within a single institution. This includes the investigator's brochure and a document that gives the history of an NME to date. It will include a description of the chemical properties and formulation of an NME and the results

Table 2.3. Requirements for a study protocol.

Section	Description
1	Introduction Including a study overview, study rationale, previous use in humans, reported toxicity, and pharmacokinetics, study context, and scope
2	Objectives Including primary and secondary objectives and endpoints
3	Study design
4	Selection of study population Inclusion criteria Exclusion criteria Removal of subjects from therapy
5	Study treatments Dose regimen Storage Packaging and labelling Administration Concomitant treatment Treatment compliance
6	Study procedures Screening visits Randomization Visit schedule Assessments Study flow chart
7	Serious adverse events and adverse-event reporting
8	Statistical considerations Determination of sample size Definition of study populations for analysis Demographic and background characteristics of study groups Analysis plan
9	Ethical considerations Indemnity
10	Quality assurance and quality control Monitoring Quality assurance of BMD data Study audit
11	Administrative procedures Data handling and record keeping Discontinuation procedures Dissemination of results
Appendices	Including ICH GCP compliance, subject information sheet, consent form, GP letters, etc

of animal testing and use in humans, including pharmacokinetic, safety, and efficacy data. This will inform the investigator about the possible risks of using an NME and the safety and monitoring required as part of the trial. The list of required documentation for regulatory approval is extensive and runs to more than 11 pages! The aim of the documents is to enable the sponsor and investigator to demonstrate

compliance with the guidelines. Many of them, however, will assist the sponsor, CRO, and local investigator to run and monitor the trial and assure the sponsor and regulatory authority of the integrity of the trial data. This documentation might also be the subject of audit by the Institutional Review Board (IRB) or Independent Ethics Committee (IEC), sponsor, or regulatory authorities. It pays, therefore, to ensure that the administrative arrangements for the trial are in good working order so that documentation is readily available and up to date.

2.8. Sampling and Randomization

In recruiting subjects to a trial, the CRO should consider the total number of subjects required for the trial. This will give an indication of the number of centres that will need to be involved in a multicentre trial. CROs, however, should be aware that centres could be participating in multiple trials, which could limit the accrual rate of subjects into the trial at particular centres. Centres that have a good research reputation in osteoporosis are already likely to have identified databases of subjects that can be used for recruitment and a high profile locally among affected subjects. The CRO must consider the ability of the local investigator to cope with the running of the trial, particularly if the trial requires a high accrual rate of subjects selected randomly from local primary care physicians. Recruitment of subjects with established osteoporosis through an osteoporosis clinic will be easier than recruitment to a prevention trial. Radio, television, and press adverts can be used, but this method of recruitment will be biased towards women from higher social classes. Even if a random sample of women is approached through primary care, there will be a better response from more affluent areas without ethnic minorities. It might be necessary to stratify recruitment for referring primary care practices to ensure that a representative group of "at-risk" women is obtained.

"Randomization" is the process whereby trial subjects are assigned to the treatment or control group. Control of the randomization process should be out of the hands of the local investigators. Randomization lists should be held by the CRO and investigators should telephone to obtain a randomization code. In order to blind the medical and nursing staff in the trial to the treatment, randomization is often managed through the pharmacy, which will either contact the CRO direct or, for a local study, hold a series of sealed envelopes containing the randomization codes for a study. It must always be possible, however, for local investigators to break the blinding if there is a medical emergency, and the mechanism for blinding, randomization, and breaking the code should be documented in the study protocol. There are three common methods of randomization:

1. Simple randomization
2. Block randomization
3. Stratified randomization.

Simple randomization is a coin-tossing approach. If the coin lands heads up, then the subject is allocated to the treatment group, but if the coin lands tails up, the

subject is allocated to the control group. Rather than use a coin, it is simpler to use random numbers generated by a computer. A separate randomization list should be held for each centre to avoid confounding factors such as a large proportion of the trial subjects at one centre entering the treatment group. The advantage of simple randomization is that the treatment allocation is unpredictable. Conversely, there is a danger that there will be unequal numbers in the treatment and control groups. If, for example, 100 subjects are recruited, there is a 5% probability that there will be only 40 subjects in the treatment group and 60 subjects in the control group.[17]

The aim of block randomization is to ensure that there are equal numbers of subjects in the treatment and control groups. Ideally, large block sizes should be avoided because this is more likely to lead to unequal numbers in the control and treatment group. Block sizes up to a maximum of six subjects should be used. Within a block of four subjects, for example, it would ensure that two subjects are in the treatment group and two subjects are in the control. There are six combinations of treatment and control in each block:

- TTCC
- TCTC
- CTCT
- CCTT
- TCCT
- CTTC,

where T is a subject in the treatment group and C is a subject in the control group. A block is chosen at random and then used to enter the first four subjects into the study. The next four subjects are entered using a second randomly chosen block, and so on. Alternatively, a random number is attached to each code in a block, for example as follows:

- T 0.173
- T 0.872
- C 0.347
- C 0.089

The block would then be sorted in ascending order of the random numbers, giving CTCT. The subjects would then be allocated to the study in this order. The order of the next block would be chosen in the same fashion. The advantage of block randomization is that there will always be equal numbers of subjects in each group, even if the study is stopped early following an interim analysis.

It might be important to use stratified randomization, where the balance of certain subject characteristics (prognostic factors) is ensured between the groups. For example, it might be important to ensure that there are a similar number of osteoporotic subjects in the treatment and control groups. Randomization would be stratified to ensure the same proportion of osteoporotic subjects in each group (i.e. those with a BMD T-score ≤ -2.5). Randomization would follow a baseline BMD measurement at a screening visit. Within each stratum, subjects would be randomized using simple or block randomization. Once recruitment to a particular

stratum of the design was complete, no further subjects with those characteristics would be recruited. Subjects can be stratified for age, sex, or other prognostic factors that might affect outcome. The prognostic factors must be relevant and a literature search can be helpful to identify them. There should not be too many prognostic factors in a stratified randomization, or accrual rates to the trial could become an issue.

2.9. Design of Data Collection

The choice of primary and secondary endpoints will determine the requirements for data collection. The endpoints should be reliable and validated in the population being studied. It is important that data collection is well organized and unambiguous. Forms to be completed by the investigator for each visit and a clear list of data requirements are a necessity. These forms should be included in the study protocol. If possible, in studies with large amounts of data, forms should be designed so that the majority of the results can be scanned into the computer. This requires the investigator to put a reasonable amount of effort into form design before the trial commences, but it is usually well worth the effort. Many of the questions can be reduced to tick boxes and simple numerical fields that can be scanned by computer. The need to produce well-designed forms will also encourage the investigator not to collect too much unnecessary data! It is always helpful to specify the position of the decimal point and the units of numerical data so that it is quite clear what is required, for example a BMD measurement as follows:

$$[\,]\,.\,[\,][\,][\,]\ \mathrm{g/cm^2}.$$

Many CROs require BMD results to be analysed by a central laboratory and will set up arrangements for the transfer of original scan data and DXA quality control data by optical disk or CD ROM. There must be good logistical arrangements for the prompt transfer of data, to ensure that any problems with the data are ironed out as soon as possible after the scan visit. It is important, however, for centres to review DXA scans immediately after acquisition to ensure that they meet basic quality criteria for subject positioning, image artefacts, and data analysis (Chapter 7) because a repeat scan is sometimes necessary.

When using questionnaires, ensure that they are valid in the population studied. There are a large number of validated questionnaires[10–15] for assessing osteoporosis and quality of life, and there is usually no need to reinvent the wheel. If a questionnaire has to be developed, take time to pilot the questionnaire and validate it. It is often possible to begin with an open questionnaire, discover the issues in a pilot sample of subjects, and then refine the answers into closed questions that can have coded answers suitable for quantitative analysis. Ensure that the questions are written in plain English and, during piloting, ensure that subjects understand the questions fully.

If qualitative methods of data collection are to be used (e.g. focus groups or semistructured interviews), use the following criteria to ensure the standard of the research:

1. Is there a clear research question?
2. Is there a clear process for data collection? What is the setting? How are participants approached? What are the inclusion criteria?
3. Is there a topic guide for the interviewers and facilitators?
4. How is confidentiality maintained? If interviews and focus groups are taped, how will tapes be stored securely and transcripts be made?
5. Check transcripts of sessions with participants (respondent validation) to see if they agree that the transcript accurately represents what was said.
6. Is there a clear process of data analysis? There are many validated methods of coding and analysing transcripts of interviews.
7. Ask an independent investigator to carry out data analysis, if possible.
8. Acknowledge your own biases in carrying out the research to avoid coming to the same conclusion as your preconceived ideas.
9. Check one qualitative method against another. Are the themes coming out of focus groups similar to those from semistructured interviews?

It is often thought that qualitative research is the easy option. It is not—you will need to set aside 10 hours of data analysis for each 1 hour of interview and get expert help. The use of qualitative research methods is growing, with greater emphasis on subject experience and subject preference, but is not the primary purpose of this book. Readers are referred elsewhere for greater detail of the use of qualitative research methods.[32,33]

2.10. Summary

The design of the study, choice of endpoints, and a good study protocol are crucial to the success of any clinical trial. In this chapter, the main stepping stones to the successful design of a trial have been outlined. It is important for the sponsor to ensure that adequate time is given to the framing of the research question, the choice of endpoint, and the writing of the study protocol. Time taken in advance of the trial will pay off in the smooth running of the trial and in the quality of data collected and available for analysis.

References

1. FDA Guidelines for the Preclinical and Clinical Evaluation of Agents used in the Prevention or Treatment of Postmenopausal Osteoporosis (1994). http://www.fda.gov/cder/guidance/osteo.pdf accessed 13/11/06
2. Deeks, J., Glanville, J. and Sheldon, T. (1996). *Undertaking Systematic Reviews of Research on Effectiveness. CRD Report 4, NHS Centre for Reviews and Dissemination*. York: University of York.

3. Liberman, U.A., Weiss, S.R., Broll, J. *et al*. (1995). Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med* **333:**1437–43.
4. Sahota, O., San, P., Cawte, S.A. *et al*. (2000). A comparison of the longitudinal changes in quantitative ultrasound with dual-energy X-ray absorptiometry: the four-year effects of hormone replacement therapy. *Osteoporos Int* **11:**52–8.
5. Cranney, A., Tugwell, P., Cummings, S. *et al*. (1997). Osteoporosis Clinical Trial Endpoints: Candidate Variables and Clinimetric Properties. *J Rheumatol* **24:**1222–9.
6. Notes for Guidance on Postmenopausal Osteoporosis in Women. Committee for Proprietary Medicinal Products (CPMP), 2001. http://www.emea.eu.int/pdfs/human/ewp/055295en.pdf accessed 13/11/06
7. Eastell, R., Cedel, S., Wahner, H.W. *et al*. (1991). Classification of vertebral fractures. *J Bone Miner Res* **6:**207–15.
8. Black, D.M., Cummings, S.R., Stone, K. *et al*. (1991). A new approach to defining normal vertebral dimensions. *J Bone Miner Res* **6:**883–92.
9. McCloskey, E.V., Spector, T.D., Eyres, K.S. *et al*. (1993). The assessment of vertebral deformity: A method for use in population studies and clinical trials. *Osteoporos Int* **3:**138–47.
10. Cook, D.J., Guyatt, G.H., Adachi, J.D. *et al*. (1999). Development and validation of the mini-osteoporosis quality of life questionnaire (OQLQ) in osteoporotic women with back pain due to vertebral fractures. Osteoporosis Quality of Life Study Group. *Osteoporos Int* **10:**207–13.
11. Randell, A.G., Bhalerao, N., Nguyen, T.V. *et al*. (1998). Quality of life in osteoporosis: reliability, consistency, and validity of the Osteoporosis Assessment Questionnaire. *J Rheumatol* **25:**1171–9.
12. Helmes, E., Hodsman, A., Lazowski, D. *et al*. (1995). A questionnaire to evaluate disability in osteoporotic patients with vertebral compression fractures. *J Gerontol A Biol Sci Med Sci* **50:**M91–8.
13. Lips, P., Cooper, C., Agnusdei, D. *et al*. (1999). Quality of life in patients with vertebral fractures: validation of the Quality of Life Questionnaire of the European Foundation for Osteoporosis (QUALEFFO). Working Party for Quality of Life of the European Foundation for Osteoporosis. *Osteoporos Int* **10:**150–60.
14. Ware, J.E., Jr and Gandek, B. (1998). Overview of the SF-36 Health Survey and the International Quality of Life Assessment (IQOLA) Project. *J Clin Epidemiol* **51:**903–12.
15. Hurst, N.P., Kind, P., Ruta, D. *et al*. (1997). Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* **36:**551–9.
16. Tilyard, M.W., Spears, G.F., Thomson, J. *et al*. (1992). Treatment of postmenopausal osteoporosis with calcitriol or calcium. *N Engl J Med* **326:**357–62.
17. Pocock, S.J. (1983). *Clinical Trials: A Practical Approach*. Chichester: John Wiley & Sons.
18. Reginster, J.Y., Deroisy, R., Lecart, M.P. *et al*. (1995). A double-blind, placebo-controlled, dose-finding trial of intermittent nasal salmon calcitonin for prevention of postmenopausal lumbar spine bone loss. *Am J Med* **98:**452–8.
19. Cranney, A., Tugwell, P., Zytaruk, N. *et al*. (2002). Meta-analyses of therapies for postmenopausal osteoporosis. VI. Meta-analysis of calcitonin for the treatment of postmenopausal osteoporosis. *Endocr Rev* **23:**540–51.

20. Hubbard, W.K. (1998). International Conference on Harmonization: Guidance on Statistical Principles for Clinical Trials. *Federal Register* **63:**49583–98 http://www.fda.gov/cder/guidance/91698.pdf accessed 13/11/06
21. Kim, J.S. (1997). Determining Sample Size for Testing Equivalence. *MDDI* May:114–17 http://www.devicelink.com/mddi/archive/97/05/020.html accessed 13/11/06
22. Ettinger, B., Pressman, A. and Silver, P. (1999). Effect of age on reasons for initiation and discontinuation of hormone replacement therapy. *Menopause* **6:**282–9.
23. Faulkner, D.L., Young, C., Hutchins, D. *et al*. (1998). Patient noncompliance with hormone replacement therapy: a nationwide estimate using a large prescription claims database. *Menopause* **5:**226–9.
24. Vestergaard, P., Hermann, A.P., Gram, J. *et al*. (1997). Improving compliance with hormonal replacement therapy in primary osteoporosis prevention. *Maturitas* **28:**137–45.
25. Cummings, S.R., Black, D.M., Thompson, D.E. *et al*. (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* **280:**2077–82.
26. Cole, R.P., Palushock, S. and Haboubi, A. (1999). Osteoporosis management: physicians' recommendations and women's compliance following osteoporosis testing. *Women Health* **29:**101–15.
27. Kurland, E.S., Cosman, F., McMahon, D.J. *et al*. (2000). Parathyroid hormone as a therapy for idiopathic osteoporosis in men: effects on bone mineral density and bone markers. *J Clin Endocrinol Metab* **85:**3069–76.
28. Stewart, A.F., Cain, R.L., Burr, D.B. *et al*. (2000). Six-month daily administration of parathyroid hormone and parathyroid hormone-related protein peptides to adult ovariectomized rats markedly enhances bone mass and biomechanical properties: a comparison of human parathyroid hormone 1–34, parathyroid hormone-related protein 1–36, and SDZ-parathyroid hormone 893. *J Bone Miner Res* **15:**1517–25.
29. Hodsman, A.B., Kisiel, M., Adachi, J.D. *et al*. (2000). Histomorphometric evidence for increased bone turnover without change in cortical thickness or porosity after 2 years of cyclical hPTH(1–34) therapy in women with severe osteoporosis. *Bone* **27:**311–18.
30. Altman, D.G. (1996). Better reporting of randomised controlled trials: the CONSORT statement. *BMJ* **313:**570–1.
31. ICH Harmonized Tripartite Guideline E6 Good Clinical Practice: Consolidated Guidance. http://www.fda.gov/cder/guidance/959fnl.pdf accessed 13/11/06
32. Rice, P.L. and Ezzy, D. (1999). *Qualitative Research Methods: A Health Focus*. Oxford: Oxford University Press.
33. Holloway, W. and Jefferson, T. (2000). *Doing Qualitative Research Differently*. London: Sage Publications.

3

Ethical Considerations

Derek Pearson and Colin G. Miller

3.1. Introduction

Whenever a clinical trial is being designed, the ethical implications must be considered. All trials must fulfil the general guidance issued in the Declaration of Helsinki (Finland), Edinburgh (Scotland) Amendment, 2000.[1] This has been enshrined in the good clinical practice (GCP) guidelines produced by the International Committee for Harmonization (ICH) and adopted by the Food and Drug Administration (FDA) in the USA[2] and the Committee for Proprietary Medicinal Products (CPMP) within Europe.[3] These guidelines cover issues that researchers must consider, such as the following:

1. The anticipated benefit of the trial to the individual subject and society must outweigh the foreseeable risks and inconveniences
2. The protection of the trial subject, which should be the most important consideration
3. The responsibilities of the Institutional Review Board (IRB) and Independent Ethics Committee (IEC)
4. The responsibilities of the investigator and sponsor
5. The informed consent of the trial subjects
6. The study protocol and investigator's brochure and the essential documentation required to undertake a clinical trial.

In the treatment of osteoporosis, there are a few further nuances that must be considered and carefully evaluated above and beyond the usual considerations.

3.2. Who Pays the Bill?

Essentially, it is the physician's responsibility to ensure that his or her primary concern is for the individual subject. However, physicians or their employers are generally reimbursed by the sponsor on a "per subject recruited" basis. This obviously provides an immediate dichotomy of interests, particularly for the site

management organizations (SMOs). There is a potential for SMOs, in particular, to see individuals as income rather than subjects. Having said that, many trial subjects attending an SMO will gain access to a far more thorough clinical work-up and, therefore, evaluation than they would in the average clinic or healthcare system. Hence, the first ethical debate starts before the subjects walk into the out-patient clinic.

The prohibitive cost of developing new therapies for osteoporosis might prevent the pharmaceutical industry from spending resources in this area. In 2004, it was estimated that the average cost of drug development was more than US$1 billion. Therapies for osteoporosis tend to be more expensive to develop because the phase III programmes are more than several years in length. It is, therefore, difficult to ensure that a reasonable payback is achieved. The regulatory agencies and, particularly, the FDA have to carefully review their guidelines for osteoporosis drug development if these investments are to be reduced. Without a reduction in research and development costs, it is difficult to envisage a healthy pipeline of new molecular entities (NMEs) in development for this indication. However, to reduce the requirements for registration, it could be argued, would require lowering of safety standards. There is obviously an ethical debate here regarding the cost–benefit–risk ratios.

Much osteoporosis research is driven by the pharmaceutical industry. This means that there is little nontherapeutic research in the field and that most clinical trials are seen as a means of bringing an NME into licensed use. Studies are not large enough to model the reduction in fracture risk and the mortality and morbidity associated with osteoporosis effectively. There are only a small number of studies that have sufficient subject numbers to inform this debate. These include the Fracture Intervention Trial,[4] the Vertebral Efficacy with Risedronate Therapy study,[5] and the Multiple Outcomes of Raloxifene Evaluation.[6] A recent meta-analysis reviewed prospective cohort studies of baseline bone mineral density (BMD) measurements and subsequent follow-up for fracture. Of >1,000 articles that were reviewed, 229 studies fulfilled their initial requirements but only 17 studies were suitable for inclusion in the analysis.[7] This means that, despite 15 to 20 years of research and development in the field of osteoporosis treatment, effective, evidence-based protocols for the management of osteoporosis have yet to be promulgated in primary care.

3.3. Placebo or not Placebo?

In any new therapeutic field, the initial studies can be ethically placebo-controlled. Although a placebo might not be in the best interest of the individual subject, without a thorough investigative programme the potential benefit of an NME is also unknown and, therefore, in this situation subjects can be considered to be ethically treated on both placebo and active treatment. However, once a good accepted treatment becomes available for the routine subject, performing new

placebo-controlled studies becomes ethically questionable. The current version of the Declaration of Helsinki[1] states the following:

> The benefits, risks, burdens and effectiveness of a new method should be tested against those of the best current prophylactic, diagnostic, and therapeutic methods. This does not exclude the use of placebo, or no treatment, in studies where no proven prophylactic, diagnostic or therapeutic method exists.

A recent consensus conference suggested that placebo-controlled trials with fracture outcomes could only be used if there is a lack of consensus regarding whether approved treatments are better than placebo, there are no serious irreversible outcomes, subjects are recruited who have rejected the current treatments because of side effects, or subjects are refractory to current treatments.[8] Although it is clear that current therapies are only partially effective in reducing fractures,[8] there is significant morbidity and mortality associated with fracture to preclude placebo-controlled trials, particularly for female osteoporotic subjects in the so-called "western world". All future NMEs will require active-comparator studies. Some researchers believe that the combination of calcium and vitamin D supplements will suffice as a comparator, but the levels of calcium and vitamin D when studied alone have been higher than those used in the control arms of published clinical trials.[8] Other researchers believe that, in general, the combination of calcium and vitamin D is an unacceptable active treatment, particularly for early postmenopausal woman. There are a number of bisphosphonates available (depending on the country), several more bisphosphonates under development, calcitonin, parathyroid hormone, and selective oestrogen receptor modulators (SERMs), all suitable as active comparator treatments. Currently, the regulatory agency guidelines are do not give clear guidance on active control studies in the osteoporosis field. Evidence of fracture reduction is still required for a treatment indication license. This requires phenomenal numbers of subjects to show an NME has superiority over another active treatment. This scenario is also untenable from a cost perspective. It is also ethically questionable to have such large numbers of subjects taking an unproven therapy, when scientifically there might be other ways of providing information on drug safety and efficacy. It might be acceptable to demonstrate that an NME is as good as, rather than better than, existing therapies in modifying BMD because there could be other aspects of the treatment that give one product advantages over another product (e.g. cost, fewer side effects, or greater acceptability to the subject).

It could be argued that in less developed countries the individual subject is still in a better position to be enrolled in a placebo-controlled clinical trial because there might not be the infrastructure or healthcare capital to treat subjects with osteoporosis. However, this position raises the question of ethics regarding the acceptability of treating subjects in one country rather than another when it comes to clinical trials. We do not have a uniform world in terms of health policy and each country has its own view of health and its treatment. Therefore, it might be possible to perform trials ethically in one country and not another, although any sponsor has ethical and moral responsibilities in the planning of a trial, regardless of the country in which it is conducted. Sponsors are also driven by the need to

obtain FDA approval for any NME under development and will, therefore, carry out trials in countries covered by the ICH's GCP guidelines.

3.4. Randomization

The randomized, controlled trial (RCT) currently seems to be the "holy grail" of clinical trial design. It is said to eliminate bias on entry to the trial. However, nonrandomized trials are thought to demonstrate a greater treatment effect than RCTs. There is a growing debate that subject preference and motivation should be taken into account when designing clinical trials.[9–11] This is much more similar to real life because subjects get increasingly involved in decisions about their care. Many subjects, for example, do not want to take hormone-replacement therapy (HRT), despite its proven beneficial effect on bone, because of the increased risk of breast cancer and continued menstruation with some of the therapies. The gastrointestinal side effects of some bisphosphonates can deter subjects from entry into a trial. If subjects are given a choice between randomization to placebo, a bisphosphonate, and another active treatment, they might refuse to enter the trial rather than risk entry into the placebo arm or treatment they do not want. With most studies having to be active-comparator studies, it might be better to take into account the subject's preference. "A well designed non-randomised study is preferable to a small, poorly designed and exclusive RCT."[9] Results of literature surveys comparing nonrandomized and randomized studies have shown that the treatment effect is not necessarily larger in the nonrandomized studies nor significantly different from the differences between RCTs. Future ethical review might consider subject preference.

3.5. Who Can Take Part?

When designing a phase III RCT, researchers must ensure that the study population is representative of the disease population. This requires careful consideration of the inclusion and exclusion criteria. If these criteria are too strict, the result of the trial might not be generalizable. Recruitment methods often rely on a high level of literacy among subjects, which results in study participants being well-educated, middle class subjects who have easy access to health information through the press and other media. In reality, there is a significant health gap and the subject at high risk of osteoporosis with poor diet, low body mass, and a history of smoking and alcohol use will bypass information relating to the prevention and treatment of osteoporosis completely. This raises the ethical issue of the availability of new, effective treatments for osteoporosis to those excluded from clinical trials. New treatments are often costly and are not provided by local health services. This is a more significant issue in relation to cancer drugs, because often chemotherapy is available only to trial subjects.

Subjects are also excluded for administrative reasons, often on the grounds of ethnicity. Trial sponsors are often unwilling to fund the translation of subject information into other languages and, even when translators are available locally, use the argument that they are unsure that subjects have fully understood the information so cannot ensure properly informed consent has been given. This excludes ethnic groups who might be at risk of osteoporosis, and in areas where there are ethnic populations of significant size it is ethical to ensure that they are fully involved in the clinical programme. This could mean that although they might be excluded from specific trials, when the development programme is taken as a whole, care must be taken to ensure that they are included. An example of this kind of effect occurred recently in the development of an ultrasonometer in the USA. A development programme required by the FDA involved the collection and development of a "normal population" database. This is obviously gender-specific and race-specific. The initial trials were set up to recruit Caucasian women. At least one IRB/IEC initially questioned the protocol and demanded that it was racially diverse. After further discussion, the IRB/IEC accepted the protocol because they appreciated it was part of a larger clinical programme, in which other protocols would be developed for other ethnic groups. Translation of the subject consent form must be carefully undertaken to ensure that language differences do not affect the character of the subjects being recruited, and that they meet the original inclusion/exclusion criteria. Therefore, it is recommended that, for all translation work, each document is also translated back into the original language to ensure the same meaning and nuances remain.

Are the inclusion and exclusion criteria for a trial biased so that only those subjects who show greater potential to benefit are included? The pharmaceutical companies must use those individuals at very high risk of developing the disease for an RCT to be cost-effective. If the data are watered down too much with subjects who are at lower risk, more subjects must be recruited. Not only does this cost significantly more, but also the ethics of treating a larger number of subjects with an unproven NME must be considered. This is one of the reasons for phase IV clinical trials. The sponsors obtain the initial data for regulatory approval through phase III trials and then use phase IV studies to see whether the treatment effect of an NME is maintained in a wider population. Furthermore, the cost of phase IV studies is considerably less because fewer measurements and endpoints must be recorded.

Researchers should take care to review inclusion and exclusion criteria when attempting to generalize the research to a wider population. Multicentre trials, particularly those based in primary care, should ensure that the choice of centrer is representative of the population.

3.6. Trial Procedures

When conducting clinical trials in any therapeutic field, there are usually a battery of clinical examinations, tests, or assessments that must be performed. These

extend from extra physical examinations or drawing more blood than is routinely performed to having more images taken [e.g. ultrasound, X-ray, magnetic resonance imaging (MRI), and dual-energy X-ray absorptiometry (DXA)], and other minimal risk measurements (e.g. height measurements). This might involve the subject in extra visits to the hospital or clinic, resulting in additional expense and inconvenience. The IRB/IEC might expect the sponsor to cover travel expenses and some trials offer a small inconvenience allowance, which is payable at the end of the study or paid on a pro-rata basis if the subject drops out before the end of the study. This is not normally considered coercive if the amount paid is reasonable (around US$130/GB £80 for completing a major trial over a period of several years, for example).

Trial subjects should be made aware of the tests and investigations that are additional to normal treatment within the informed consent process. A flow chart outlining the trial procedures can be helpful, for example (Table 3.1). This should include a clear, layman's explanation of any procedures and risks involved. The IRB/IEC will make an assessment of the additional investigations from an ethical standpoint. This might include, for example, weighing the drawing of 20 ml of blood every 3 months for biochemical markers (which is generally acceptable) against the large number of blood samples that could be required in phase I pharmacokinetic studies, for which there might be concern at the volume of blood taken.

One major area of concern is the radiation dose given to subjects. For a trial in which vertebral deformity is an endpoint, lateral spine radiographs must be acquired. To obtain the baseline data adequately, the requirements might be that two antero-posterior (AP) spine radiographs must be acquired for vertebral identification, in addition to a minimum of two, or sometimes three, lateral spine radiographs. If any of these are acquired incorrectly or the subject moves during acquisition, they must be retaken. DXA measurements are also obtained, minimally postero-anterior (PA) spine and femur measurements, but forearm, lateral lumbar spine, and total body scans can also be requested. Furthermore, it has been recommended in some guidelines that, to aid precision, duplicate DXA measurements should be obtained. For some sites, quantitative computed tomography (QCT) can be used instead of DXA, which further increases the radiation dose. Table 3.2 outlines the potential radiation dose that a subject could receive. At baseline, the complete radiological assessment is not dissimilar to a routine clinical work-up a physician might request for a subject. The repeat measurements are more problematic ethically. Generally, because the radiation dose from DXA is very low, this is not a significant problem. However, repeat radiographs at the same visit are not acceptable within the current ethical guidelines, nor are routine lateral spine radiographs at 6-month intervals.

The additional radiation dose and associated risks must be explained to the subjects in simple terms. It is best to use the effective dose (ED; in mSv) rather than the entrance skin dose (ESD; in mGy) or organ dose (mGy) because the ED can be related to the additional risk that the subject is exposed to. Calculating the ED for DXA is not easy because of the high, nonstandard filtration in the beam and the high X-ray tube voltages, but estimates of the ED are available in the

Table 3.1. A typical study flow chart for the sample study used in this book.

Procedure	Screening	Baseline	6 months	1 year	18 months	2 years	30 months	3 years	42 months	4 years
Clinical assessment	X	X	X	X	X	X	X	X	X	X
Dietary questionnaire	X					X				X
Lumbar spine X-ray	X									
Bone density scan (DXA)	X			X		X		X		X
Ultrasound scan of heel (QUS)	X			X		X		X		X
Blood tests including:	X	X	X	X	X	X	X	X	X	X
FBC, ESR										
PTH										
Bone-specific alkaline phosphatase										
Osteocalcin										
Study medication	Throughout									
Adverse event monitoring	Throughout									

Table 3.2. The typical effective dose and estimated lifetime risk of fatal cancer from common radiographic and DXA examinations.

	Typical effective dose	Estimated lifetime risk of fatal cancer
Thoracic AP-spine radiograph	0.4 mSv	1 in 50,000
Lumbar AP-spine radiograph	0.7 mSv	1 in 29,000
Thoracic lateral spine radiograph	0.3 mSv	1 in 67,000
Lumbar lateral spine radiograph	0.3 mSv	1 in 67,000
PA spine DXA (pencil beam)	<1 μSv	1 in 20,000,000
PA spine DXA (fan beam)	2–70 μSv	1 in 10,000,000 to 1 in 285,000
Femur DXA (pencil beam)	<1 μSv	1 in 20,000,000
Femur DXA (fan beam)	1–60 μSv	1 in 20,000,000 to 1 in 330,000
Total body DXA	4–75 μSv	1 in 5,000,000 to 1 in 270,000
Forearm DXA	<1 μSv	1 in 20,000,000
QCT spine	30–250 μSv	1 in 660,000 to 1 in 80,000

literature.[12,13] The main radiation risk from X-ray or DXA investigations is of cancer induction. Table 3.2 also gives the lifetime risk of fatal cancer for subjects aged between 16 and 69 years, based on a risk coefficient of 5% Sv^{-1}.[14] To apply this to paediatric subjects, double the risk, and for geriatric subjects, divide by five. To put these risks into context, approximately one in three of the population will develop cancer in their lifetime and one in four of the population will die from cancer. For the newborn, it is better to use risk estimates for the foetus and consider childhood cancer risks and the risk of hereditary disease, for which the risk coefficients are 3.0% Sv^{-1} and 2.4% Sv^{-1}, respectively. The ED of a whole-body DXA using a fan-beam DXA (Hologic QDR 2000, Bedford MA, USA) is approximately 8 μSv. Thus, the risk of fatal childhood cancer is approximately 1 in 3 million compared with a natural incidence of childhood cancer of 1 case in 650 children, of which 50% of cases are fatal. The risk of a hereditary disease is similar (1 case in 5 million children) compared with a natural incidence of 1 case in 100 children or 1 case in 20 children if minor abnormalities are considered.

Three categories of radiation risk have been proposed,[15] with corresponding levels of benefit to society:

Category I. Trivial risk (<0.1 mSv), requiring minor benefit to society from the research.

Category II. Minor-to-intermediate risk (0.1–10 mSv), requiring intermediate-to-moderate benefit to society from the research. This has been subdivided into the following:

Category IIa. Minor risk (0.1–1 mSv).
Category IIb. Intermediate risk (1–10 mSv).

Category III. Moderate risk (>10 mSv), requiring substantial benefit to society.

The majority of DXA machines give an ED that falls into the lowest category of risk, except for one of the newer fan-beam machines. If subjects receive additional X-rays of the lumbar spine and repeat DXA measurements, however, most

osteoporosis trials fall into category II. The IRB/IEC must consider the benefit to society from the research in the light of additional radiation dose. If the study design is poor, with insufficient subject numbers or a poor research hypothesis, for example, the study should be rejected.

A further consideration is how to explain the radiation risk to trial subjects as part of informed consent. Many use the chest X-ray as a unit of radiation risk (ED, 10–20 μSv), but this is unhelpful because it is not an expression of the additional risk. Perhaps it is better to explain the risk in terms of the equivalent number of days or months of natural background radiation. This is, on average, 2.4 mSv/year or 7 μSv/day in the UK. Thus, a fan-beam DXA might be equivalent to 1 day of natural background, whereas a lumbar spine X-ray is equivalent to approximately 3 months of natural background. This, at least, associates the risk with something that subjects can understand.

Genetic testing of blood samples is becoming increasingly common. The IRB/IEC will review the study to ensure that safeguards are in place to protect the interests of the subject in this sensitive area. If genetic testing is part of a study, there should be a clear hypothesis for the test. It is not acceptable to take a blood sample and retain it indefinitely to test the sample for a whole range of genes as new tests become available. The subject should be made aware of the genetic test in the subject information sheet. They should be told if the blood sample is going to be stored for future use, and further informed consent should be obtained if a new genetic test becomes available. Will the subject be told the result of the genetic test? Generally, results of such tests should not be fed back to the subject because genetic testing in osteoporosis is speculative and nonspecific at present. If there is feedback, it is unlikely that any genetic counselling will be required, but investigators might want to consider how to approach other family members if the study is to be widened. Now that specific cell lines are the subject of patents, this is clearly a sensitive area. Trial subjects might need to be made aware of the transfer of samples to external organizations for genetic testing, particularly those that might develop patents commercially. Trial subjects should be made aware that they will not gain from the commercial development of cell lines or patents that result from genetic testing of their samples. A clear statement regarding the implications of genetic testing should be included on the consent form.

3.7. Ethical Review

The role of the IRB/IEC is to protect the subject and ensure the scientific integrity of the study. When reviewing a study, they will consider the following issues:

1. Has the trial a clear research question?
2. Is the trial designed so that it is capable of answering that research question?
3. Are arrangements in place to deal with the interference in the management of subject care?
4. Are there adequate arrangements in place for identifying and monitoring adverse events?

5. Is the subject information adequate and written in clear, nontechnical language? Is it coercive in any way?
6. Are the consent arrangements adequate? Do subjects have a "cooling off" period between receiving the subject information and being asked to consent?
7. Do the benefits of the research outweigh the risk to the subjects?
8. Are the financial arrangements ethical?

In the USA, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1974 was charged with the task of identifying the basic ethical principles concerning human subjects involved in clinical research.[16] The three basic principles that the commission published in the Belmont report are as follows:

Respect for persons. Individuals are capable of making informed choices about taking part in research. Researchers must acknowledge this and seek to protect individuals who might not be able to take informed decisions because of illness or other incapacity. The application of this is in ensuring trial subjects receive adequate information about the trial that enables them to make an informed choice of whether to take part (see Section 3.8).

Beneficence ("do no harm"). Not only should investigators respect the decisions of the individual, but there is also an obligation to protect them from harm and secure their well-being, that is maximize the benefit while minimizing the harm.

Justice. The selection of trial subjects should be fair and be representative of the population that is likely to benefit from the research. The burden of research should not fall on any one subject group more than others. In the field of osteoporosis research, there are often too many trials looking for too few subjects, and investigators should ensure that subjects are not recruited to multiple trials at the same time.

The FDA guidelines pertaining to ethical standards for research on human subjects are primarily derived from this report. However, an IRB/IEC has to weigh not only these ethical considerations, but also the competing principle of the social benefits derived from scientific research.

In a multicentre trial, the committee will also consider the suitability of the local investigator, the institution in which the trial is to be carried out, and the local research subjects. They will ensure that subjects are not recruited into multiple studies in centres at which many clinical trials are operating at the same time.

The IRB/IEC will usually have lay members on them. A clear, nontechnical lay summary of the project is vital. The lay members see their role as protecting the subjects' interests and will undertake review from the subjects' viewpoint.

3.8. Informed Consent

Clear subject information is a priority. Many studies are rejected by IRB/IECs simply because of poor subject information. It is wise to spend time producing a good subject information sheet. Take advice from those who are involved in the

Table 3.3. Informed consent.

Basic elements of informed consent:

1. A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures that are experimental.
2. A description of any reasonably foreseeable risks or discomforts to the subject, including the risks associated with the NME, the risk of not being on active treatment, and the risks of additional investigations, DXA and multiple X-rays.
3. A description of any benefits to the subject or to others which might reasonably be expected from the research. Therapeutic research might be of no direct benefit and might not reduce the subject's risk of fracture, but there could be societal benefit if the study adds to knowledge of the disease or treatment of the disease.
4. A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.
5. A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the US Food and Drug Administration, other regulatory authorities, or the trial sponsors might inspect the records.
6. For research involving more than minimal risk, explanations of whether any compensation and whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information can be obtained.
7. Explanations of whom to contact for answers to pertinent questions about the research and research subjects' rights, and in the event of a research-related injury to the subject.
8. A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject can discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.

Additional elements of informed consent:

9. A statement that the particular treatment or procedure might involve risks to the subject (or to the embryo or foetus, if the subject is or might become pregnant) that are currently unforeseeable.
10. Anticipated circumstances under which the subject's participation could be terminated by the investigator without regard to the subject's consent.
11. Any additional costs to the subject that might result from participation in the research.
12. The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject.
13. The approximate number of subjects involved in the study.

production of subject leaflets and do not assume that something that is obvious to you will be obvious to the subject. Guidance is available from the FDA (Table 3.3).[17] A sample subject information sheet is shown in Appendix 3.1 for the example trial used throughout this book. The sheet is based on the current guidelines issued by the Central Office of Research Ethics Committees (COREC) in the UK. It is broken into sections, rather than being a large block of text. Subjects should be invited to take part in the trial in the opening paragraph. The language should be nontechnical and the risks and benefits of participating in the trial should be clearly stated. Explain what will happen to the subject if they take part and give details of any investigations that might be involved in the trial. A picture of the equipment to be used can be helpful because many subjects have no idea what DXA equipment is like. The information should explain any financial arrangements,

including the fact that the local investigator might be receiving funds from the sponsor to carry out the trial. Subjects should be made aware of their rights, any compensation available to them if something goes wrong, and the fact that participation in the trial might affect their insurance policies, including health insurance. IRB/IEC approval of a study should be omitted from the subject information. A study has shown that subjects understand that the role of the IRB/IEC is to ensure that subjects come to no harm.[18] Informing them of IRB/IEC involvement in the review process might imply that the trial is safe and likely to be of benefit.

In addition to providing a subject information sheet specific to the trial, it can be useful to include some general information on osteoporosis, with contact information on local and national self-help groups. Depending on the study design, it might be useful to include dietary and exercise information to assist in fracture prevention, although this could compromise the study design in some trials. In our example study, subjects were given information on calcium supplementation, but were responsible for assessing their own dietary calcium intake and supplementation. It would be wrong to assume that the control and placebo groups used this information in the same way and it might be necessary to assess calcium intake at baseline and completion of the study.

Subjects were recruited to the example study through advertisements in primary care clinics and local radio, television, and newspapers. Advertisements should be reviewed by the IRB/IEC. In this case, subjects were invited to an initial information session at the local hospital, where a general talk on osteoporosis was given, in addition to a talk on the trial. Only then were subjects given the formal subject information; they were contacted at a later date, once they had been given time to consider their involvement in the trial. Such a complex recruiting process can be expensive and time-consuming and results in only limited[19] recruitment, but it ensures that subjects are well informed and raises awareness of the problem of osteoporosis in the local community.

It is important that subjects demonstrate that they fully understand the information they have received. The consent form is, therefore, important. An example of GCP is given in Appendix 3.2. Investigators should ensure subjects complete the form for themselves. It is unlikely in osteoporosis trials that subjects will be unable to consent, so this issue of relative or caregiver assent need not be considered.

3.9. Dissemination

The dissemination of trial findings is crucial to the ethical conduct of research and is fraught with the danger of bias. The IRB/IEC will want to know how the findings are to be disseminated as part of the review process. Researchers have a responsibility to ensure that the findings are available for publication in peer-reviewed journals, even if the findings are negative. This is one area of bias because negative studies are not often published or take longer to reach publication.[20] The use of overoptimistic language in the report, the ease of publication of reports from high profile clinicians and high-profile centres, and the abuse of the peer-review process

all add to the bias at publication. A well-structured report, in a peer-reviewed publication, is the only way to reduce the risks of such bias (see Chapter 9).

3.10. Summary

The ethics of performing clinical trials are complex and variable. Some of the further considerations in the field of osteoporosis have been provided and must be evaluated for each protocol. It is important for the trialist to consider this aspect of the trial before submission to the IRB/IEC.

References

1. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland, 1964, and revised by the 52nd World Medical Assembly, Edinburgh, 2000. http://www.wma.net/e/policy/b3.htm accessed 13/11/06.
2. Expert Working Group (Efficacy) (1997). International Conference on Harmonization: Guidelines for Good Clinical Practice. *Federal Register* **62:**25691–709. http:// www.fda.gov/cder/guidance/959fnl.pdf accessed 16/11/06.
3. The Committee for Proprietary Medicinal Products: Step 5 Note for Guidance on Good Clinical Practice (CPMP adopted July 1996): CPMP/ICH/135/95. http:// www.emea.eu.int/pdfs/human/ich/013595en.pdf accessed 16/11/06.
4. Cummins, S.R., Black, D.M., Thompson, D.F. *et al*. (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* **280:**2077–82.
5. Harris, S.T., Watts, N.B., Genant, H.K. *et al*. (1999). Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis. *JAMA* **282:**1344–52.
6. Ettinger, B., Black, D.M., Mitlak, B.H. *et al*. (1999). Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. *JAMA* **282:**637–45.
7. Marshall, D., Johnell, O. and Wedel, H. (1996). Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ* **312:**1254–9.
8. Brody, B.A., Dickey, N., Ellenberg, S.S. *et al*. (2003). Is the use of placebo controls ethically permissible in clinical trials of agents intended to reduce fractures in osteoporosis? *J Bone Miner Res* **18:**1105–9.
9. Britton, A., McKee, M., Black, N. *et al*. (1998). Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* **2(13):**1–124.
10. Brevin, C.R. and Bradley, C. (1989). Patient preferences and randomised clinical trials. *BMJ* **299:**313–15.
11. Cooper, R.G., Grant, A.M. and Garratt, A.M. (1997). The impact of using a partially randomised patient preference design when evaluating alternative managements for heavy menstrual bleeding. *Br J Obstet Gynaecol* **104:**1367–73.
12. Njeh, C.F., Apple, K., Temperton, D.H. *et al*. (1996). Radiological assessment of a new bone densitometer the Lunar Expert. *Br J Radiol* **69:**335–40.
13. Njeh, C.F., Fuerst, T., Hans, D. *et al*. (1999). Radiation exposure in bone mineral density assessments. *Appl Radiat Isot* **50:**215–36.

14. ICRP Publication 60 (1991). Recommendations of the International Commission on Radiation Protection. *Ann ICRP* **21:**1–77.
15. Radiation Protection 99 (1998). *Guidance on medical exposures in medical and biomedical research*. Luxembourg: Office for Official Publications of the European Communities.
16. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979). Belmont Report: notice of report for public comment. *Federal Register* **44:**23191–7. http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.htm accessed 16/11/06
17. 21 Code of Federal Regulations 50.20 General requirements for informed consent. http://www.fda.gov/oc/ohrt/irbs/accessed 16/11/06
18. Kent, G. (1997). The views of members of Local Research Ethics Committees, researchers and members of the public towards the roles and functions of LRECs. *J Med Ethics* **23:**186–90.
19. Grainge, M.J., Coupland, C.A., Cliffe, S.J. *et al*. (1998). Cigarette smoking, alcohol and caffeine consumption, and bone mineral density in postmenopausal women. The Nottingham EPIC Study Group. *Osteoporos Int* **8:**355–63.
20. Jadad, A. (1998). *Randomised Controlled Trials*. London: BMJ Books.

Appendix 3.1

St Elsewhere's Hospital, Metabolic Bone Clinic

A randomized, double-blind, placebo-controlled trial of nasal salmon calcitonin in the prevention of bone loss in postmenopausal women.

Investigators: Dr Smith and Dr Jones

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives, and your primary care physician if you wish. Ask us if there is anything that is unclear or you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

What is the purpose of the study?

The purpose of the study is to test the effectiveness of nasal salmon calcitonin in preventing bone loss in postmenopausal women. The study will take 4 years to complete. During that time, you will be taking either nasal salmon calcitonin or a dummy drug (placebo) by nasal spray. All subjects will be offered calcium and vitamin D supplements. Eighty subjects are being recruited to the study.

Why have I been chosen?

This study has been advertized in clinics, on local radio and television, and through the press. As a result, you have attended one of our information sessions on osteoporosis. You have been asked to take part in the study because you are within 5 years of the menopause and have no other bone-related diseases.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time and without giving a reason. This will not affect the standard of care you receive.

What will happen to me if I take part?

This study will last for 4 years. At the start of the study, you will need to come to the Metabolic Bone Clinic at St Elsewhere's Hospital for a screening visit. At this visit, we will ask you a number of questions about your medical history, diet, smoking, and exercise and give you some more general information about osteoporosis, diet, and exercise. We will take 10 ml (two teaspoons) of blood to measure chemicals that are markers of bone metabolism. You will also have an X-ray of your spine and a special scan to measure the bone density of your lumbar spine and hip. This scan involves a small amount of X-rays (less than a conventional X-ray) and is called a dual-energy X-ray absorptiometry scan (DXA) scan. You will be asked to lie on a bed for your hip and spine scan. A picture of the

scanner is shown below. We will also use a special ultrasound machine to measure the bone density of your heel. This involves putting your foot into a water bath for a few minutes while a reading is taken. A picture of the ultrasound machine is also shown below.

The screening visit will take approximately 2 hours. At the end of the screening visit, it might be decided that you are not suitable to continue in the study because of your medical history. You will be paid travel expenses at the end of the visit.

If you are suitable to continue in the study, you will be asked to attend for a second visit within 1 month of the screening visit. The format of the visit will be similar to the screening visit, but no more X-rays will be taken. You will, however, be required to have further DXA and ultrasound scans.

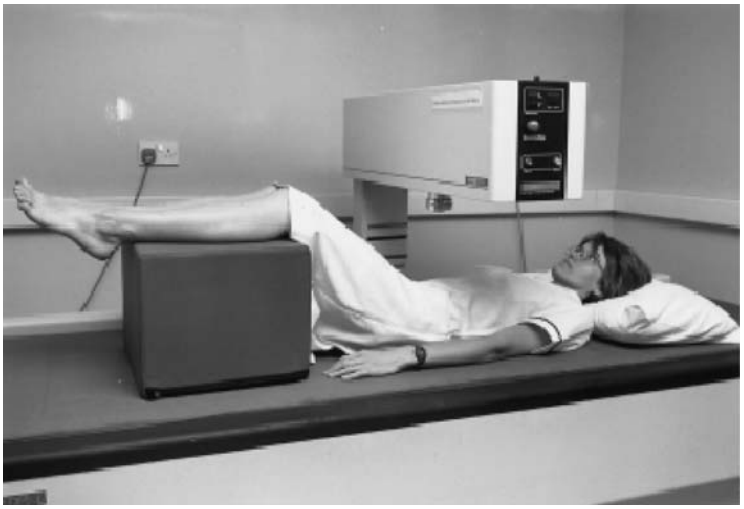
At that visit, you will be randomly assigned to a treatment group or a group of subjects who receive a dummy drug (a drug that looks exactly like the real thing but contains no active ingredient, called a "placebo"). Neither you nor your doctor will know which group you are in, but you have a one in two chance of receiving nasal salmon calcitonin. Your doctor will be able to find out which group you are in, if necessary.

The study lasts for 4 years, and you will need to come to clinic every 6 months for approximately 1 hour. Every year, you will be asked to have repeat DXA and ultrasound scans, answer questions about your diet and exercise, and have further blood samples taken.

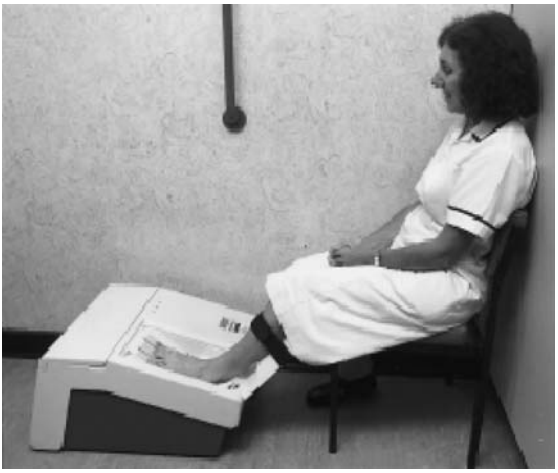
You will be given reasonable travel expenses for attending the clinic and an inconvenience allowance after completing the study. If you do not complete the study, the inconvenience allowance will be paid on a pro-rata basis.

What do I have to do?

It is important that you take the trial medication as directed by the doctor in clinic. You will be given instructions regarding how to take the medicine: it is given as a nasal spray. There are no other precautions that you have to take and you can take any other medication you normally use. You will be given general information about prevention of osteoporosis,



A dual energy X-ray absorptiometer.



Ultrasound of the heel.

including advice on diet and exercise, but it is up to you to decide whether or not to follow this advice.

What is the drug that is being tested?

The form of nasal salmon calcitonin being tested is called [tradename] and has been used for a number of years to help women control menopausal symptoms with serious side effects. This is the first time the effect on bone loss has been studied. The drug is taken as a nasal spray and you will be given the drug in a dated spray pack. You will be given a card (similar to a credit card) with details of the trial on it, which you should carry at all times.

What are the alternatives for treatment?

Other forms of treatment are available that are known to prevent bone loss at the spine or hip. You can discuss these with your primary care physician or the clinic doctor if you wish.

What are the side effects of taking part?

[Tradename] can, in a very small number of people, cause an allergic reaction. More common, but generally mild, side effects include a dry or blocked nose, sneezing, inflammation of the nasal lining, flushing, headache, diarrhoea, nausea, and flu-like symptoms. If you experience any of these, please report it to the doctor organizing the study.

What are the possible disadvantages and risks of taking part?

The disadvantages of taking part are that you have a one in two chance of not receiving the treatment drug when there are known treatments for preventing bone loss in the spine or hip already available. That is why we are only recruiting women who are at low risk of developing osteoporosis during the study and offering calcium and vitamin D supplements to all study participants. If, during the study, you lose bone rapidly, the doctor will withdraw you from the study, investigate the causes of your bone loss, and recommend appropriate treatment.

The X-rays and DXA scans both involve exposing you to additional radiation. The additional radiation dose (<1 mSv) is equivalent to less than 5 months of natural background radiation.

If you have private medical insurance, you should check with the company before agreeing to take part in the trial. You will need to do this to ensure that taking part will not affect your medical insurance.

What are the possible benefits of taking part?

Taking part in this trial might be of no direct benefit to you. The information we get from this study could help us to prevent osteoporosis in the future.

What if new information becomes available?

Sometimes, during the course of a research project, new information becomes available about the drug that is being studied. If this happens, your research doctor will tell you about it and discuss with you whether you want to continue in the study. If you decide to withdraw, your research doctor will make arrangements for your care to continue. If you decide to continue in the study, you will be asked to sign an updated consent form.

Also, on receiving new information, your research doctor might consider it to be in your best interests to withdraw you from the study. He or she will explain the reasons and arrange for your care to continue.

What happens when the research study stops?

At the end of the study, you will not be able to continue on the study drug. You will be able to discuss treatment options with your primary care physician or the research doctor. Occasionally, the company sponsoring the research might stop it. If this is the case, the reasons that the study has been stopped will be explained to you.

What if something goes wrong?

Compensation for any injury caused by taking part in this study will be in accordance with the guidelines of the Association of the British Pharmaceutical Industry (ABPI). [The sponsor], without legal commitment, will compensate you, without you having to prove that the sponsor is at fault if it is probable that such injury results from giving [tradename] or any other procedure carried out in accordance with the protocol for the study. [The sponsor] will not compensate you if such injury results from any procedure carried out that is not in accordance with the protocol for the study. Your right, in law, to claim compensation for injury if you can prove negligence is unaffected. You can also use the standard complaints mechanism and contact the St Elsewhere's Complaints Officer on [telephone number].

Will my taking part in this study be kept confidential?

If you consent to take part in the research, you will be identified by study number alone. Any of your medical records might be inspected by [the sponsor] for purposes of analysing the results. They might also be looked at by people from [the sponsor] and the regulatory authorities, to check that the study is being carried out correctly. Your name, however, will not be disclosed outside the hospital. Your primary care physician will be told that you are taking part in the study.

What will happen to the results of the research study?

The results of this study will be published in a journal and a copy of the results will be available on request after the study has closed.

Who is organizing and funding the research?

[The sponsor] are funding this study, which is being organized through the Metabolic Bone Clinic at St Elsewhere's Hospital.

Contact for further information

If you want to discuss this further, please contact Dr Smith on [contact details].
Thank you for taking the time to read this information.

Appendix 3.2

Consent Form

St Elsewhere's Hospital, Metabolic Bone Clinic

A randomized, double-blind, placebo-controlled trial of nasal salmon calcitonin in the prevention of bone loss in postmenopausal women.

Investigators: Dr Smith and Dr Jones

The subject should complete the whole of this sheet herself.

Please cross out as necessary

- Have you read and understood the subject information sheet? YES/NO
- Have you had opportunity to ask questions and discuss the study? YES/NO
- Have all the questions been answered satisfactorily? YES/NO
- Have you received enough information about the study? YES/NO
- Who have you spoken to? Dr/Mr/Mrs/Ms

Do you understand that you are free to withdraw from the study:

- at any time? YES/NO
- without having to give a reason? YES/NO
- without affecting your future medical care? YES/NO
- Do you agree to take part in the study? YES/NO

Signature (Subject): Date:
Name (in Block Capitals):
I have explained the study to the above subject and she has indicated her willingness to take part.

Signature (Doctor): Date:
Name (in Block Capitals):

4

Standardization and Pretrial Quality Control

Derek Pearson

4.1. Introduction

Before initiating a clinical trial, the sponsor must be assured that the equipment and techniques used can adequately answer the research question. In multicentre trials, each centre might have instruments from different manufacturers, compounding the issues involved in initiating the trial. The investigator must be assured of the following:

1. The instruments and anatomical site chosen are adequate to classify the osteoporotic status of subjects on entry to the trial and monitor the anticipated change in bone mineral density (BMD) or quantitative ultrasound (QUS)
2. Staff are adequately trained on the equipment to be used
3. Long-term precision is known within the subject group studied
4. Differences between instruments are known and, as necessary, a crosscalibration is derived
5. Subjects radiation doses are known.

The aim of this chapter is to summarize the equipment tests required before establishing a clinical trial to reassure the investigator that each centre is adequately prepared to begin clinical work. It will cover the following points:

1. Differences between dual energy X-ray absorptiomerty (DXA) instruments or ultrasonometers
2. The choice of phantom for crosscalibration
3. The review of daily quality control (QC) before the trial commences
4. Accuracy
5. Precision: both long-term and short-term precision, *in vitro* and *in vivo*
6. Crosscalibration and standardized BMD
7. The radiation dose.

The chapter will also consider the preparations for using X-ray morphometry (either radiographic or DXA morphometry). Only Lunar (GE Healthcare,

Madison WI, USA) and Hologic (Bedford MA, USA) densitometers will be considered because this covers 95% of the world market.

4.2. Equipment Differences: DXA

The calibration differences in BMD measured on different manufacturers' DXA equipment are known and documented[1–4], in addition to the intramanufacturer instrument differences.[5–8] The Lunar densitometers are calibrated against an ashed bone standard[9] and give measured results some 10–15% higher than those obtained for the same subject measured with either Hologic or Norland (Cooper Surgical, Trumbull, CT, USA) instruments, which are calibrated against a hydroxyapatite standard.[10] Figure 4.1 demonstrates the calibration difference between Lunar and Hologic devices *in vivo* on a Bland and Altman plot. The difference in BMD measured on the two instruments is plotted against the mean BMD, to demonstrate any systematic differences between the instruments. The mean difference between the two systems was measured as 0.12 g/cm^2. This implies that, when establishing a multicentre clinical trial, investigators tend to choose centres with equipment from the same manufacturer. However, if studies exceed 10 to 12 instruments this is not necessary because crosscalibrations can be obtained that are adequate to compare groups of subjects at baseline and demonstrate a treatment effect within the subject group. It should be noted, however, that

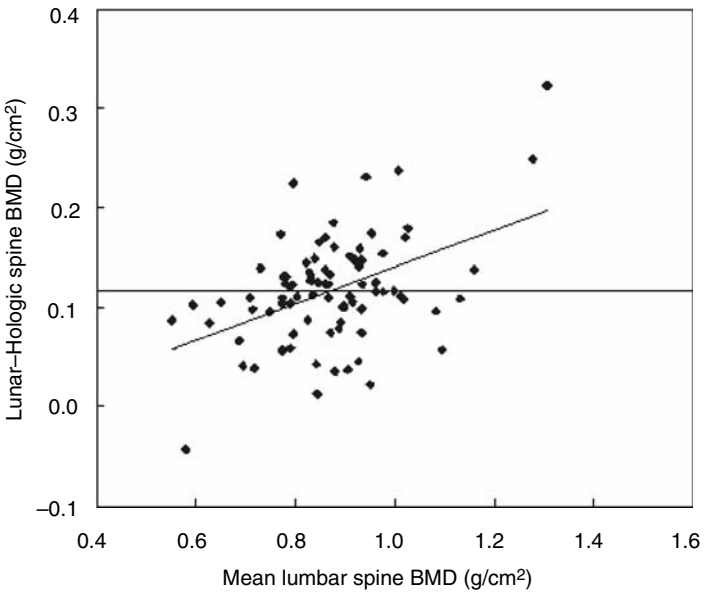


Figure 4.1. Bland and Altman plot of Lunar DPX-L and Hologic QDR 2000 lumbar spine BMD *in vivo*. The mean difference is 0.12 g/cm^2. There is a significant regression of the difference in BMD on the mean BMD ($r = 0.46$; $P < 0.0001$).

the crosscalibration is not good enough to enable clinicians to transfer individual subjects from one instrument to another, even of the same manufacturer, owing to the standard error of estimate of the crosscalibration. At 10 instruments or less, the error and cost of having instruments from a number of manufacturers is probably not worth the gain of recruiting additional centres to the trial. The differences between manufacturers are large, but instruments from the same manufacturer can still be up to a few percentage figures different.[11,12] Investigators will be keen to avoid cohort effects from individual centres within the trial and will still want to crosscalibrate instruments from the same manufacturer at different centres.

Comparison of BMD in the lumbar spine is the easiest because there are a range of phantoms that can be used to crosscalibrate. Femoral BMD presents more of a problem, because there are no good anthropomorphic phantoms for crosscalibrating (the limitations of the Hologic hip phantom are discussed later in this chapter) and the regions of interest (ROIs) vary between manufacturers. The Hologic QDR software sets the neck box at a right-angle to the midline of the neck, tangential to the trochanter, whereas the Lunar DPX algorithm identifies the narrowest part of the neck. Both Lunar and Hologic instruments determine the position of Ward's triangle by searching for the site of minimum BMD, but the ROI is a different size. Figure 4.2 shows the different ROIs superimposed. The difference in position of the ROI and poorer precision of BMD in the hip mean that the correlation coefficient between manufacturers is worse than in the lumbar spine (correlation coefficients between 0.876 and 0.953 for

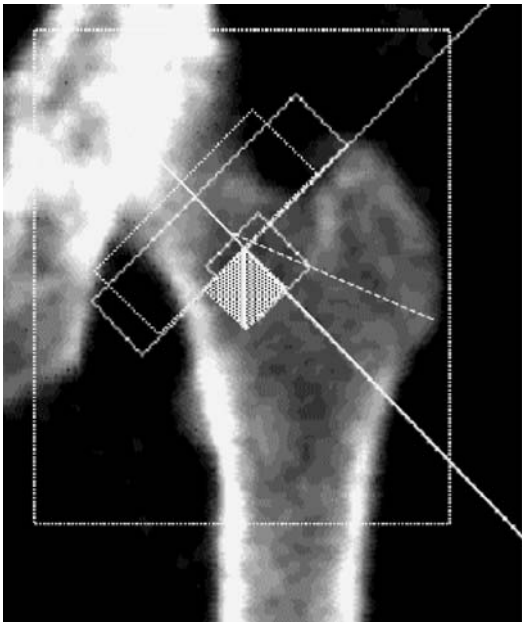


Figure 4.2. The Lunar and Hologic hip regions superimposed. The Hologic Ward's triangle area is the shaded area and the Hologic femoral neck box shown with a broken line. The Lunar regions are shown with an unbroken line.

the hip compared with 0.99 for the lumbar spine[4,13]), with a standard error of the estimate for the regression that is almost doubled (0.04 g/cm^2 to 0.079 g/cm^2 for the femur compared with 0.03 g/cm^2 for the lumbar spine[4,13]). This means that, in clinical trials that use femoral BMD as a surrogate endpoint for hip-fracture risk, investigators would be best advised to use equipment from the same manufacturer, although they would still need to compare results between centres.

4.3. Equipment Differences: Ultrasound

A phrase has been coined in bone densitometry circles: "Ultrasound is not ultrasound is not ultrasound." It is important that investigators keep this in mind when considering the use of ultrasonometers in clinical trials. Anatomical sites differ (e.g. cortical versus trabecular), ultrasound frequencies used differ, techniques differ (e.g. water bath versus contact and imaging versus nonimaging), and variables differ (e.g. heel velocity, time of flight, and bone velocity). In a phantom study with four different instruments, broadband ultrasound attenuation (BUA) varied between 61 dBMHz^{-1} and 116 dBMHz^{-1},[14] and the speed of sound (SOS) varied between 1418 ms^{-1} and 1686 ms^{-1}. *In vivo*, the mean difference between Lunar Achilles Plus (GE Healthcare, Madison WI, USA) Plus and McCue Cuba Clinical II (Cooper Surgical, Trumbull, CT, USA) instruments is approximately 30 dBMHz^{-1} for BUA and 90 ms^{-1} for SOS.[15] It is possible to crosscalibrate *in vivo* and phantoms are now becoming available[16–19] to crosscalibrate *in vitro* to a limited extent, because not all phantoms are suitable for every type of equipment. Investigators are advised to use equipment from one manufacturer, but be aware of differences between equipment from the same manufacturer. For example, an osteoporotic phantom measured on five different versions of the Lunar Achilles device gave BUA results ranging from 98 dBMHz^{-1} to 109 dBMHz^{-1},[19] with a coefficient of variation of about 4%, differences that the authors concluded were potentially clinically significant. An upgrade from a Lunar Achilles device to an Achilles Plus device gave a mean difference in stiffness of 2.7%.[16] (Stiffness is a linear combination of BUA and SOS reported by Lunar ultrasonometers. It has been derived to correlate to BMD better than BUA or SOS alone and is scaled to a young normal value of 100.) This is similar to the interinstrument variation observed with DXA instruments from the same manufacturer. Investigators should ensure that crosscalibration is carried out to avoid differences between centres.

4.4. DXA Phantoms

A number of lumbar spine phantoms are available:

1. The Hologic anthropomorphic lumbar spine phantom (Figure 4.3a) consists of four moulded lumbar vertebrae, each containing similar densities of calcium hydroxyapatite, with no differentiation of cortical or trabecular bone. The vertebrae are embedded in a block of epoxy resin, a pseudo-soft-tissue-equivalent material

that is supraphysiological and, therefore, substandard for crosscalibration purposes. Each phantom is provided with a specified (factory-calibrated) mean BMD value. It is easily measured on a daily basis by most QDR instruments, and is probably the most commonly used phantom in clinical trials. It has been designed primarily for use on Hologic instruments and does not have vertebrae above and below those being evaluated, which is a requirement for Lunar systems, for which some operator intervention is required to analyse the phantom results. The major drawback for this phantom is that it does not have a range of BMD values, so it is inadequate for cross-calibration purposes, although the addition of an aluminium mask to provide a range of BMD values has been used.[20]

2. The Lunar aluminum spine phantom (Figure 4.3b) is a rectangular aluminium bar, representing the first to fourth lumbar vertebrae (L1 to L4). The base of the 12th thoracic vertebra (T12) and the top of the fifth vertebra (L5) are also mimicked in the phantom. Different vertebral densities are provided by varying the thickness of the aluminium in a series of steps. The area of each vertebra is also varied. Each step has known values of bone mineral content (BMC), area, and BMD. The spine phantom is placed in a 15-cm deep water bath to simulate soft tissue, making it less easy to use. However, it has been available encased in a soft-tissue mimic of epoxy resin since 1998. The phantom is scanned from the midpoint of L5 up to the midpoint of T12. The mean BMD of the second to fourth lumbar vertebrae (L2 to L4) is precalibrated by the manufacturer, as 1.256 ± 0.025 g/cm^2. The BMDs of individual vertebrae between L1 to L4 are 0.92 g/cm^2, 1.076 g/cm^2, 1.239 g/cm^2, and 1.403 g/cm^2, respectively. The range of BMD values does not include values as low as those found in the clinical setting. It is straight-edged, providing no assessment of edge-detection algorithms, and can have poor precision on some fan-beam instruments owing to the partial volume effect.

3. The European spine phantom ESP (QRM GmbH, Möhrendorf, Germany) (Figure 4.3c) was developed in response to a need for a universal standard for crosscalibrating DXA and Quantitative Computed Tomography (QCT) systems from different manufacturers. The ESP consists of three geometric, pseudoanthropomorphic vertebrae, which are made of bone-equivalent plastics and calcium hydroxyapatite. Each vertebra has a different BMC and subsequent standardized BMD of 500 mg/cm^2, 1000 mg/cm^2, and 1500 mg/cm^2, respectively.[12] The vertebrae are embedded in a tissue-equivalent plastic, which is moulded into an oval phantom with flattened sides and dimensions of 28 cm by 18 cm. It is compact and easy to measure. It combines most of the good properties of both Lunar and Hologic spine phantoms, being pseudoanthropomorphic, and contains a range of densities. The ESP was designed to provide a test of the instruments' edge-detection algorithms, which neither the Lunar nor the Hologic spine phantom provides. It is difficult to obtain a very precise measurement of L2 because of the low BMD and high attenuation of the soft tissue, which becomes a problem when performing crosscalibration. The ESP is also the most expensive of all the phantoms commercially available.

4. The Bona Fide phantom (BFP; Bio-Imaging Technologies Inc, Newtown, PA, USA) shown in Figure 4.3d is a calcium hydroxyapatite step wedge embedded

in acrylic, which provides a soft-tissue equivalent of 26% fat (the normal physiological range is approximately 6–40%). In comparison, the Hologic phantom gives a nominal 60% fat. It offers a range of densities between 0.7 g/cm^2 and 1.5 g/cm^2. The shapes of the vertebrae are designed to mimic those found in real life and test edge-detection algorithms better than the Lunar spine phantom. It is easy to measure and comes in its own carrying bag, which remains *in situ* for scanning. Similar to the ESP, it is also not manufacturer-specific, which makes it ideal for crosscalibration purposes.

The Hologic spine phantom is adequate for a daily check of consistency, but it is not adequate for crosscalibration purposes because only one value of BMD can be

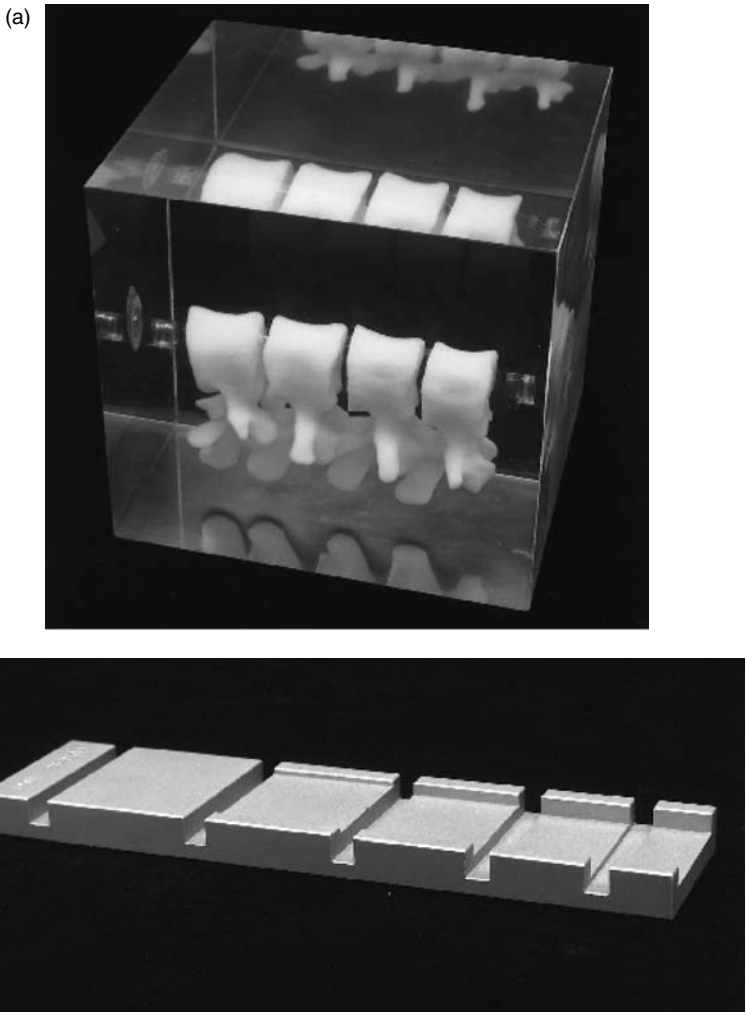
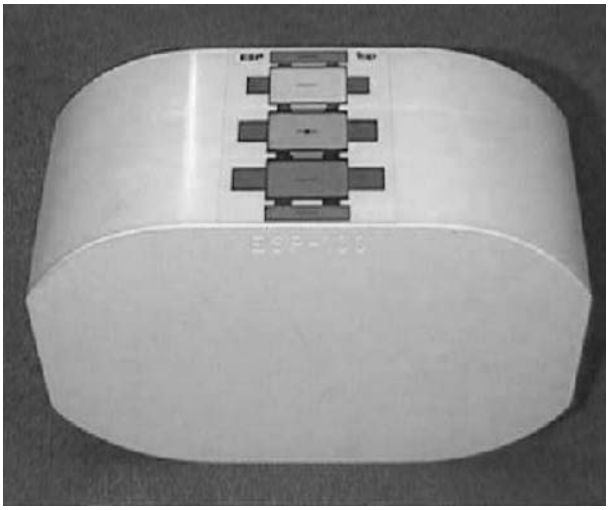


Figure 4.3. (Continued)

(c)



(d)

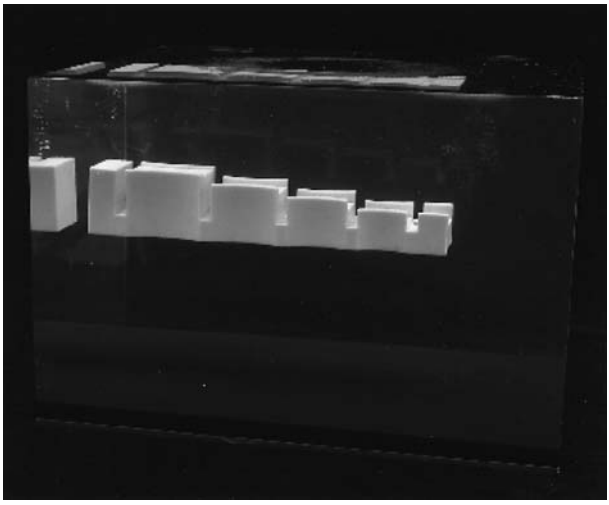


Figure 4.3. (a) Hologic spine phantom. (b) Lunar aluminium spine phantom. (c) European spine phantom. (d) Bona Fide phantom.

measured. Although a good, consistent calibration with the Hologic spine phantom can be obtained, calibration changes can be found when checked against a linearity phantom. Figure 4.4 shows the results from a Lunar DPX-L instrument for a period of time during which the Hologic phantom gave a stable result, but there was a significant fall at L2 and the third lumbar vertebra (L3) using the Lunar aluminium spine phantom. The ESP tends to diverge from an *in-vivo* crosscalibration at high values of BMD.[20] It is adequate to provide a crosscalibration at BMD values between 0.5 g/cm^2 and 1.0 g/cm^2. Its role in standardized BMD is discussed in the section on

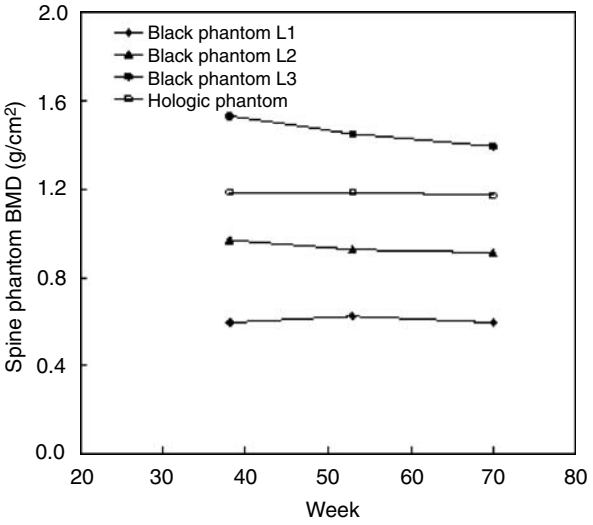


Figure 4.4. Linearity data for a Lunar DPX-L densitometer. The results from the Hologic phantom show no change, but there is a significant fall in BMD at L2 and L3 of the Lunar aluminum spine phantom.

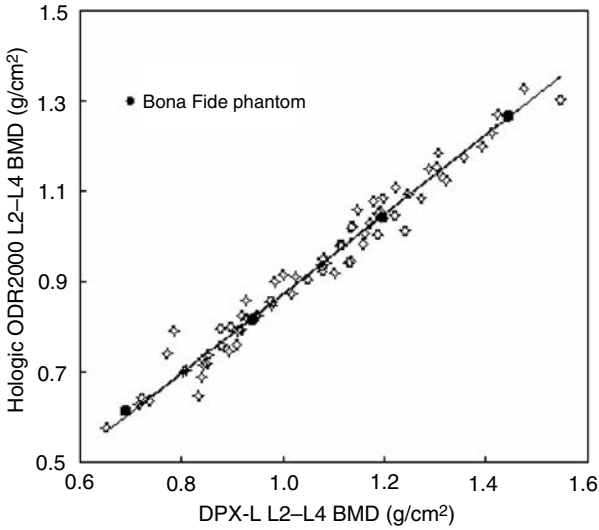


Figure 4.5. *In-vivo* crosscalibration data showing the regression line and Bona Fide phantom vertebrae.

standardization later in this chapter. The Lunar phantom will also provide an adequate crosscalibration, but lacks low values of BMD and is not made of calcium hydroxyapatite, which the US Food and Drug Administration (FDA) prefers. The BFP provides good crosscalibration data. Figure 4.5 shows an *in-vivo* regression, involving 73 subjects, measured on Lunar DPX-L and Hologic QDR 2000 devices. The BMD for each vertebra of the BFP is shown superimposed on the data and the subject

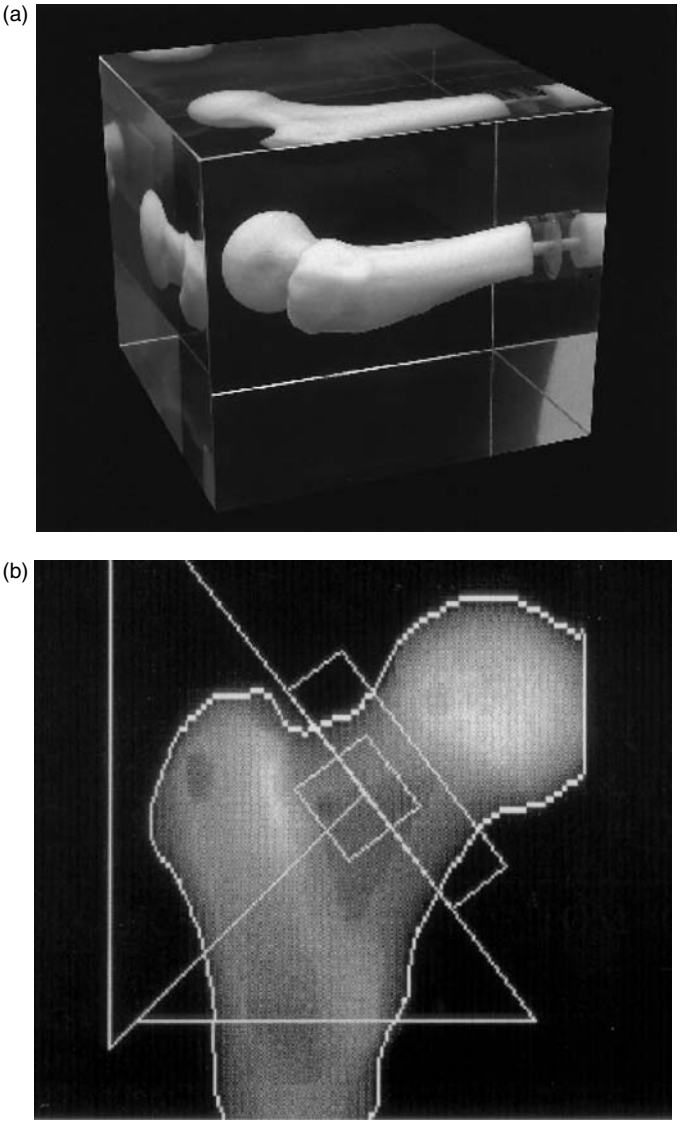


Figure 4.6. (a) Hologic hip phantom. (b) A scan of the Hologic hip phantom on the Lunar Expert DXA.

regression. It can be seen that the phantom results lie close to the subject regression for the whole range of data. In conclusion, both the ESP and Lunar phantom will provide an adequate crosscalibration, but the results using the BFP make this the prime candidate for crosscalibrations.

There is only one commercially available hip phantom, the Hologic hip phantom (Figure 4.6a). The problem with the phantom is that neither the QDR nor the DPX algorithms place the Ward's triangle ROI correctly or in the same place.

Both instruments identify areas of minimum BMD within the femoral neck, and this is not physiological. The gradient in BMD down the femoral neck does not mimic real life either; this is illustrated in an image of the phantom taken on a Lunar Expert (Figure 4.6b). The area of the Ward's triangle ROI on Lunar instruments is about three times larger than that found on the Hologic device. Because the length of the neck is short on the phantom, the position of the neck box is similar and the area is similar (5.5 cm^2 on the Hologic device compared with 5.3 cm^2 on the Lunar device). The phantom provides only a single calibration point because only one value of BMD is available. However, it depends on the question to be answered by the femur phantom. The operation of the DXA hardware can be adequately checked and crosscalibrated by a spine phantom; it is only the edge-detection algorithms that cannot be evaluated. There is no evidence to date to suggest that instrument failure occurs by edge-detection breakdown, because this is under software control and is a robust part of the instrument. All changes that have ever been documented have their basis in the hardware. This makes it suitable for monitoring consistency, but not suitable for crosscalibration purposes. This presents a problem when comparing instruments from different centres. A crosscalibration for lumbar spine cannot be applied. An *in-vivo* crosscalibration in one centre might give a reasonable crosscalibration between instruments from different manufacturers, to remove the gross differences in BMD at the hip. There is no adequate way at present, however, of taking into account interinstrument variation of equipment from the same manufacturer at different sites, apart from the single-point crosscalibration this phantom provides.

4.5. Ultrasound Phantoms

A restricted range of ultrasound phantoms is available, including the Leeds phantoms (University of Leeds, Leeds, UK) and the phantoms provided by manufacturers for daily quality-assurance (QA) checks.

4.5.1. The Leeds Phantoms

The Leeds phantoms are made of an epoxy material enclosed in a Perspex® cylinder. The epoxy material mimics the bone component and can be produced with different porosity values.[18] The pores are filled with vegetable oil and the ends of the tube are sealed. A "normal" and an "osteoporotic" phantom are available. The phantoms are limited by the fact that the seal is made of relatively thin plastic card and the phantoms cannot be used in dry contact systems without a special rig being constructed to ensure there is no pressure on the ends of the phantom. They are prone to leaking, and although they will be stable in the short term and can be used for a one-off crosscalibration, they are not suitable for measuring the long-term consistency of ultrasound instruments.

4.5.2. Phantoms Supplied by Manufacturers

All ultrasound phantoms suffer from two technical challenges: they are, generally, temperature-sensitive and the results they provide will also be temperature-sensitive, and they are hygroscopic to some degree. Both these factors further challenge the development of a useful phantom, regardless of the instrument type. Temperature inertia contributes to the site-to-site variation in the measured BUA and velocity of sound (VOS) or SOS of ultrasound phantoms. The trabecular structure also affects *in-vivo* ultrasound results, but this is difficult to mimic in a phantom.

4.6. Pretrial QA

Investigators need assurance that the equipment they are going to use is consistent. No trial should be started with brand-new equipment. A QA history for the equipment of at least 3 months is necessary to demonstrate long-term performance stability. This is usually in the form of the QC tests recommended by the manufacturer. However, a daily test of one of the phantoms described above is also helpful in judging the performance of the instrument. Monitoring data for equipment, as described in Chapter 8, is also helpful, because this will assure the investigator that the staff and equipment are capable of supporting the trial.

QA centres should look at the following aspects:

1. Systematic changes in performance from baseline, either as a step change or gradual trend
2. A baseline established over a period of a number of days rather than a baseline established using multiple measurements on 1 day
3. Increased scatter about the baseline compared with other similar equipment
4. Do results change from site to site when moving portable equipment?
5. Does temperature have an effect?
6. Daily phantom measurements on Lunar DXA equipment (other than the black QC phantom)—the manufacturer's QA only provides "PASS" or "FAIL" against a number of tests and provides primary instrument calibration.

4.7. Accuracy

Subjects are often entered into a trial because of their BMD, as measured at a screening visit. This will depend on the inclusion and exclusion criteria for the trial, which might require a certain proportion of women to match the World Health Organization (WHO) criteria for osteoporosis and osteopaenia. In classification, it is accuracy that is important, not precision. This raises a number of issues, such as the following:

1. Is the equipment accurate?
2. Why is classification being used within the study design? Is it necessary?

3. The use of normal ranges. Are the normal ranges used appropriate to the population to be studied?
4. Are the population databases, and hence cut-off, congruent across all sites?
5. Are subjects being classified for osteoporotic status or fracture risk?

The accuracy of equipment could be compared if there was an agreed standard with which to calibrate the equipment. The differences between equipment calibrations and the results obtained are well documented and there is no such standard. Standardization is discussed later in this chapter, but in this context accuracy must be considered in terms of diagnostic accuracy rather than the absolute accuracy of a BMD or QUS measurement. The debate is thus centreed on the use of normal ranges in classification.

Classification is used within study design for a number of reasons:

1. The study only involves subjects with a particular degree of risk, for example "normal" subjects or ostoporotic subjects
2. The study is to be stratified according to BMD or QUS results, to ensure adequate subject numbers within each stratum and thereby demonstrate a significant treatment effect
3. The study involves decision-making according to the BMD or QUS results; for example, the study might assign subjects to different treatments depending on the measurement or the role the measurement has in deciding the care pathway for the subject might be the purpose of the study itself.

Classification can be easily justified because of the indications for using a particular drug. However, care must be taken not to bias the study population and extrapolate the results to a more general population if the drug moves from the research stage into wider clinical practice.

Investigators must classify on the basis of a measurement that is logical in terms of the study design rather than the convenience of the technology available. For example, it would be inappropriate to use ultrasonometry of the heel to classify subjects as osteoporotic in a study for which vertebral fracture was the endpoint. When monitoring treatment response, the classification must use the site and technology that will be used throughout the study.

The WHO criteria for defining osteoporosis and osteopaenia were based on T-scores at the spine, hip, or midradius. The example given within the WHO report[21] is derived from forearm data, although it implies that this methodology can be transferred to other sites. It quotes a study of white women aged 50 years or over[22] in which 32% of women had a lumbar spine BMD $>$ 2 standard deviations (SDs) below the young normal mean, 29% of women had low BMD of the femur, and 26% of women had low BMD in the midradius. However, 45% of the women had low bone mass at the spine, hip, or midradius using this cut-off. Most clinical trials use inclusion criteria based on BMD at the spine or femur. It is clear, however, that a cut-off applied at one anatomical site will identify a different subject group to those identified using a different anatomical site. This is even more evident when using peripheral measurements to attempt to classify osteoporosis in

the central skeleton, because the sensitivity of QUS for diagnosing spinal osteoporosis might be as poor as 40% in some situations. The T-scores for different ultrasound instruments vary differently with age because the age-related bone loss at the heel is different from that at the lumbar spine or hip. Investigators must beware if attempting to relate a QUS T-score to a DXA T-score because of the poor agreement in classification for osteoporosis between QUS and DXA.[23] For example, the midpoint of the normal range of estimated BMD using the Hologic Sahara does not cross the $T = -2.5$ threshold before the age of 95 years compared with 74 years for lumbar spine BMD.[24]

Normal ranges vary between different ethnic groups, although it seems that white American and European normal ranges are interchangeable.[25–28] Appropriate normal ranges for an ethnic group must be used and investigators should be aware of the source of normal data. Often, normal ranges are derived from "hospital normals" and are not derived from random population samples, and differences have been reported when random populations are selected.[29] This is particularly true of the National Health and Nutrition Examination Survey (NHANES) study and the femoral BMD normal range using Hologic instruments.[30,31] However, this has now been corrected on the later versions of the Hologic software.

The SD of the normal range has a major effect on the classification of subjects. In the majority of cases, manufacturers have chosen a SD that is constant across the whole age band. A small difference in the SD is magnified when determining the T-score <-2.5 cut-off and can lead to the differences in classification discussed above. The differences might result from the sample used to derive the normal range not being large enough, arbitrary decisions by manufacturers in setting the SD, the sample not reflecting the normal population, and statistical manipulation of the data.

4.8. Precision

Good precision is often considered the holy grail of equipment manufacturers marketing DXA or QUS equipment. Better than 1% performance is offered with DXA. Although good precision is important in clinical trials, one should be wary of manufacturers' claims. Better than 1% precision is possible *in vivo*, in the short term, with relatively young, fit subjects. Phantom results in the long term of better than 0.5% are easy to obtain. The investigator must know the long-term, in-vivo precision in the study population because this will affect the study design. What it is possible to achieve in the short term with those young, fit subjects cannot be maintained over a period of 4 years in elderly osteoporotic subjects. As a rule of thumb, the SD of repeated measurements of BMD will remain the same in absolute terms at approximately 0.01 g/cm^2, whatever the mean BMD of the population studied. Thus, in young normal subjects the short-term precision will be approximately 1% because the mean BMD is approximately 1 g/cm^2, whereas in elderly osteoporotic subjects the mean BMD will be approximately 0.7 g/cm^2,

giving a precision of approximately 1.4%. Investigators should ensure they know how the precision was measured, because different methods will give different answers. Were multiple measurements carried out on subjects or only two? Who was in the subject group? Were enough subjects included for the results to be reliable? *In vitro*, as few as 10 samples might be adequate, but *in vivo* a minimum of 30 subjects should be used if only two measurements are to be obtained.[32] Were the measurements made and/or analysed by the same observer? Precision quoted in the literature should be reviewed with care.

Precision is usually measured as the coefficient of variation (CV%), which is defined as follows:

$$\text{CV\%} = \frac{\text{SD of repeated measurements}}{\text{mean}} \times 100\%. \tag{4.1}$$

Although this is simple to calculate in phantoms in the short or long term, it is often not possible or ethical to carry out more than two repeated measurements on subjects or normal volunteers. To measure short-term precision *in vivo*, at least 30 subjects should be measured twice, ideally on separate days. If the measurements are made on the same day, the subject should be asked to move away from the equipment and begin again. Many centres use the same observer to make the measurements, in order to improve precision, but this is unlikely to represent the variation in real life, when different operators might measure subjects. Study sponsors, however, often make quite stringent demands that one or, at the most, two operators only are used per site. There is usually excellent compliance with this request. When two measurements are made for each subject the CV% is defined as follows:

$$\text{CV\%} = \frac{\sqrt{\frac{\sum(a-b)^2}{2n}} \times 100\%}{(\bar{a}+\bar{b})/2}, \tag{4.2}$$

where a and b are the two measurements on each of the n subjects and a and b are the mean values of the first and second measurements.

When combining CV% from a number of individuals, a better estimate of the CV% can be obtained by calculating the root mean squared (RMS) SD and dividing by the average of the mean BMD for each individual involved in the study. The RMS SD and mean BMD ($\bar{a}$) are calculated as follows:

$$\text{RMS SD} = \sqrt{\frac{\sum_i \sigma_i^2}{n}} \tag{4.3}$$

$$\bar{a} = \frac{\sum_i \bar{a}_i}{n}, \tag{4.4}$$

where a_i is the BMD for the ith subject, σ_i is the variance of repeated measurements in that individual, and n the number of subjects. Thus, the CV% becomes the following:

$$\text{CV\%} = \frac{\text{RMS SD} \times 100\%}{\overline{a}}. \tag{4.5}$$

Although the CV% is useful in comparing precision between different DXA systems in which the variation between subjects is similar, it does not enable an easy comparison between different technologies because the variation between subjects is widely different. For example, SOS measured on the Lunar Achilles Plus has a CV% of 0.3%. This would seem to make it an ideal measurement for monitoring changes in bone, giving much better precision than DXA in the spine, for which the CV% is approximately 1%. However, the dynamic range of the measurement must be considered. For SOS, this is relatively small. This cancels out the effect of good absolute precision and means that it is not as good for monitoring change as DXA.

The standardized precision (SCV%) has been developed as a means of expressing the precision as a function of the dynamic range of the instrument. It has been defined as follows:[32,33]

$$\text{SCV\%} = \frac{\text{SD of repeated measurements}}{\text{Clinical range}} \times 100\%. \tag{4.6}$$

Another expression to overcome the problem is the so-called "annualized precision" (ACV%):

$$\text{ACV\%} = \frac{\text{SD of repeated measurements}}{\text{rate of change per annum}} \times 100\%. \tag{4.7}$$

The problem comes in the definition of the clinical range or rate of change per annum in any measurement. The original use of SCV% was to enable comparison of different equipment using a population that was measured on all the instruments. The definition then took the 5% to 95% of that range,[33] which was the optimum for that type of study. Subsequently, some have defined the clinical range as the SD of the normal range,[34] whereas others have defined it as the clinical range within the subject population measured, with many variations in between. Where authors have quoted SCV%, if they have not given the absolute value of the clinical range or how it was derived, the figures are meaningless for trying to make a comparison between centres. There is a similar problem with those authors who use the rate of change per annum. There is no agreement on what this is. Be wary when trying to interpret results.

Having issued this health warning, the SCV% is a useful quantity that takes into account the between-subject variation, in addition to the within-subject

variation. The young normal mean and range of results for a variety of different techniques is given in Table 4.1.[35] The clinical range is taken as 6 SD of the young normal mean and gives a measure of the variation between subjects. The CV% and SCV% have also been calculated. BUA and stiffness for the Lunar Achilles, for example, have worse CV% than SOS. As stated previously, however, the range of SOS values, and the small change in SOS with treatment, means that the SCV% is comparable with BUA, but worse than stiffness. This would confirm stiffness as the measurement of choice for monitoring change in ultrasound and the one agreed by the FDA. The SCV% makes comparison between instruments from different manufacturers easier. Taking BUA as an example: for CV%, BUA measured on the Lunar Achilles Plus device seems better than the Cuba Clinical II device. The mean BUA on the Achilles Plus device is higher, with a proportionately smaller clinical range, than that on the Cuba Clinical II device. If that is taken into account, the SCV% shows that the performance of the two instruments is much closer.

Often, good short-term precision is used as a basis for proceeding with a clinical trial that might last many years. Because open extensions are common, subjects are followed for up to 10 years. There can be many changes to the equipment within this time and it is important to know the long-term precision *in vivo*. The long-term precision can be measured by fitting a linear regression to the results from an individual subject measured over a number of years (Figure 4.7). The variation is defined as the standard error of estimate (SEE) of the regression[10,34] and the CV% is calculated as follows:

$$\text{CV\%} = \frac{\text{SEE} \times 100\%}{\text{Mean}}. \tag{4.8}$$

To combine the CV% from a number of individuals, the RMS average SEE can be used, as in Equation 4.5 above. Table 4.2[37] gives the long-term precision of DXA

Table 4.1. CV% and SCV% for three ultrasonometry instruments: the Lunar Achilles Plus, McCue Cuba II, and Hologic QDR2000 DXA.

	CV%	SCV%	Clinical range	Typical young normal mean
Achilles Plus				
BUA	1.60	5.10	39.1 dBMHz^{-1}	125 dBMHz^{-1}
SOS	0.33	5.10	100 ms^{-1}	1560 ms^{-1}
Stiffness	2.40	4.10	49.6	100
Cuba II				
BUA	3.30	3.30	74 dBMHz^{-1}	90 dBMHz^{-1}
VOS	0.50	6.80	114 ms^{-1}	1800 ms^{-1}
DXA (Hologic QDR 2000)				
AP Spine	1.40	2.00	0.66 g/cm^2	1.047 g/cm^2
Total hip	1.00	1.52	0.72 g/cm^2	0.975 g/cm^2

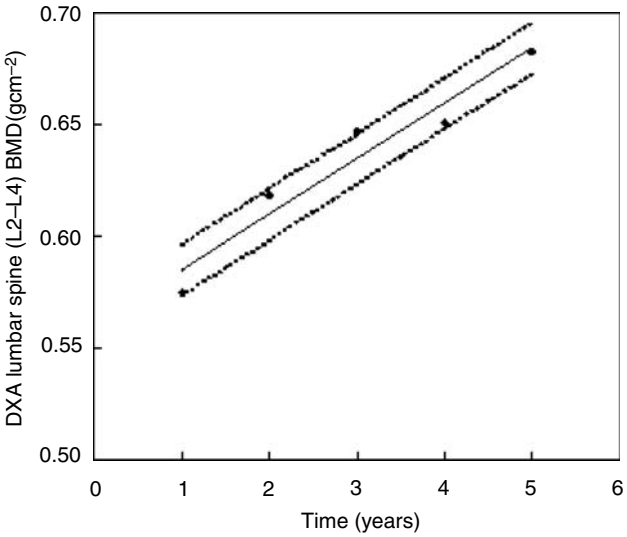


Figure 4.7. Regression through 4-year subject data to determine long-term precision. The broken lines represent the regression ±1 SEE about the regression.

(Hologic QDR 2000) and QUS (Lunar Achilles Plus). A standardized long-term CV% can also be calculated because it can be seen from Table 4.2 that SOS again seems to be the measurement of choice for monitoring long-term change in skeletal status. Here, it is more appropriate to use an ACV%, because change with time is important in monitoring follow-up. The ACV% and average rate of change measured in a control group are also given in Table 4.2. DXA of the lumbar spine is now superior to QUS for monitoring long-term changes.

Good precision is important, but this does not mean that a technique that does not have the best precision cannot be used to demonstrate a treatment effect within a clinical trial. There is a great deal of confusion between the precision required to monitor changes in BMD or QUS in an individual and that required to demonstrate a treatment effect within a group of subjects. Within an individual, a change in BMD or QUS is only considered significantly different to the baseline value if the change is greater than three times the long-term CV%. Thus, for BMD of the spine, a 3% change would be significant if the long-term CV% was 1%, whereas a change of 7.5% might be required before the change in a BUA result with a long-term CV% of 2.5% is regarded as significant. This affects the monitoring period, depending on the rate of change of BMD within an individual and the treatment they receive. If BMD and BUA change at a rate of 2% per year in the example above, the BMD measurement will detect a change after 18 months, whereas follow-up of almost 4 years will be required to demonstrate significant progress using QUS.

In a clinical trial, change is being monitored in a group of subjects, and the main source of variation is the variation in BMD or QUS results between individuals.

Table 4.2. Long-term CV% and annualized CV% for the Hologic QDR 2000 DXA and Lunar Achilles Plus QUS. The average rate of change in BMD and QUS results in a control group is also given.[36]

	Region/measure	Long-term CV%	ACV%	Average rate of change
DXA	Lumbar spine	1.8	1.1	0.015 g/cm^2
	Total hip	1.5	1.9	0.007 g/cm^2
QUS	BUA	2.8	2.1	1.7 $dBMHz^{-1}$
	SOS	0.7	3.5	3.1 ms^{-1}
	Stiffness	4.7	2.5	1.6

The CV% is of less importance. If the overall variation is 10% when the CV% is 1%, doubling the CV% to 2% increases the overall variation to 10.1%. Thus, a treatment effect can be demonstrated sooner than that in an individual. For example, in a placebo-controlled study of hormone-replacement therapy (HRT),[36] the long-term precision of BUA was 2.8%. It would require a change in BUA of 8.4% before it would be deemed significant in an individual. The rate of change in the control group was, on average, 1.5% per year. More than 5 years would have to elapse before a significant difference could be demonstrated. A significant treatment effect, however, was demonstrated at 2 years, with a mean difference between the treatment and control group of $<6\%$.

4.9. Crosscalibration

In a multicentre trial, it is clearly impossible (not to say unethical) to derive an *in-vivo* crosscalibration and, therefore, a phantom crosscalibration must be used instead. The phantoms available have been described earlier in this chapter. It is important that the phantom chosen contains a range of BMD or QUS values so that a multipoint calibration can be used. Although they do not completely mimic real life, the slope of the phantom regression is usually not significantly different from an *in-vivo* regression[37] and will provide an adequate crosscalibration to cope with cohort effects from different centres, even when instruments from different manufacturers are being used.

In such a crosscalibration, the phantom should be circulated to the centres involved and measured at least 10 times on each instrument. Often, this takes place on a single day when a representative from the sponsor is present. It is better to measure the phantom over a number of days, if possible, to take into account the day-to-day variation in the equipment. Once measurements have been made on all instruments, the calibration can be made in one of the two following ways:

1. One instrument can be chosen to be the "master" and all other instruments calibrated to that standard. There is a probability, however, that the "master" could change its own calibration during the study period, when further crosscalibrations might be required.
2. All instruments are calibrated to the known BMD of the phantom.

A third option is that the equipment at different centres is close enough not to require a crosscalibration to be applied.[38]

To calculate the crosscalibration, a regression analysis is usually used. The problem with regression is that the model assumes that there is no random error in the independent variable. This results in an underestimate of the slope of the true linear relationship between two instruments. For example, Figure 4.8 shows an *in-vivo* crosscalibration. The regression of y on x and x on y are both shown. The true, linear relationship lies somewhere between the two and can be determined in a number of ways. The line that bisects the angle between the two regressions can be calculated,[39] a principal-components analysis can be used,[40] or the line fitted by eye between the two can be used.[41] This is more important when an *in-vivo* crosscalibration is being used. For a phantom, the results lie close to the regression line and the difference between the regression of y on x and x on y is so small as not to matter.

DXA has been considered in this section, but the principles apply to crosscalibration for any type of equipment. With the growth in peripheral DXA, peripheral QCT (pQCT), and QUS, the comparisons become much harder and sponsors are recommended to use one model of equipment. This is because measurement sites vary, ROIs vary, and there are fundamental differences in the way manufacturers measure what seems to be the same quantity. This is highlighted in QUS by the differences in SOS, VOS, time of flight velocities, heel velocity, and bone velocity and the difference in BUA of up to a factor of 2 if measured at the same site over what is, apparently, the same frequency range.

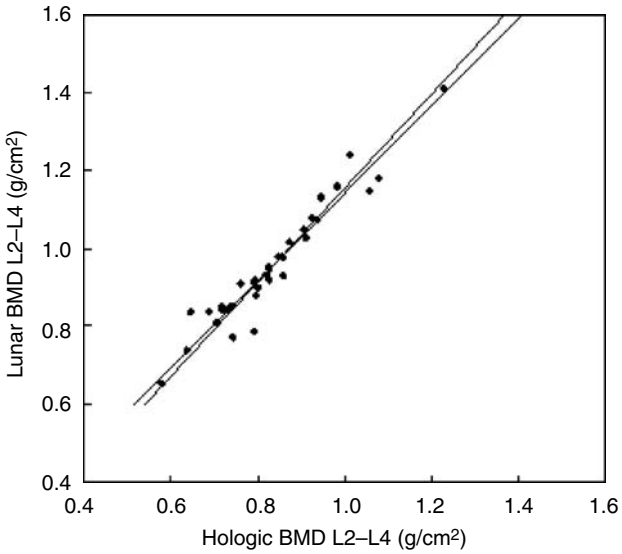


Figure 4.8. *In-vivo* cross-calibration showing regression of both x on y and y on x.

4.10. Standardization

DXA instruments from the three main manufacturers now report standardized BMD (sBMD) for both spine and hip. This was developed on behalf of the International DXA Standardization Committee (IDSC).[42] The basis of this study was the measurement of lumbar spine and hip BMD in 100 women on three different instruments: the Hologic QDR 2000, Lunar DPX-L, and Norland XR26 MkII. The ESP was also measured on each instrument. Using the three linear regressions from each pair of instruments, it is mathematically possible to derive a set of equations that enable the sBMD to be calculated from the BMD measured on each instrument. For the sake of convention, L2 to L4 was the region chosen for the postero-anterior PA spine and total hip was the region chosen for the femur. The equations for the PA spine are as follows:

$$\text{sBMD} = 1000(\text{BMD}_{\text{Hologic}} \times 1.0755)$$
$$\text{sBMD} = 1000(\text{BMD}_{\text{GE}} \times 0.9522)$$
$$\text{sBMD} = 1000(\text{BMD}_{\text{Norland}} \times 1.0761).$$

These, and the equations for the femur, have been adopted as an international standard by the IDSC[43–45] and are used by manufacturers within their reports. The units of milligrammes per square centimetre are used for sBMD so that the results are not confused with BMD. The equations have been tested prospectively in 56 subjects at one centre.[39] Prospectively, the root mean squared (RMS) difference in sBMD measured on the Hologic QDR 2000 and Lunar DPX-L devices was 30.34 mg/cm^2. The RMS difference of approximately 3.5% was not thought to be clinically significant by the authors proposing standardization.

The problems with these equations are as follows:

1. They have been derived on a small subject group on one set of equipment and do not take into account the interinstrument variation. The RMS differences give users a feel for the size of the effect of interinstrument variation (3.5%). This is too large to transfer an individual subject from instrument to instrument and is similar to the size of the interinstrument variation reported elsewhere.[38]

2. The mathematical technique used to derive the equations has nonunique solutions. The ESP has been used to "peg" the equations to a known BMD. The phantom was used because it contains known amounts of hydroxyapatite and the midvertebra was chosen because "it alone shows a very good proximity to all three human regression lines." This weakness must be accepted, and any bone standard that lay on or close to the line of human regressions could have been used to provide a useable solution to the equations.

3. The regression model has been used. This means that the slope of the relationship between any pair of instruments has been underestimated, as outlined above. The regressions in the original model were also forced through the origin, even when there were significant nonzero intercepts. This has been corrected by

the authors of the original paper,[39] but these corrected equations have not been adopted by the IDSC.

Despite this, if phantom checks carried out in the pretrial phase show that there is little difference between the equipment at different centres, sBMD is probably adequate for comparing groups of subjects within a clinical trial in which the overriding aim is to demonstrate a treatment effect.

4.11. Radiation Dose

There is probably enough information in the literature and from the manufacturers for it to be unnecessary for investigators to measure the radiation dose from existing DXA, pDXA, and pQCT equipment. Table 3.2 gives a summary of entrance skin dose (ESD) and effective dose (ED) for current techniques. If a new application of this technology is proposed, investigators should consider measuring ESD and estimating the ED. The ED is the weighted sum of absorbed dose to each irradiated organ or tissue[46] and enables a comparison of radiation risk between different X-ray examinations to be made. Estimating the ED is difficult because of the high level of filtration used in DXA instruments. The best method is to measure ESD and then the variation of dose with depth in a water or anthropomorphic phantom,[47,48] although there are methods to calculate the ED directly from ESD.[49] Ionization chambers or thermoluminescent dosimeters can be used, but, because dose rates are very low, multiple scans might need to be made to get an accurate measure of ESD or depth dose. Once the ESD and depth dose are known, the dose at the centre of each irradiated organ can be calculated and used to estimate the ED. On one visit, subjects might have a number of regions scanned (lumbar spine and femur, for example). The total ED per visit will be the summation of the ED for each region.

4.12. Morphometry

There are two common methods of morphometric analysis of the spine:

Radiographic vertebral morphometry. Measurements are made on standard antero-posterios (AP) and lateral radiographs of the thoracic and lumbar spine using a digitizing tablet connected to a computer, and the anterior, posterior, and midvertebral heights are measured.[50] Alternatively, heights can be measured directly from film[51] or a qualitative index of vertebral deformity can be calculated.[53] From the measurements, indices of vertebral deformity can be calculated.[52–54]

Morphometric absorptiometry (MXA). A lateral image of the spine is obtained using DXA instruments.

In both techniques, six points are placed on each vertebral body[55], as shown in Figure 7.16. There are a number of issues to consider before establishing a clinical trial:

- Choice of technique. A choice must be made between radiographic morphometry and MXA. It is not possible to mix the techniques between centres.
- Consistent, high-quality radiographs must be obtained that are suitable for identifying the points on the vertebra. A procedure for taking the radiographs must be agreed before the study commences that ensures the following:
  - A common film focus distance (100 cm)
  - The spine is parallel to the long axis of the X-ray table
  - Lumbar views are centred on L3 and include T12 to the sacrum
  - Thoracic views are centred on the seventh thoracic vertebra (T7) and include the second thoracic vertebra (T2) to T12
  - T12 is included on all views
  - Lateral views ensure subjects can maintain a comfortable lateral position with knees, shoulders, and elbows flexed to 90°, with appropriate support
  - Lateral views ensure the long axis of the spine is parallel to the table
  - Lateral thoracic spine views are obtained with the subject breathing during a long exposure to blur overlying rib and lung details and give better visualization of the vertebra
- MXA positioning. Subject positioning is important and the manufacturer's instructions on positioning should be carefully followed (Chapter 7). Figure 7.9 shows a typical MXA scan.
- Choice of normal range. Most manufacturers provide a normal range that is derived from radiographic morphometry rather than MXA.[55] This does not take into account the magnification inherent in a radiograph, which results in differences of between 2 mm and 16 mm in anterior vertebral height between the two techniques. MXA reference ranges must be developed or vertebral fracture will be overdiagnosed using MXA with radiographic-derived reference ranges.
- How important are the upper thoracic vertebrae? MXA does not image these vertebrae well compared with radiographic morphometry.
- MXA is not as sensitive as radiographic morphometry in identifying vertebral fractures because of poor image quality.[55]
- What is the definition of vertebral fracture? There is some debate on this issue in the literature. A >15% decrease in vertebral height has been used to define fracture.[56] However, this has been show to overdiagnose vertebral fracture.[51–54] Others have used the SD of the vertebral height to define fracture[51,53] and concluded that the definition of fracture should depend on the level specific SD,[51] necessitating the creation of a normal range derived from a large sample size. An alternative has been developed that requires two criteria to be fulfilled at each vertebral site to define fracture,[54] minimizing false-positive results.

Staff training and the agreement of a protocol for point placement are important in MXA and should be agreed between centres before establishing the trial.

4.13. Summary

In summary, there are a significant number of technical issues that must be considered before establishing a clinical trial that are not trivial. It is always worthwhile taking time to carry out pretrial QC and come to a conclusion on these issues before starting the study. It is too late to sort out issues of standardization and crosscalibration once the trial is established.

References

1. Mazess, R.B., Trempe, J.A., Bisek, J.P. *et al*. (1991). Calibration of dual-energy X-ray absorptiometry (DEXA) for bone density. *J Bone Miner Res* **6:**799–806.
2. Laskey, M.A., Flaxman, M.E., Barber, R.W. *et al*. (1991). Comparative performance *in vitro* and *in vivo* of Lunar DPX and Hologic QDR-1000 dual energy X-ray absorptiometers. *Br J Radiol* **64:**1023–9.
3. Laskey, M.A., Crisp, A.J., Cole, T.J. *et al*. (1992). Comparison of the effect of different reference data on Lunar DPX and Hologic QDR-1000 dual-energy X-ray absorptiometers. *Br J Radiol* **65:**1124–9.
4. Genant, H.K., Grampp, S., Gluer, C.C. *et al*. (1994). Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res* **9:**1503–14.
5. Economos, C.D., Nelson, M.E., Fiatarone, M.A. *et al*. (1996). A multicenter comparison of dual-energy X-ray absorptiometers: *in vivo* and *in vitro* measurements of bone mineral content and density. *J Bone Miner Res* **11:**275–85.
6. Orwoll, E.S. and Oviatt, S.K. (1991). Longitudinal precision of dual-energy X-ray absorptiometry in a multicenter study. The Nafarelin/Bone Study Group. *J Bone Miner Res* **6:**191–7.
7. Khan, K.M., Henzell, S.L., Broderick, C. *et al*. (1997). Instrument performance in bone density testing at five Australian centres. *Aust N Z J Med* **27:**526–30.
8. Kolta, S., Ravaud, P., Fechtenbaum, J. *et al*. (1999). Accuracy and precision of 62 bone densitometers using a European Spine Phantom. *Osteoporos Int* **10:**14–19.
9. Mazess, R.B., Sorensen, J.A., Hanson, J.A. *et al*. (1989). Dual-energy X-ray absorptiometry. In: Ring, E.F.J., Evans, W.D. and Dixon, A.S. (eds). *Osteoporosis and Bone Mineral Measurement*, p. 62–71. York: IPSM.
10. Blake, G.M., Wahner, H.W. and Fogelman, I. (1999). *The Evaluation of Osteoporosis*, 2nd edn. London: Martin Dunitz.
11. Lees, B., Garland, S.W., Walton, C. *et al*. (1997). Evaluation of the European Spine Phantom in a multi-centre clinical trial. *Osteoporos Int* **7:**570–4.
12. Pearson, J., Dequeker, J., Henley, M. *et al*. (1995). European semi-anthropomorphic spine phantom for the calibration of bone densitometers: assessment of precision, stability and accuracy. The European Quantitation of Osteoporosis Study Group. *Osteoporos Int* **5:**174–84.
13. Laskey, M.A., Crisp, A.J., Cole, T.J. *et al*. (1992). Comparison of the effect of different reference data on Lunar DPX and Hologic QDR-1000 dual energy X-ray absorptiometers. *Br J Radiol* **65:**1124–9.
14. Strelitzki, R. and Truscott, J.G. (1998). An evaluation of the reproducibility and responsiveness of four "state-of-the-art" ultrasonic heel bone measurement systems using phantoms. *Osteoporos Int* **8:**104–9.

15. Horton, B.R., Pearson, D. and Masud, T. (1998). A comparison of ultrasound measurements using the Achilles Plus and Cuba II. In: *Current Research in Osteoporosis and Bone Mineral Measurement V*, p. 69. London: British Institute of Radiology.
16. Pearson, D., Cawte, S.A., Horton, B. *et al*. (1997). Cross calibration of an Achilles Plus and Achilles ultrasound machine. *Osteoporos Int* **7:**302.
17. Diessel, E., Fuerst, T., Njeh, C.F. *et al*. (2000). Comparison of an imaging heel quantitative ultrasound device (DTU-one) with densitometric and ultrasonic measurements. *Br J Radiol* **73:**23–30.
18. Clarke, A.J., Evans. J.A., Truscott, J.G. *et al*. (1994). A phantom for quantitative ultrasound of trabecular bone. *Phys Med Biol* **39:**1677–87.
19. Strelitzki, R., Clarke, A.J., Truscott, J.G. *et al*. (1996). Ultrasonic measurement: an evaluation of three heel bone scanners compared with a bench-top system. *Osteoporos Int* **6:**471–9.
20. Tothill, P., Fenner, J.A. and Reid, D.M. (1995). Comparisons between three dual-energy X-ray absorptiometers used for measuring spine and femur. *Br J Radiol* **68:**621–9.
21. Kanis, J.A. (1994). Assessment of Fracture Risk and its Application to Screening for Postmenopausal Osteoporosis. Report of a WHO Study Group. Geneva: World Health Organization.
22. Melton, L.J., 3rd, Chrischilles, E.A., Cooper, C. *et al*. (1992). Perspective. How many women have osteoporosis? *J Bone Miner Res* **7:**1005–10.
23. Grampp, S., Genant, H.K., Mathur, A. *et al*. (1997). Comparisons of noninvasive bone mineral measurements in assessing age-related loss, fracture discrimination, and diagnostic classification. *J Bone Miner Res* **12:**697–711.
24. Faulkner, K.G., von Stetten, E., and Miller, P. (1999). Discordance in patient classification using T-scores. *J Clin Densitom* **2:**343–50.
25. Lehmann, R., Wapniarz, M., Randerath, O. *et al*. (1995). Dual-energy X-ray absorptiometry at the lumbar spine in German men and women: a cross-sectional study. *Calcif Tissue Int* **56:**350–4.
26. Oldroyd, B., Stewart, S., Truscott, J.G. *et al*. (1990). Bone mineral density in normal women living in the UK. In: Ring, E.F.J. (ed.) *Current Research in Osteoporosis and Bone Mineral Measurement*, p. 53. London: British Institute of Radiology.
27. Truscott, J.G., Oldroyd, B., Simpson, M. *et al*. (1993). Variation in lumbar spine and femoral neck bone mineral measured by dual energy X-ray absorption: a study of 329 normal women. *Br J Radiol* **66:**514–21.
28. Ryan, P.J., Spector, T.P., Blake, G.M. *et al*. (1993). A comparison of reference bone mineral density measurements derived from two sources: referred and population based. *Br J Radiol* **66:**1138–41.
29. Petley, G.W., Cotton, A.M., Murrills, A.J. *et al*. (1996). Reference ranges of bone mineral density for women in southern England: the impact of local data on the diagnosis of osteoporosis. *Br J Radiol* **69:**655–60.
30. Looker, A.C., Wahner, H.W., Dunn, W.L. *et al*. (1995). Proximal femur bone mineral levels of US adults. *Osteoporos Int* **5:**389–409.
31. Looker, A.C., Wahner, H.W., Dunn, W.L. *et al*. (1998). Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* **8:**468–89.
32. Gluer, C.C., Blake, G.M.K., Lu, Y. *et al*. (1995). Accurate assessment of precision errors: how to measure the reproducibility of bone densitometry techniques. *Osteoporos Int* **5:**262–70.

33. Miller, C.G., Herd, R.J.M., Ramalingam, T. *et al*. (1993). Ultrasonic velocity measurements through the calcaneus: which velocity should be measured? *Osteoporos Int* **3:**31–5.
34. Blake, G.M. and Fogelman, I. (1997). Technical principles of dual energy X-ray absorptiometry. *Semin Nucl Med* **17:**210–28.
35. Horton, B.R., Pearson, D. and Masud, T. (1998). A comparison of ultrasound measurements using the Achilles Plus and Cuba II. In: Ring, E.F.J. (ed.). *Current Research in Osteoporosis and Bone Mineral Measurement V*, p. 69. London: British Institute of Radiology.
36. Sahota, O., San, P., Cawte, S.A. *et al*. (2000). A comparison of the longitudinal changes in quantitative ultrasound with dual-energy X-ray absorptiometry: the four-year effects of hormone replacement therapy. *Osteoporos Int* **11:**52–8.
37. Cawte, S.A., Pearson, D., Green, D.J. et al. (1999). Cross-calibration, precision and patient dose measurements in preparation for clinical trials using dual energy X-ray absorptiometry of the lumbar spine. *Br J Radiol* **72:**354–62.
38. Faulkner, K.G. and McClung, M.R. (1995). Quality control of DXA instruments in multicenter trials. *Osteoporos Int* **5:**218–27.
39. Hui, S.L., Gao, S., Zhou, X.H. *et al*. (1997). Universal standardization of bone density measurements: a method with optimal properties for calibration among several instruments. *J Bone Miner Res* **12:**1463–70.
40. Feldmann, U., Schneider, B., Klinkers, H. *et al*. (1981). A multivariate approach for the biometric comparison of analytical methods in clinical chemistry. *J Clin Chem Clin Biochem* **19:**121–37.
41. Armitage, P. and Berry, G. (1971) *Statistical Methods in Medical Research*, 3rd edn. Oxford: Blackwell Science.
42. Genant, H.K., Grampp, S., Gluer, C.C. *et al*. (1994). Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res* **9:**1503–14.
43. Genant, H.K. (1995). Universal standardization for dual X-ray absorptiometry: patient and phantom cross-calibration results. *J Bone Miner Res* **10:**997–8.
44. Steiger, P. (1995). Standardization of measurements for assessing BMD by DXA. *Calcif Tissue Int* **57:**469.
45. Hanson, J. (1997). Standardization of proximal femur BMD measurements. International Committee for Standards in Bone Measurement. *Osteoporos Int* **7:**500–1.
46. IRCP Publication 60 (1991). *Recommendations of the International Commission on Radiological Protection*. *Ann IRCP* **21:**11–21.
47. Njeh, C.F., Samat, S.B., Nightingale, A. *et al*. (1997). Radiation dose and *in vitro* precision in paediatric bone mineral density measurement using dual X-ray absorptiometry. *Br J Radiol* **70:**719–27.
48. Njeh, C.F., Fuerst, T., Hans, D. *et al*. (1999). Radiation exposure in bone mineral density assessment. *Appl Radiat Isot* **50:**215–36.
49. Huda, W. and Morin, R.L. (1996). Patient doses in bone mineral densitometry. *Br J Radiol* **69:**422–5.
50. Minne, H.W., Leidig, G., Wuster, C. *et al*. (1988). A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. *Bone Miner* **3:**335–49.
51. Eastell, R., Cedel, S., Wahner, H.W. *et al*. (1991). Classification of vertebral fractures. *J Bone Miner Res* **6:**207–15.
52. Melton, L.J., 3rd, Kan, S.H., Frye, M.A. *et al*. (1989). Epidemiology of vertebral fractures in women. *Am J Epidemiol* **129:**1000–11.

53. Black, D.M., Cummings, S.R., Stone, K. *et al*. (1991). A new approach to defining normal vertebral dimensions. *J Bone Miner Res* **6:**883–92.
54. McCloskey, E.V., Spector, T.D., Eyres, K.S. *et al*. (1993). The assessment of vertebral deformity: A method for use in population studies and clinical trials. *Osteoporos Int* **3:**138–47.
55. Cawte, S.A. (1999). *Morphometric X-ray Absorptiometry using the Hologic QDR 2000*, MPhil Thesis. Nottingham: University of Nottingham.
56. Barnett, E. and Nordin, B.E.C. (1961). Radiological assessment of bone density. *Br J Radiol* **34:**683–91.

Section 2

Performing the Trial

5

Organization of the Clinical Trial by the Sponsor

Colin G. Miller

5.1. Introduction

There is a standard set of start-up, ongoing monitoring, and close-out requirements and procedures for all clinical trials. These generic requirements are discussed in other books and are not within the remit of this book. However, because of the nature of this therapeutic area, it is important to consider the particular extras and details that have to be evaluated, and this chapter provides an overview of these items. The sponsor must be aware of these items in a timely manner—all too often the novice will overlook an important detail until it becomes a crucial issue. This chapter will help ensure these kinds of errors are avoided.

5.2. Regulatory Guidance

Regulatory approval is the ultimate goal within the pharmaceutical industry. Therefore, knowledge of the rules and guidelines is crucial before embarking on any clinical programme. We also live in a global market, so with the high cost of development, any new molecular entity (NME) must address most of the world's markets in one programme. No longer can companies afford to run studies to appeal to single countries. This chapter will not address the generic good clinical practice (GCP) issues, *per se*, but the ones related to the field of osteoporosis.

At the time of writing, within the USA the regulatory field is changing significantly in the area of osteoporosis. Phase III studies are now being conducted that strictly do not fulfil the requirements laid out in the US Food and Drug Administration (FDA) guideline document (1993).[1] The FDA guidelines were written in response to the development primarily of antiresorption compounds, such as the bisphosphonates. They do not allow for the truly anabolic compounds such as parathyroid hormone (PTH), whereas the European [Committee for Proprietary Medicinal Products (CPMP)] and World Health Organization (WHO) guidelines,[2,3] which were written later, do allow for this scenario a little more

adequately. Similar guidelines are also available in Japan. It is not just government agencies that have designed the so-called "regulatory environment". The Group for the Respect of Ethics and Excellence in Science (GREES) have also published guidelines.[4] The GREES do not have any true regulatory power but have presented at numerous meetings and have had significant had an influence on the way the regulatory agencies evaluate compounds in the osteoporosis arena. This section will provide a brief comparison of the salient clinical points from each of the guidelines, which are obviously open to interpretation by each of the agencies in Europe.

Table 5.1 shows the expectations of each group for the phase II clinical programme. It can be seen that the CPMP expectations are dramatically different from the rest, requiring 2-year phase II data. This works if phase III can be started before the completion of phase II. However, this is challenging if a bone mineral density (BMD) endpoint is really required.

Table 5.2 outlines the differences in the inclusion and exclusion criteria for the different agencies and the duration of the treatment of established osteoporosis. Table 5.3 explains the endpoint requirements and Table 5.4 describes the type of fracture analysis that is required. The differences seem relatively benign on the surface, but they make a large difference to the number of subjects required and the study design.

There is currently a debate regarding whether the distinction between treatment and prevention should be maintained, because this depends on the definition of disease and assumes that prevention is an acceptable increase in BMD and treatment is the reduction in fractures. The regulations, as they currently exist, are shown in Table 5.5.

Table 5.1. Regulatory requirements for primary endpoints of phase II studies.

Primary endpoints of Phase II studies	Agency	Study duration
BMD	CPMP	2 years
	GREES	1 year
	WHO	3–12 months
BMD + biochemical markers	FDA	1 year
	Japan	Not precise
Biochemical markers (in some circumstances)	WHO	3–12 months

Table 5.2. Regulatory requirements for the recruitment and duration of trials for the treatment of established osteoporosis.

	Agency
Selection criteria:	
BMD T-score	
below −2 SD	FDA, GREES
below −2.5 SD	CPMP, Japan
Fragility fractures	CPMP, Japan
Duration:	
2 years	GREES
3 years	FDA, WHO, CPMP

Table 5.3. Regulatory requirements for endpoints of trials for the treatment of established osteoporosis.

Primary endpoints	Agency
Fractures	Japan, GREES, CPMP
Fractures	
• significant 3 years	FDA*
• 2 years and 3 years, phase IV†;	FDA*
• 5 years	FDA, WHO*
BMD	WHO*
Duration	
• 2 years	GREES
• 3 years	FDA, WHO, CPMP

* If normal data is obtained from animal studies.
† Historically, fracture data required at 2 years and 3 years, with a follow-up of 5 years, as a phase IV study.

This has changed with the advent of PTH, where 18-month data are now being accepted.

Table 5.4. Fracture analysis requirements in trials for the treatment of established osteoporosis.

Definition	Agency
Number of subjects with new fracture	CPMP
Incident fractures—worsening of previous fractures	FDA
Number of bone fractures—severity	Japan
"Validated criteria"	GREES
"No gold standard": Number of subjects with new fractures	WHO
Worsening of previous fractures	
Number of new fractures for time period	
Cumulative number of fractures	

Table 5.5. Regulatory requirements for the selection of women into trials for the prevention of osteoporosis.

	Agency
Prevention bone loss:	
Within 3 years of menopause	FDA
Within 3 years of menopause to prevent bone loss	WHO
After ovariectomy or early menopause	GREES
Within 5 years of the menopause in women with at least one increased risk factor for development of osteoporosis	CPMP
Osteopenia: BMD T-score between −1 and −2.5 SD:	
Within 5 years of menopause	GREES
No time limit	WHO
More than 5 years after the menopause in women with at least one increased risk factor for development of osteoporosis	CPMP

The regulatory environment is further clouded by the ethical environment and the success of drugs such as the bisphosphonates. The regulations were set up to expect placebo-controlled studies. Furthermore, the revised Declaration of Helsinki, Finland,[5] states that a new method should be tested against the best current proven methods. Osteoporosis is obviously within this category, which makes the use of a placebo an ethical issue if there are already proven treatments. Active comparator studies are, at first glance, the obvious way forward; however, the endpoints for such trials are tough to determine from both a practical standpoint and what will be acceptable to the regulatory agencies. For example, does an NME have to be more efficacious than an existing compound or can it have similar efficacy? Does efficacy have to be compared using a fracture endpoint or will BMD suffice? How should the safety profile be compared? The answers will influence significantly the trial design and size of the subject population required in studies. For example, it would be an extremely high-risk strategy to show that an NME had superior fracture prevention properties compared with, for example, the bisphosphonates, which have demonstrated a >50% reduction capability in 3 years. The cost and size of the study alone would be prohibitive.

It is, therefore, impossible, at the time of writing, for the author to provide any definitive guidance on the regulatory issues for the treatment and prevention of osteoporosis. The best information that can be provided is addresses of websites for the reader to refer to for the latest developments. It is perhaps easier to gain a good understanding in the USA because the FDA expect "face-to-face" meetings in the planning and development of each programme. Unfortunately, it is not possible to do this with the European agencies, and, therefore, companies are left with difficult decisions when designing a new clinical programme.

5.3. Special Data Collection

As part of the planning stages of the clinical trial, the sponsor must consider the organization of the dual-energy X-ray absorptiometry (DXA) measurements and, if they are being collected, the samples for the biochemical markers and the X-ray films for vertebral morphometry.

If the study is part of a submission to a regulatory agency, an Imaging Core Laboratory (ICL) or a DXA quality assurance (QA) centre will have to be selected, because they are a requirement written into the guidelines for both the European agency and the FDA. Even if the study is not for regulatory submission, a QA centre is highly cost-effective, not only in savings on the direct cost of data entry, but also by ensuring a higher quality of data. From the author's personal experience, if a QA centre is not involved, between 3% and 25% of the data could be invalid or analysed incorrectly. Cost is one component of this issue; the other component is the ethical implications of having lost subject data that could have been salvaged.

A DXA QA centre will also be able to provide input into the study design and ensure the correct terminology is used. For example, there is often confusion in the use of the terms "femoral neck", "trochanter", "total hip", and "hip". All of these measurements are obtained from a single scan of the proximal femur or hip, but

each component provides a slightly different piece of information about the femur, because each measurement is of a different subanatomical area and includes a differing percentage of cortical versus trabecular bone. While writing the protocol, a protocol appendix dedicated to the DXA measurements and collection should be seriously considered.

Selection of a DXA QA centre should be carefully considered. Historically, the DXA data used to be handled by academic centres. Now, there are professional QA centres or ICL's set up to provide this kind of service. Sponsors should visit the QA centre before initiating any contract with them, although all too often this step is missed in the interests of time. Sponsors would not expect to conduct a study at a trial site that they have not visited, so similarly, the same level of detail should be applied to the QA centre. Although you might not be familiar with DXA measurements, a good QA centre should be able to show you around their facilities and demonstrate the software for all the makes of densitometers you anticipate using in your study. Furthermore, they should be able to answer any technical questions and give you a basic understanding of the measurement techniques. If the software cannot be demonstrated or questions are left unanswered, it would suggest that your study might be the launch of a new service for the QA laboratory, which is probably not want you want! Many places advertise their experience with DXA and either do not have the experience or subcontract the work out. Another question to ask is how many DXA technologists they employ and what their experience in performing DXA QA is. If they have only one technologist, what happens if he or she leaves midway through the trial? Find out whether they understand the items they will need to check on a DXA image before selecting the centre (Table 5.6).

Another aspect of the lumbar spine measurement that has to be considered is the number of allowable vertebrae. The best precision will be obtained with four

Table 5.6. Items to be reviewed for analysis by a DXA QA centre.

AP spine:
1. Correct identification of the four lumbar vertebrae at baseline
2. Identification of the same four vertebrae at follow-up
3. Identification of the top of L5
4. Identification of the base of T12
5. Are the bone edges OK?
6. For Hologic—is the ROI correct?
7. At follow-up, is the ROI within 5 pixels the same size as baseline?

Femur:
1. Is the femur vertical in the scan window?
2. Are the bone edges OK? (Only very, very rarely are these wrong.)
3. Is the femoral neck box positioned correctly?
   Lunar—midpoint of the femoral neck.
   Hologic—tangential to the greater trochanter.
4. Is the midline positioned correctly?
5. Is the analysis sufficiently far down the femoral shaft?
6. For Hologic—is the ROI correct?
7. At follow-up, is the ROI within 5 pixels the same size as baseline?

vertebrae, and this will become poorer with each successive loss of a vertebra. Therefore, ideally four vertebrae would be used. However, in a deformity study, it would be anticipated that there will be some subjects with a deformity in the lumbar vertebrae. Recruitment would be very challenging if no deformities were allowed. Therefore, the suggested optimum allowance is three evaluable vertebrae at baseline, so that during the course of the study, if a second lumbar deformity occurs, two vertebrae would still be measurable. The measurement of only one vertebra would be considered of too poor precision to be acceptable under normal circumstances.

5.4. Biochemical Markers

Laboratories that can assay bone biochemical markers might need to be selected, in addition to the choice of specific assays. There are several laboratories in the world that can perform these specialized assays, and the list is growing. Again, the sponsor should evaluate the laboratory by personal visit and going through their standard operating procedures (SOPs). As with the DXA QA laboratory, it should be possible to conduct a systems audit without an in-depth knowledge of the technology being employed.

5.5. Vertebral Morphometry

The evaluation of Vertebral Morphometry, as a measure of efficacy, should be performed by an independent third party. Most investigator sites do not have the experience to perform the careful evaluation of vertebral deformity, plus an element of investigational bias is also introduced if the evaluation is left with the local investigators. Historically, vertebral morphometry was completed using a very manual approach, by reading the films and marking the morphometry points with a wax pencil. The distance between the points was then measured with a ruler or pair of callipers. Nowadays, everything is digitized and, although the morphometry points still have to be applied, there are now software packages that enable this to be semiautomated. However, this electronic application of points still requires some human intervention and interpretation.

A decision must be made regarding the methodology for vertebral deformity assessments, whether to use a so-called "semiquantitative" (or "pseudoquantitive") scoring system or a fully quantitative methodology, or both. Both methodologies have their advantages and disadvantages, but if a trial is for registration of the NME, both techniques will have to be employed. Historically, this has also required an adjudication process if there has been discordance in the deformity determination between the two techniques for individual vertebrae. This then becomes a very expensive and time-consuming endeavour. A newer approach has been used, in which the radiologist reviewing the digitized radiographic images scores them using a semiquantitative technique. The results are entered into the computerized system in real time. The second step is to electronically toggle on the preassigned morphometry points. Once the radiologist is in agreement with the morphometry placement, the computer system displays both sets of results, that is the semiquantitative

and quantitative techniques, and an adjudication, by the same reader, can take place. This methodology requires fewer readers, all the data are available in real time, and the sponsor does not have to wait for final adjudication to take place.

The choice of reader and radiographic QA centre or ICL is, again, important. Historically, this was always performed at academic centres, but now there are professional ICLs' that perform this function and have the full infrastructure to enable this to be performed seamlessly, either as a stand-alone service or with the bone radiologist(s) of your choosing. In either case, as a sponsor, ensure the laboratory has the capabilities and the SOPs in place to perform the work. The time it takes at the front end of the study, to ensure everything is in position and will operate smoothly, far outweighs the time that could be spent once the study is underway trying to salvage a poorly handled QA process. The same ICL should be selected for both DXA and radiographs, not just for the convenience, but so that the radiographic assessment of the lumbar vertebrae can be used to ensure the correct vertebrae are evaluated for DXA. Furthermore any abnormalities seen on the radiograph are automatically fed back to the DXA results to remove vertebrae that may have disease that would preclude them from DXA evaluation.

5.6. Couriers and Data Transfer

With all the data that must be handled and sent to various laboratories, it is important to ensure a good courier system is in place for the study. Most of the laboratories or service providers described above will suggest a vendor according to their experience. Although you might have a preference for a particular courier company that might seem less expensive, it might not be able to handle the requirements for the study. Generally, the laboratories handling this specialized kind of data have tried most courier companies and experience will have shown which ones are most suitable. It is advisable to always use the courier companies that are recommended. Since the DXA data is all digital and many X-ray facilities are using digital X-ray systems, it could be anticipated that the image data could be sent via File Transfer Protocol or FTP. Unfortunately, many investigator sites do not have this capability, and those that do require someone other than the technologist to perform this operation. Therefore, since the critical step is to move the images to the ICL as swiftly as possible, the courier system is still usually the simplest and most reliable method for the site. The delay in the image transfer over an FTP is only a few hours, so the cost benefit in most situations still resides with courier transfer, at the time of writing.

5.7. Investigator Meetings

These are now standard start-up procedures for any clinical trial, so what is different in the field of osteoporosis? It is certainly advantageous and cost-effective to have a representative of the laboratories (both imaging and biochemical) give a presentation at the meeting, not only to provide an overview of their services, but to provide a data flow and answer any technical and specific questions. For some of the samples, there might be special storage and shipping requirements, which must be explained. Another facet of the investigator meeting to consider is

whether to include training for the DXA technologists. Most QA centres recommend this for several reasons:

1. The DXA technologist is going to be collecting your primary and safety efficacy data—make sure they are properly trained and know what they are doing.
2. It provides the technologists with an opportunity to talk to the representatives from your QA laboratory directly and build a rapport. This can help tremendously at the start-up when some of the issues might be unfamiliar for the technologist. They are more likely to call the QA laboratory for support and ask questions if they know someone at the laboratory. This is far more preferable than the technologist muddling on and sending scans that are incorrect or transmittal forms that are not properly completed.
3. The DXA technologist is made to feel part of the site investigational team. Unlike seasoned investigators, they will rarely travel, and the bonus of travelling to an investigator meeting will go a long way to help motivate and encourage a crucial member of the support staff. It also ensures the investigator and subinvestigator get to know their team and can remove some of the barriers to good internal cooperation.

5.8. Cross Calibration of DXA Scanners

There is a need to cross calibrate the densitometers used at different centres taking part in the trial. Essentially, a so-called "gold standard" phantom (or phantoms) must be measured on each densitometer. This can be performed either by a representative of the QA centre visiting every site in turn or by sending the phantom around by courier. The cheapest methodology is obviously sending the phantom by courier. The disadvantage of this option, however, is that this takes considerable time. As a rule of thumb, you should allow 1 week per instrument for each phantom that is on rotation. The reason for this length of time is purely logistical. Let us assume that the phantom is sent out by courier on a Monday, it gets to the site on Tuesday or Wednesday, and the site scans the phantom on Wednesday or Thursday. At best, it will be sent out again on the Wednesday or Thursday, for arrival at the next site by Friday or the following Monday. This assumes that the site is primed and has allotted sufficient time for this to be accomplished. This method works more efficiently within the USA, where there are no borders to cross between countries, but is more problematic in Europe and further afield.

Site visits, although quite costly, can provide some additional benefits:

1. A DXA site audit, or prequalifying visit, can be performed.
2. Training can be given to the technologists there. This can help supplement the training at the investigator meeting, if the technologist attended, or be an alternative to the training that would have been performed at the meeting. This also enables the technologist to ask questions one-on-one rather than in the group setting, which some find intimidating. Furthermore, if more than one technologist is at the site, they can also be involved with the training, which they would otherwise miss if they had been unable to attend an investigator meeting.

3. Time is obviously a major gain. On average, it should be possible for one representative to complete one site visit per day. It should, therefore, be possible for all but the largest studies to have all the site visits completed either before or within a short timeframe of subject recruitment at each site.

Ideally, the phantom should visit each site at the start of the study, before subject enrolment, but this is not practical in many cases, so other instrument data must be obtained to ensure the DXA equipment is operating optimally.

5.8.1. Selection of Phantoms for Cross Calibration

The choice of phantoms for cross calibration is fairly limited. It should be remembered that phantoms are, at best, subject mimics that are designed to provide an assessment of how a machine is operating. DXA instruments are designed to measure people, not phantoms, and, therefore, the challenge has been to produce a phantom that works and provides precise and accurate measurements but is not too costly to make or transport. A number of phantoms have been made and each DXA manufacturer makes their own, but only three phantoms have regularly been used for cross calibration. These are discussed in detail in Chapter 4.

Some QA centres believe that each site should have the same type of phantom for their daily QA. This often varies from study to study, with one QA centre preferring the European spine phantom (ESP), for example, and another the Bona Fide phantom (BFP). It, therefore, ends up that some sites have to measure a series of different phantoms each day to satisfy the demands of different clinical trials. Within the realms of the precision and measurement statistics, this is unreasonable and unnecessary.

5.8.2. Cross Calibration Statistics

The concepts here differ between QA centres, but are rarely discussed or publicized. Essentially, there are two main methodologies available.

5.8.2.1. Methodology 1

All of the subject data are converted to a single DXA machine and every minor calibration shift is corrected. This methodology works on the premise that a single absolute BMD is to be used for assessing the drug effect on each subject. Essentially, the cross calibration data from the phantom rotation provide the basis for an equation to be developed for each site. This then enables the adjustment of that site's subject BMD to a "standard BMD". Therefore, all data are corrected to a single calibration system. During the course of a trial, if a DXA instrument goes out of calibration, further minor corrections are applied to the subject data. This also makes the assumption that the phantom data are a good subject mimic and reflects the subject data.

5.8.2.2. Methodology 2

For most clinical trials, the endpoint is the percentage change from baseline for each subject averaged as a group. Each site is block-randomized with regard to treatment,

and, therefore, the site effect on BMD is minimized. Furthermore, there is a precision or measurement error on every phantom measurement. Therefore, a minimalist approach is taken to changes in subject data. The data remain unchanged, which, if nothing else, helps the audit trail. The only time the subject data are altered is when the calibration shift at any individual site is twice that of the error around the phantom measurement. The cross calibration data provided by the phantom are only used for providing comparable baseline demographics and changes in calibration with an individual instrument.

5.8.2.3. Methodology Summary

Although both methodologies are employed, the second methodology has the greater merit, and has now been used in many recent publications. It keeps the data cleaner and provides the minimal introduction of further measurement error. It is also less work for the QA centre or sponsor, and yet arguably achieves the same outcome. There are disadvantages to using the percentage change from baseline in the analysis, however, and statistical methods are provided in Chapter 9 that cope well with the intersubject and intermachine variation in BMD.

5.9. Inclusion and Exclusion Criteria

For most trials in osteoporosis, BMD is an inclusion and exclusion criterion. Although in one respect it seems obvious to define BMD using T-scores, there is a major potential for error here. T-scores are derived from a population that is either defined by the DXA manufacturer or one the local site has created. Each of these databases is, therefore, influenced by the population used and the statistics used to provide this population curve. Therefore, it is important to identify the population curve to be used and then provide the absolute BMD to the sites for the cut-off criteria, for example the Hologic female Caucasian or the Lunar European male normal population curve. The problems of not doing this were first recognised in 1994.[6] Several studies had instructed principal investigators to use the Caucasian American female reference population that was supplied with their instrument. Because of the different populations used, one manufacturer's database had a slightly higher young healthy normal population than the other for the neck of femur region. This enabled subjects at these sites to be more easily recruited and had been noted by sites that had more than one type of densitometer. This was supposedly changed with the advent of the standardization to the National Health and Nutrition Examination Survey (NHANES)[7–9] data. Sadly, this is still not the case because of the population statistics that have been employed. The standard deviation (SD) for Hologic populations for spine L1–L4 measurements is 0.11 g/cm^2 compared with 0.12 g/cm^2 for Lunar equipment. Because inclusion and exclusion criteria are normally derived from the number of SDs from peak bone mass, it can be readily appreciated that sites using Hologic equipment will have an easier time recruiting subjects with, for example, an SD of -3 or more compared with Lunar sites in terms of absolute BMD (even taking the calibration differences between Hologic and Lunar instruments into account). The reason for

this discrepancy is the use of different statistical algorithms by each manufacturer. Both have valid reasons for their use.

The best way around this problem is to select a cut-off from one reference database, define the absolute BMD that this represents, and then calculate this using a cross calibration equation for the other densitometer. Although it should be arbitrary which database is selected, most studies have used the Hologic Caucasian database and converted it for Lunar systems. If the reverse is done, principal investigators using Hologic equipment tend to complain because the numbers do not match their database and it is harder for them to recruit subjects on this basis. Lunar sites, by contrast, tend to have less issue with converted BMD criteria, because a converted BMD from a Hologic instrument provides a slightly higher value than using a strict T-score defined by the Lunar database. By providing these values, a sponsor ensures a similar demographic spread between sites regarding BMD and equal recruitment between sites using different manufacturers' equipment.

Z-scores (i.e. the number of SDs a BMD measurement is above or below the age-matched normal range) are generally not used for inclusion or exclusion criteria for the very reason that they are age-related and, therefore, become very complex to detail. As an endpoint, if age is a confounding factor in the analysis of the study, Z-score changes can be an excellent methodology for removing the confounding variable. This has certainly become the standard practice in paediatric studies.

5.10. Data Flow

Some serious consideration must be given to the data flow with respect to the biochemical, DXA, and morphometry data. The priorities will differ, depending on the phase and complexity of the study. For blinding reasons, it is recommended that the DXA technologists are instructed merely to acquire the data and not to analyse it. The exception might be at baseline for inclusion and exclusion criteria, but it is now becoming standard practice to have the QA centre analyse or reanalyse all scans. From a subject safety viewpoint, the QA centre should be monitoring and flagging any DXA scans that show percentage or absolute change from baseline that could be considered a subject management issue. These limits should be determined *a priori* at the start of the study for these data and similarly for vertebral deformities. A data safety monitoring board (see below) should also be involved in setting out these guidelines and reviewing the data.

There is often a discussion regarding whether the sites should receive a copy of all the data at the end of the study. If the central laboratories and QA centres are considered the repository of all the data, this should be sufficient to ensure data integrity and satisfy external audit. Some sponsors believe copies should be returned to the site, for complete data sets. This is a lot of extra work for both the centre and the sites, and does not guarantee everything will get back to the sites' central documentation. There is a different argument for subject management, because there is need for the treating physicians to know the full subject records for the study, depending on the trial and the treatment.

The other question regarding subject flow is who makes the final decisions on inclusion and exclusion if DXA or vertebral deformity is involved. For all major studies, this should probably be handled by the QA centre, but it then requires the centre to be able to operate a good, rapid, and responsive turnaround time with the incoming data. If the QA centre fails in this respect, the investigational sites rightly become very annoyed and lose motivation for the study.

5.11. The Data Safety and Monitoring Board (DSMB)

The sponsor can appoint a DSMB for the trial. Their role is to primarily review unblinded data and ensure the safety of the individual subjects during the course of the study. This is achieved by the following means:

1. Reviewing the monitoring reports from the clinical research associate (CRA)
2. Reviewing adverse event (AE) and serious adverse event (SAE) reports
3. Monitoring accrual rates into the trial
4. Deciding on changes to the inclusion and exclusion criteria
5. Agreeing changes in sample size
6. Overseeing interim analyses of the trial data
7. Agreeing to early stopping of the trial, if necessary
8. Reviewing individual subject data from a safety perspective (optional).

Ideally, the inclusion and exclusion criteria and sample size should not change as the trial progresses.[10] Such changes might be necessary in light of the monitoring reports, AE reports, and accrual rates. Sample sizes can be changed, particularly in long-term trials, because new information about the magnitude of a possible treatment effect becomes available. Alternatively, the outcome of an interim analysis might enable the sample size to be changed.

Interim analyses should be handled with care. They should be planned in the protocol before the trial starting, rather than succumbing to the temptation to see how the results are coming along. The statistical methods should be defined in advance and all the staff involved in the day-to-day running of the trial should remain blinded to the results of such analyses. Investigators should only be made aware of the changes to the protocol that arise because of the results of the interim analysis.

Stopping a trial early should only be considered on ethical grounds, either because of the safety data or because it becomes apparent that the power of the trial is not acceptable according to an interim analysis. If it is intended to stop the trial because an adequate treatment effect has been demonstrated, the interim analysis should be planned and in the protocol.

5.12. Trial Audit

All trials sponsored by the pharmaceutical industry will be audited and there is good documentation in the regulations regarding how these should, generally, be

conducted. However, the question arises as to how these audits might differ in a study on osteoporosis compared with other therapeutic areas, and what, if anything, has been carried out differently by the FDA.

From a sponsor's perspective, audits should be carried out much earlier on in the study than for other therapeutic areas because of the long duration of these studies. If there were issues at a site, laboratory, or QA centre, it could go undetected for a significant period of time if these are not performed in a timely fashion. It has now become a recognised practice that the centralized facilities can become the holders of source data, in addition to the investigator sites themselves. This, therefore, negates the issue of having all the results returned to the sites for audits. However, the *sequela* to this is that the central facilities also need to be audited, and this is where the challenge comes in from the sponsor's point of view. Investigator site audits should pose no unusual problems to a seasoned auditor, the only difference might be the need to see the DXA instrument and review the documentation that is stored by the technologist there, or visit the X-ray department if spine radiographs have been performed. For most studies, the technologists are involved in the transmission of the data to the QA centre. Therefore, there should be some documentation present to show the transmission of all images. Furthermore, there is likely to have been some correspondence between the QA centre and radiographers during the course of the trial, which, again, should be documented. Finally, if instruction manuals for scan and image acquisition have been issued, the question is raised as to whether they in the department where the staff who are involved in obtaining the DXA or radiographs can easily access them?

How to handle the core laboratories? Blood or biochemistry laboratories are required to have SOPs in place, participate in external quality control, and have usually obtained external accreditation. This makes it relatively easy for the sponsor to carry out an external audit of such laboratories. QA centres will also be well established, with good procedures although there are no external accreditation, enabling similar audit by the sponsor. A full systems audit should be performed and a review of the documentation for a sample of subjects. When performing an audit on the QA centre, it is unusual for the sponsor to have an auditor who can read DXA and X-ray films. A good QA centre will have a second review process built into their data flow to ensure that two pairs of eyes look at each piece of data. However, more recently there are some unpublished (as yet) examples of another independent group with DXA and/or morphometry experience auditing, for example 10% of the primary QA centre's scans and films. This provides a very robust methodology and, although it obviously increases study costs, it should provide a good sense of how the imaging components have been managed.

At the time of writing, the FDA has not audited any QA centres directly for DXA or X-ray handling, although there is helpful guidance available. However, in other therapeutic areas this has been conducted and in fields such as rheumatoid arthritis and oncology the FDA has often requested a complete set of digital images in a format in which they can review them. If this trend continues, we can expect to see the same requirements being requested in the field of osteoporosis.

5.13. Trial Closure

At the end of the trial, the CRA will ensure that each centre has submitted all the required documentation on all trial subjects. The clinical research organization (CRO) appointed by the sponsor will then review all the data to ensure that there is a clean data set available for analysis. The CRA should produce a final monitoring report for each participating centre.

It is then the responsibility of the sponsor to carry out the statistical analysis of the data and produce a final clinical study report and any peer-reviewed publications. Guidance on the format and content of clinical study reports is available from the FDA.[12] Chapter 9 deals with the analysis and presentation of results for peer-reviewed publication.

5.14. Summary

Although there is a standard set of tasks that have to be conducted in clinical trials, those in osteoporosis or bone-related diseases require some unique extra details that have to be addressed. Arguably, the most crucial of these is the assigning of a QA and biochemistry laboratory, because they will need to be ready to provide kits to the trial sites before subject enrolment. Furthermore, having these teams identified early in the trial process will provide some technical and consulting support for protocol development, if that is required.

Although all trials should have good documentation and audit trails, this is crucial in trials in which BMD or fracture is an endpoint. The trials will normally be running for several years and, even with a normal turnover of staff, this will mean that very few people involved in the trial at the start will still be working on it at the end. Therefore, without good documentation, it will be difficult for staff writing up the final reports to follow some of the decisions that were made and the rationale for them.

Because trials evaluating bone measurements are some of the longest in duration in the pharmaceutical world, it is crucial to ensure that there is good planning at the front end. Not only is it costly to repeat a trial, but also in this arena it could be several years before the errors in the planning are finally noticed. However, if everything is set up well, after the busy subject recruitment period is over, most trials get into a "natural stride" or rhythm; this is particularly so within this field, because the normal running period is quite extended. From a sponsor's perspective, as with all trials, good up-front planning is crucial to the good execution of the trial, and no more so than in the field of osteoporosis.

References

1. FDA Guidelines for the Preclinical and Clinical Evaluation of Agents used in the Prevention or Treatment of Postmenopausal Osteoporosis (1994). http://www.fda.gov/cder/guidance/osteo.pdf accessed 13/11/06

2. CPMP Notes for Guidance on Postmenopausal Osteoporosis in Women (2001). http://www.emea.eu.int/pdfs/human/ewp/055295en.pdf accessed 13/11/06
3. World Health Organization (1998). *Guidelines for Preclinical Evaluation and Clinical Trials in Osteoporosis*. Geneva: World Health Organization.
4. Reginster, J. and Compston, J. (1997). GREES recommendations for the registration of new drugs in the prevention and treatment of osteoporosis. *Calcif Tissue Int* **60:**488.
5. Recommendations Guiding Medical Doctors in Biomedical Research Involving Human Subjects. Adopted by the 18th World Medical Assembly, Helsinki, Finland, 1964, and revised by the 52nd World Medical Assembly, Edinburgh, 2000. http://www.wma.net/e/policy/b3.htm accessed 20/11/06
6. Miller, C.G. and Barden, H.S. (1994). Entrance criteria for clinical trials with DEXA. *J Bone Miner Res* **9(Suppl. 1):**S209.
7. Looker, A.C., Wahner, H.W., Dunn, W.L. *et al*. (1998). Updated data on proximal femur bone mineral levels of US adults. *Osteoporos Int* **8:**468–89.
8. Faulkner, K.G., Roberts, L.A. and McClung, M.R. (1996). Discrepancies in normative data between Lunar and Hologic DXA systems. *Osteoporos Int* **6:**432–6.
9. Chen, Z., Maricic, M., Lund, P. *et al*. (1998). How the new Hologic hip normal reference values affect the densitometric diagnosis of osteoporosis. *Osteoporos Int* **8:**423–7.
10. Hubbard, W.K. (1998). International Conference on Harmonization: Guidance on Statistical Principles for Clinical Trials. *Federal Register* **63:**49583–98. http:// www.fda.gov/cder/guidance/91698.pdf accessed 13/11/06
11. Blunt, B.A., Jones, D., Svensen, R.K. *et al*. (1998). Good clinical practice and audits for dual X-ray absorptiometry and X-ray imaging laboratories and quality assurance centers involved in clinical drug trials, private practice, and research. *J Clin Densitom*. **1**:323–37.
12. Expert Working Group (Efficacy) (1996). International Conference on Harmonization: Structure and Content of Clinical Study Reports. *Federal Register* **61:**37320–43. http://www.fda.gov/cder/guidance/iche3.pdf accessed 20/11/06

6

Local Site Organization

Dorothy Adams and Colin G. Miller

6.1. Introduction

When your clinical centre has been notified that you have been selected as a site for a new clinical trial in osteoporosis, you can be assured that the sponsor has made an assessment of the subject population and facilities available. It is now up to the investigator, site administrator, and staff to prove the sponsor made the right decision in placing the study at their site. Because delays in study start-up can cause a sponsor to lose millions of dollars, it is imperative to get the study up and running as soon as possible. During this process, the emphasis is on speed and accuracy. Always remember that there are many qualified sites able and willing to take on this project should your commitment falter. You can congratulate yourself that you have been chosen to participate, but the site must be organized to get to work on the project quickly!

It is important to ensure that all staff, equipment, and facilities are in place to get the study up and running as soon as possible. The investigator must be willing to have a flexible schedule, in order to handle clinic visits for evaluation of potential subjects once recruitment starts. The scheduling staff, the DXA technologist, pharmacist, and laboratory staff must be included in study planning and trained for their roles in the study.

6.2. Administration

Although the study has been placed at the site, it is still important to show commitment to the project and get the paperwork started. Two actions should be put in place immediately: the Independent Review Board/Independent Ethics Committee (IRB/IEC) submission, and negotiation and signing of contracts. The time required to accomplish these processes will vary from institution to institution. The site must be able to predict the time required and facilitate the process because this is, generally, a major cause of delay in study start-up.

6.2.1. IRB/IEC Submission

Once you have a final (rather than draft) copy of the protocol, immediately determine the next IRB/IEC meeting date and deadline for submission. Contact the IRB/IEC administrator with a request for the project to be on the agenda for that meeting; if you do not already have submission requirements, make the request for them at this time. If you are aware of any specific requirements of your IRB/IEC that might need negotiation with the sponsor (e.g. "compensation for injury clause" in the consent form), get the discussions started before submission.

The compensation for injury clause is, generally, the most controversial for the legal staff of both the institution and the sponsor. To remedy the problem before submission, contact the sponsor immediately, providing them with the institution's required language. This can take time so it is best to get this settled *before* the IRB/ IEC protocol submission. In the UK, there are now standard no-fault compensation arrangements with the Association of British Pharmaceutical Industries (ABPI) and national guidance concerning the compensation clause (see Appendix 3.1).

One other issue to consider before IRB/IEC submission is whether or not the protocol calls for storage of blood or tissue specimens for later study, or whether it calls for study of genetic markers. This is a particularly sensitive issue at present, particularly if blood or tissue specimens are to be used for genetic studies, or stored indefinitely for future research. The storage and future use of samples must be made clear in the subject information sheet and explicit on the consent form. Most IRB/IECs have special consent form requirements for these issues so make every effort to deal with them before submission.

In studies of osteoporosis, the IRB/IEC might be concerned about radiation exposure owing to the BMD tests. Address this issue in your IRB/IEC submission documents by pointing out the amount of radiation in BMD tests compared with that from natural background radiation. This should be included within the subject information so that they can make an informed choice about the study.

Prepare your submission exactly as required by your IRB/IEC, and get it to them on time! On the IRB/IEC submission cover sheet, you should indicate which attachments are included. For example, these attachments might consist of the following:

1. A copy of the protocol and date of the protocol
2. The subject information sheet, with version number and date
3. The consent form, with version number and date
4. The investigator's brochure for the new molecular entity (NME) under investigation
5. Recruitment advertisements, if appropriate.

This is helpful because the IRB/IEC staff, generally, deal with a lot of documents and you might need proof of what you have submitted in case any items are misplaced. The cover sheet, with its list of attachments, should serve as back-up evidence for the documents reviewed in the IRB/IEC approval process. Many trials

are held up at the IRB/IEC stage because of simple administrative errors. Are the information sheets and consent form on the right headed notepaper? Have they been spell-checked and grammar-checked? Are they the right versions? Have the correct brand names been used for branded drugs? Getting the submission right eases the process for the investigator and IRB/IEC.

6.2.2. The Study Contract or Clinical Trial Agreement (CTA)

The study contract or CTA can also be responsible for delays in study execution, so deal with it as soon as it arrives from the sponsor. This is not as complicated if you are at a site where the principal investigator has the ability to take the responsibility and sign off on the document. However, many sites do not have this luxury and must deal with the Finance or Corporate Affairs Department of their institution.

The delays in signing off the CTA vary from institution to institution but some more progressive institutions have worked out a general contract agreement with several sponsors and have been able to speed up the process. It is important to know whether, and with whom, this has been done at your institution. Unfortunately, commonly this has not been previously negotiated and will take a lot of time because it can involve both financial and legal staff.

Osteoporosis trials are always lengthy and involve several years of follow-up evaluations. It is important to confirm that the study budget allows for an annual increase in the cost of services necessary to conduct the study. It is also important to ask the sponsor to provide a letter of indemnification, assuring they will cover you for any liability during the conduct of the study.

6.2.3. Regulatory Documents

Once the IRB/IEC submission and contract negotiations are put in motion, you can start on other documents that are more under your control, including the following:

1. US Food and Drug Administration (FDA) Form 1572
2. *Curriculum vitae* (CV)
3. Laboratory documents
4. Equipment calibration documentation
5. Protocol signatures.

Prepare your own Form 1572 rather than using the one prepared by the sponsor. As a study progresses, there might be changes of coinvestigators, requiring the form to be updated. The forms can be kept on computer, which makes changes easier, if necessary, during the study.

Keep current CVs for all investigators easily available on computer, in addition to a file with a signed and dated copy, so they can be duplicated for submission with Form 1572. Keep a current copy of the investigators' medical licenses or statutory registrations attached to their CVs, which should also be included with the submission.

For each local laboratory involved with the study, maintain a packet of documents that includes the following:

1. A copy of current laboratory certification
2. A list of laboratory normal ranges
3. A copy of the CV of the laboratory director.

Keep this information up to date and duplicate it, as required, for submission. You will find that most studies use a central laboratory so you will not need to submit these documents. However, you will need to get these items from the central laboratory to complete your study files.

Obtain documentation that gives the calibration of any dual-energy X-ray absorptiometry (DXA) instrument used. It is important in a multicentre trial to define how the crosscalibration was obtained and which instrument was used as the master calibration. Even in single-centre trials, calibration against a standard phantom will reassure the sponsor and can enable the trial data to be combined with other studies more easily in future.

Once the FDA Form 1572 is completed and you have the necessary attachments (CVs of the investigators and laboratory documents), it is time to meet with the principal investigator, who must sign the Form 1572 and protocol. It is important to note the date of the protocol and be sure you are working with a final copy rather than a draft.

6.2.4. Document Submission

It is recommended that you send a covering letter to the sponsor with all document submissions. This provides a tracking mechanism, giving you a record of what has been submitted and when. The first such letter will, generally, state that the following documents are enclosed:

1. Protocol signature page
2. FDA Form 1572
3. CVs and copies of medical licenses for the investigators
4. Laboratory documents (if necessary)
5. Documentation of equipment calibration (if necessary)
6. The signed CTA.

Tell the sponsor of the expected IRB/IEC review date and any other information, to assure them you are still committed to the project. Do not delay the submission if the CTA is not ready, but update the sponsor on its progress in the covering letter.

6.2.5. IRB/IEC Review

In the week following the IRB/IEC meeting and your project review, it is a good idea to follow up with the IRB/IEC staff regarding the status of your documents. Commonly you will have to answer some questions for them or make revisions to the consent form. The sooner you know what they want, the sooner you can get issues resolved and the final IRB/IEC approval documents. Unfortunately, in most

cases, you will need to deal with the sponsor's legal staff if changes are required to the consent form. Anything you can do to keep the process moving is highly recommended, but do not forget to show appreciation for IRB/IEC staff's efforts because not only do they deserve it, but also their cooperation is crucial to a speedy study start.

6.2.6. IRB/IEC Approval

Once you have the IRB/IEC letter of approval and the approved information sheet and consent form, write to the sponsor again with the following enclosures:

1. A copy of the IRB/IEC letter of approval
2. A copy of the approved information sheet and consent form
3. Any conditions of approval
4. The IRB/IEC membership.

Use a courier to send all study documents to the sponsor because you can then rest assured that they will get them the next day and it will also provide a tracking system for the documents, if necessary.

6.3. Study Initiation

Immediately following submission of the final documents, the supplies and study drug will arrive and initiation of the project at the site is ready to take place. Study initiation is an important time-point in any study. This is your chance to get all the staff involved in the study at the site together for training on the study requirements and it serves as a signal that the project is now ready to get underway. It is important to include as many staff members as possible:

1. Principal investigator (mandatory)
2. Subinvestigators
3. Coordinators
4. Technologists responsible for DXA and any laboratory testing
5. Pharmacists.

It is important that each individual clearly understands their role and responsibility within the project. Involving them in the project at the time of initiation by making them aware that they are part of the team does this best!

The study initiation visit is another chance to show off the quality of your study site to the sponsor. You can count on at least one person from the sponsor being present for the study initiation and they will need no less than 4 hours of time with the study staff. They usually require at least 1 hour of that time with the study physicians. These studies are best performed with physicians who have a busy subject schedule, so it can be a challenge to get the initiation on their schedule. Schedule the time with the physicians for their lunch hour and provide some sandwiches, which is a treat appreciated by the physicians, in addition to the sponsor staff!

Ensure the arrangements for storing and issuing the study drug and maintaining blindness (if appropriate) are in place. The storage space for the study drug must be secure and environmentally appropriate. Ensure that a log of environmental conditions (e.g. temperature and humidity readings) is maintained. The pharmacy is often crucial in the randomization and blinding process and, therefore, in the smooth running of the trial.

6.3.1. *Training*

Training for the conduct of a specific clinical trial usually takes place at the investigators' meetings. Sponsors make every effort to involve the key people from each site in the training sessions. The invitation to the meeting for osteoporosis trials would, generally, include the person responsible for the BMD scans. It might be difficult to arrange the physicians' schedules so they can attend these meetings, but it is important to make an effort. The sponsors must be assured that the physicians are really interested and committed to the study.

Basic training for coordinators is available through several organizations, such as the Association of Clinical Research Professionals (ACRP) (www.acrpnet.org). The ACRP is the organization that provides a certification programme for Clinical Research Associates (CRAs) and investigators (or clinical research coordinators) who have 2 or more years' experience. In the UK and Europe a similar organization is the Institute of Clinical Research (www.instituteofclinicalresearch.org).

Because the primary evaluation tool for osteoporosis studies is BMD, proper training and education of the DXA technologists is crucial. It is best to have a lead technologist for the study, but it is also important to have a back-up technologist properly trained. The technologists will need to understand the necessity for frequent calibration of the equipment and maintenance of a log-documenting calibration. They must be trained in the use of phantoms in running the quality assurance tests on a daily basis. They must be trained to archive the study subjects' scans on a daily basis, to prevent the potential loss of data in the event of a computer failure. The DXA technologists are crucial for producing a successful osteoporosis trial and must be treated as equal contributors to the study.

6.4. Recruitment Methods

The most effective method of subject recruitment for clinical trials is to work with the investigators in reviewing their existing database. With much of this information in a computer database, one can, generally, request a list of subjects according to diagnosis or age, in addition to several other protocol requirements. Once a list of potential subjects is secured, a fine-tuned screen can be performed with the physicians' records.

The legal position on contacting potential subjects on behalf of a physician varies from country to country. The minimum requirement is to have the physician's

approval to contact the subject on their behalf. In some countries (particularly in the European Union where there are stricter laws controlling access to confidential subject information), the first contact in recruiting a subject must come from someone who has a legal right to access the subject's medical record. Usually, this is the physicians themselves. In most circumstances it is illegal to release names of potential subjects to the CRA before the subjects have expressed an interest in participating in the trial.

Recruitment through the physician's practice database is more effective because the subjects have an established relationship with the physician and the clinic staff. They are more likely to be reliable in terms of follow-up and retention in the clinical trial.

However, it is not possible to rely on this method of recruitment alone. To meet timelines and avoid issues of confidentiality, it is possible to advertise for potential subjects. Sites have been very successful in advertising for subjects with diagnoses in most therapeutic areas. Women with osteoporosis are very likely to respond to an advertisement for a clinical trial because most recognise that there are still many controversial issues in how best to treat women during their postmenopausal years. The process of controlled clinical trials is the only way to make information available to answer these controversial questions.

It is crucial that any form of advertisement be reviewed and approved by your IRB/IEC. Articles for newspapers or scripts for radio and television also require review. Recruitment through the Internet is beginning to have a role and might be very effective in some therapeutic areas. Whatever method you use, it is important to be prepared for the telephone calls by making sure you have staff informed and available to handle them.

Many sponsors want to be proactive in subject recruitment and contract a central subject recruitment company to orchestrate the process. This has been very successful in many clinical trials. In most studies (such as osteoporosis), people are more likely to respond to an advertisement if the site's identity is visible and recognised as part of their community. Medical care (and study participation) is a very personal matter and better provided at a local level.

Make sure your study design takes into account the accrual rate of subjects into the study in realistic terms. Often, the response to requests for subjects to take part in a long, complex trial is small. For example, one trial recently approached almost 8000 women through primary care. Less than 6% were randomized.[1] Another study in primary care aimed to recruit almost 200 women per month. In reality, only 500 were recruited in the first 12 months because the logistics of the trial had not been thought through at the design stage. The response in primary care will also depend on the social and ethnic mix of the area surrounding the clinic.

If accrual into the trial is slower than expected, this can cause financial problems, with grant funding running out before the trial is completed. It can also be very demoralizing for study staff and, if the trial fails to recruit the full number of subjects, means that the study becomes invalid and, therefore, unethical.

In multicentre trials, it is important to monitor accrual at each centre. It might be possible to stop or suspend the study in centres that do not have sufficient recruitment, while maintaining the overall rate of accrual at other centres.

Some sponsors encourage the practice of competitive recruitment between centres in a multicentre trial. This can be unethical if there are large sums of money to be paid to centres on a per-subject basis. It will encourage unethical recruitment methods and can result in subjects being entered on the study who do not adequately meet all the inclusion criteria.

6.5. Subject Retention and Compliance

Once you have secured the subject's consent to participate in the research study, you now have the challenge of subject retention and follow-up. All your efforts in recruitment are wasted if you fail to retain subjects for the duration of the clinical trail. This is a greater challenge in studies of osteoporosis because these studies have to be designed with several years of follow-up. Of crucial importance here is the personal touch and caring of the investigator and the study staff.

Many of the women recruited for clinical trials in osteoporosis are still fully employed and could have difficulty in allowing for the extra time involved for study participation. Make every effort to schedule their follow-up appointments at times convenient to them and keep to the schedule as much as possible. Ensure that access to the clinics is convenient, avoiding the bustle of the regular out-patient clinic.

Contact between clinic visits is often important in long studies. Write to those involved with newsletters, or telephone to ensure that they are still taking the study medication. If designing study documentation and newsletters, make sure there is an identity for the trial. This includes using a logo on all study documentation and a short title for the trial that is easy to remember.

Sponsors of many trials will provide small token gifts to the study subjects, such as mugs and bags with study logos. Remember those special occasions, such as birthdays and Christmas, by sending a card. Subjects not only enjoy these extra items but also are reminded that they have an important role in a very important trial. Their sense of belonging will help maintain their compliance. An inconvenience allowance or travelling expenses can be paid. Make sure, however, you obtain IRB/IEC approval for such gifts and allowances because they might be viewed as coercive.

It is important to maintain the subjects' compliance with the study treatment. The shorter the study period and the simpler the intervention, the easier it will be to maintain compliance. However, some osteoporosis trials could involve follow-up of 10 years. Giving good information about osteoporosis and the study drug at the start of the study will ensure subjects recruited to the trial will be well informed and well motivated. This can include information on diet and exercise and general health issues. Compliance falls if the time between clinic visits is too long. Keeping up the contact with subjects through newsletters, telephone calls,

and clinic visits will all help, particularly if there is a good relationship between study staff and subjects.

There are a number of ways of monitoring compliance. Laboratory tests can be used to monitor levels of the study drug or active metabolites in the blood. In some studies, very low doses of inactive, slow-turnover chemical markers have been added to treatments to enable easy detection by blood sampling. Other studies have used electronic dispensers for the study drug. This is expensive and does not ensure that the drugs are taken once dispensed. Alternatively, subjects can be asked to return all unused drugs when attending for clinic visits. Again, it is not possible to work out whether or not the drug has been taken. The best method is to ensure that there is a good relationship between study staff and participants. Make sure that participants are well motivated by providing a flow of information about the trial and about the lifestyle issues associated with osteoporosis, and ensure that side effects and adverse events (AEs) are discussed at clinic visits. If subjects do not comply with the study medication, it could mean that they will have to be withdrawn from the study and offered alternative treatment. If noncompliance is a serious problem, there are significant implications for the design of the study and the analysis of the results.

6.6. Monitoring

The sponsor of clinical trials is obliged to have oversight of the project activities at the local level and so it is not uncommon to expect a visit from the CRA every 4 to 6 weeks during the progress of the study. It is important to make space and time available for these individuals, and this does not mean a draughty hallway where they have no privacy! The purpose of monitoring a clinical trial is to ensure that the trial is being carried out in compliance with the protocol and good clinical practice (GCP). It is important to ensure that the trial data are accurate, complete, and can be traced back to the source documentation, including the subjects' medical records. Monitoring visits should also ensure that the rights and well-being of the trial subjects are protected. This might include audit of the informed consent documentation and interviewing some trial subjects to ensure the quality of the consenting process.

CRAs are appointed by the sponsor and should have appropriate training and scientific and clinical knowledge to monitor the trial properly. The amount of monitoring will depend on the complexity of the trial. In some circumstances, by ensuring that there are good written procedures, the investigators receive good training, and there are frequent review meetings, it might be possible to reduce the amount of on-site monitoring and carry out monitoring centrally. There is a comprehensive list of the CRA's responsibilities in monitoring the trial in the International Committee on Harmonization ICH guidelines on GCP,[2] which includes ensuring that the investigator, their staff, and the local facilities are adequate for the trial, only eligible subjects are recruited for the trial, and trial documentation is accurate and up to date.

Don't forget that there are many qualified sites anxious to do clinical trials. Not having space available for the CRAs during the conduct of the trial can cause a site to lose the opportunity to participate in a study. It is important to remember that these monitors are "on the road" as much as 80% of their working week. Find them some desk space where they can shut the door, review the necessary documentation, and work in peace and quiet. If you can make their visit to your site pleasant and comfortable, you will find they will recommend your site for future trials.

At the end of a monitoring visit, the CRA is responsible for completing a monitoring report. This should include details of everything that was reviewed at the trial site, including accrual to the study. The report might include recommendations to improve the operation of the trial at the site, in order to meet the requirements of the ICH guidelines on GCP.

6.7. Study Files

Most sponsors provide a "regulatory document binder" to assist the sites in maintaining proper records for the study. These binders are helpful, but most experienced and committed study sites have their own system. However the files are kept, it is imperative that all the information is easily accessible when the monitor comes to visit. When maintaining these files, you must remember that 2 to 3 years after the study is closed, you might need to pull these files for an audit by the FDA or other regulatory authorities. The better job you do in keeping proper records during the conduct of the trial, the less painful the audit will be for you.

6.8. AE Reporting

AE monitoring is important, in order to establish the safety profile of an NME. In active comparator trials, for example, an NME might be designed to have an improved safety profile compared with existing treatments, even if it only demonstrates a treatment effect that is similar to an existing drug. An AE is any event that occurs but is not expected, e.g. unusual symptoms, drug reactions, or abnormal laboratory findings. It does not have to be caused by the study drug to require reporting. Adverse drug reactions (ADRs) are a subset of AEs for which a causal link to the NME can be established. The causal link is usually coded as follows:

1. Probably not related to the study drug
2. Possibly related to the study drug
3. Probably related to the study drug
4. Definitely related to the study drug.

An AE or ADR is considered serious if it results in death, is life threatening, requires the subject to be admitted to hospital or kept in hospital longer than expected, results in a significant disability, or is a congenital abnormality or birth defect.[3] Many drugs are known to have adverse reactions or side effects. If these are expected as part of the treatment, it might not be necessary to report them (e.g. the side effects of chemotherapy treatment are well known and do not require reporting in oncology-based clinical trials). The investigator's brochure will record the known ADRs that do or do not require reporting. There is standard information that it is necessary to report for each AE.[3] Serious adverse events (SAEs) or serious ADRs (SADRs) must be reported to the regulatory authorities within 7 calendar days by the sponsor, followed by a full report within 8 calendar days after the first notification—so the organization of the trial has to be slick and well oiled.

It might be necessary for the sponsor to break the blinding for an individual subject in the case of a SADR. The subject might have to be taken off the study and put on another standard treatment because of the SADR. It is important, however, that the local study staff who are recording clinical information or undertaking DXA do not know the outcome of the unblinding if the subject is to remain in the trial. Breaking the blinding will ensure that SADR reports are not filed for the NME, active comparator, or placebo unnecessarily. This will ensure that the investigator's brochure is accurate when it is updated with new safety information as it becomes available.

AE reports should be reviewed by the Data Monitoring Committee (see Chapter 5). A summary of AE reports should be sent to investigators in each centre. The investigator should review the AE reports and consider the impact of the reports on the trial at their centre. It is unlikely in osteoporosis trials that it will be necessary to stop a trial at any one centre in a multicentre trial because of local issues arising from AE reports, but it does happen in other disciplines. The investigator should send a copy of the AE reports, together with their assessment of the local impact, to the IRB/IEC for information.

6.9. Summary

Good organization and attention to detail is vital for the successful management of a clinical trial. The amount of paperwork involved is enormous[2] and can seem unending. Getting it right at every stage will smooth the path to regulatory approval of your NME. It is particularly important that good relations are established between the study subjects and the staff at the local centres, and between the staff and the CRA who will monitor them. This will ensure that there is adequate recruitment, good compliance with treatment, and good communication between the sponsor and the investigators. It might seem a great effort, but it will be worth it in the long run.

Remember to send the results of the study back to participating centres. It is easy, once a study has closed, to forget those who have been most closely involved in the day-to-day management of your study. A summary report for the IRB/IEC is essential, and a summary in lay terms that can be sent out to the research subjects themselves is important as a "thank you" for taking part.

References

1. Grainge, M.J., Coupland, C.A., Cliffe, S.J. *et al*. (1998). Cigarette smoking, alcohol and caffeine consumption, and bone mineral density in postmenopausal women. The Nottingham EPIC Study Group. *Osteoporos Int* **8:**355–63.
2. Expert Working Group (Efficacy) (1997). International Conference on Harmonization: Guidelines for Good Clinical Practice. *Federal Register* **62:**25691–709. http://www.fda.gov/cder/guidance/959fnl.pdf accessed 13/11/06
3. Expert Working Group (Efficacy) (1995). International Conference on Harmonization: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting. *Federal Register* **60:**11284–7. http://www.fda.gov/cder/guidance/iche2a.pdf accessed 20/11/06

7

Clinical Trial Quality Assurance

Susan A. Earnshaw

7.1. Introduction

Clinical trial quality assurance (QA) covers a whole range of issues in bone mineral density (BMD) assessment. In addition to the regular review of quality control procedures, it also encompasses adequate subject preparation, technologist preparation, observance of rigorous scanning and analysis protocols, and the use of regular, organized data shipments to dual-energy X-ray absorptiometry (DXA) QA centres. All members of the research team should regard these aspects as necessary and important considerations both at the time of clinical trial set-up and throughout the duration of the trial. It is only by implementing and maintaining clear, organized protocols and procedures that accurate and reproducible BMD data will be obtained. Although it is primarily the densitometer technologist who will be responsible for ensuring subject data is adequately acquired, principal investigators and study monitors have a role in ensuring that the technologist is adequately prepared and included as a member of the clinical trial team. The preparation will include both relevant training and trial documentation.

In this chapter, we review the preparation required for subjects and technologists, general scanning principles, and protocols, and then discuss lumbar spine, proximal femur, and forearm DXA scans in further detail. Lateral lumbar spine DXA was developed as a method for assessing the metabolically active trabecular vertebral body by excluding the predominantly cortical posterior elements of the vertebrae. However, precision has been demonstrated to be very poor;[1,2] thus, the technique is not widely used. It is, therefore, not discussed in this chapter. Whole-body DXA is another technique that is not widely available. It is used for noninvasive assessment of body composition and is more readily available than traditional methods. However, it is yet to be accepted as a "gold standard" technique.[3] Morphometric X-ray absorptiometry (MXA) is a technique developed for the measurement of vertebral body heights in clinical trials monitoring fracture incidence or prevalence and will be briefly discussed. Newer methods of identifying vertebral fractures, such as

lateral vertebral assessment (LVA) or vertebral fracture assessment (VFA), will also be discussed.

We have not discussed each manufacturer in today's market, owing to the range of equipment available, particularly the vast range of peripheral DXA and ultrasound systems. However, techniques and principles are broadly similar across the entire field of bone densitometry. This chapter is restricted to those examinations carried out on axial bone densitometers.

The integrity of both subject and bone densitometer quality control data must be ensured at all stages of the clinical trial. This entails regular data archiving, record keeping, with the associated study paperwork, and preparations for the auditors' visit, which are discussed in the final sections of this chapter.

7.2. Preparation Before Bone Mineral Assessment

Adequate preparation of bone densitometry equipment, the subject, and the technologist is essential if a diagnostic scan is to be achieved. Only once all routine equipment quality control (QC) checks are deemed within acceptable limits should trial QC scans be carried out and subjects scanned.

7.2.1. Bone Densitometry Equipment Daily QC

Ensure that daily QC of the scanner is carried out at the start of each day. This includes use of both the routine calibration block and the aluminium spine phantom for Lunar (GE Healthcare, Madison WI, USA) densitometers. For Hologic (Bedford MA, USA) densitometers, this includes use of the Hologic spine phantom. Review the results and ensure that the BMD lies within the predetermined limits, as discussed in Chapter 8. Should the results lie outside the limits, the initial step is to repeat the QC scans, ensuring that the correct procedures are followed. If this second scan also fails, contact the manufacturer or their distributor for further advice. Do not scan any subjects until a valid QC result is obtained, because subjects' scan data will be invalid.

7.2.2. Before the Subject's Arrival

Subjects should receive details of the scan procedure before they attend their appointment. For a clinical trial, this should be included in the trial-specific subject information letter, plus any additional subject letters or leaflets routinely distributed by the scan department. Many subjects are worried that the scan involves them lying in a long, dark tunnel for 30 minutes [as in a magnetic resonance imaging (MRI) scan]. Fears such as these can be allayed by ensuring each subject receives information in advance that includes a picture of a subject lying on a bone densitometer with a concise description of the procedure and the duration of both the individual scans and the total appointment length. The subject should also be advised on whether they will need to undress (if so, they might prefer to bring

their own dressing gown), that they will not require an injection, and on whether X-rays are involved in the procedure.

Consideration should be given to obtaining, in advance, copies of the subject's previous radiographic examinations and reports, particularly if they were at the same hospital and are thus readily available. Previous scans and radiographs should always be available for those subjects attending for follow-up trial scans. These should be reviewed before the subject enters the scanning room so that problems can be identified and discussed with other staff if necessary.

If the subject has had a previous DXA scan, it is important to review the scan printouts. Check which scan mode was previously used, whether there were any problems or modifications with the subject positioning and scan acquisition, and whether the analysis was carried out according to the manufacturer's guidelines. For future reference, it is helpful to note any modifications or difficulties experienced during positioning, scanning, or analysis of the scan printouts.

If the subject has had no previous BMD scans but, as is often the case, has had radiographs of the lumbar spine, these will be helpful in assessing the anatomy of the lumbar spine and determining whether positioning and analysis are likely to present difficulties. Previous radiographs of the abdomen will provide useful information on lumbar spinal anatomy, whereas pelvic radiographs will demonstrate proximal femur anatomy.

7.2.3. Subject Preparation

There are several contraindications to BMD measurements, as follows:

1. If the subject does not meet the trial inclusion and exclusion criteria, this could be decided in the clinic before the subject attends for the DXA scan. If the study is dependent on subjects having BMD values within specific limits, this will be determined by confirmation from the DXA QA centre.
2. Because the technique involves the use of ionizing radiation, adequate measures should be taken to ensure that any woman of childbearing age who is, or might be, pregnant does not undergo a scan.
3. Recent barium examinations will result in barium artefacts overlying either the bone or the soft tissue of the lumbar spine. Following the barium examination, a period of 2 or 3 weeks could be required before the lumbar region can be clearly assessed.
4. Recent nuclear medicine examinations; depending on the isotope used, this can vary between 24 hours and several days.
5. The presence of metal implants, such as a total-hip-replacement prosthesis. In many cases, it will be possible to scan the contralateral proximal femur, but occasionally the subject will have bilateral hip replacements.
6. Subjects who are claustrophobic might be unable to tolerate a quantitative computed tomography (QCT) BMD scan.
7. Subjects who are in considerable pain or suffering from severe scoliosis might be unable to lie in the correct position for the duration of the scan.

When the subject attends for a scan, they could be apprehensive or nervous about the procedure. A detailed explanation of the procedure and what is required of the subject should resolve their fears and generally ensures better cooperation during positioning and scanning, such that accurate consistent subject positioning is achieved.

It is preferable to ask subjects to remove their own clothing and wear a hospital gown for an axial DXA scan. This is unpopular at some centres, because it can be considered inconvenient or embarrassing for some subjects. It is time-consuming, which could be a problem at busy centres with a high subject throughput, and requires additional changing facilities for subjects. However, this ensures that all artefacts, such as zips, buttons, belts, and coins, are removed from the scan field. It is considerably more time-consuming and embarrassing (for the subject) to ask a subject halfway through the scanning procedure to unzip a pair of trousers and remove the zip from the scan field while still lying on the scanner table! The other alternative is to ask subjects to attend the clinic wearing a sweat suit or jogging suit without metal fastenings; however, this might not be acceptable for elderly people.

The subject's height and weight should be accurately measured and recorded on the request form, for entry in the subject's biography. If this is a repeat visit, whether the subject has gained or lost considerable amounts of height and/or weight should be observed. This should be recorded on the scan printout for the attention of the requesting clinician or clinical trial DXA QA centre. Height loss could indicate that the subject has suffered a vertebral fracture, and it might be necessary to exclude fractured vertebrae from the analysis. Studies have shown excessive weight increase to be associated with a decrease in BMD.[4]

The subject's height and weight are also used by Lunar DXA software to calculate a subject's body-mass index (BMI), which is, in turn, used to automatically set the scan mode for each scan region. Additionally, normal reference ranges might include a weight correction factor that is applied to a specific range of subject weights. Lunar software applies a weight correction to reference data for subjects between 25 kg and 100 kg.[5,6]

Once the subject enters the scanner room and begins to relax, they might talk about aspects of their medical history that are not on the request form. Listen for comments such as those reflecting a loss of height since they were younger (possibility of vertebral fractures or scoliosis) or previous fractures of the hip or wrist (scan the contralateral side). They might be a long-term asthma sufferer (i.e. the subject might be unable to hold their breath during a scan or lie flat). Always check with the subject whether they have had previous fractures or operations at the scan sites, to avoid unnecessary scanning of fractures or metal implants.

7.2.4. Establishing a Clinical Trial Database

All bone densitometers store electronic information (both subject biographies and scan data) in a database. The manufacturer establishes the database format so that either there is a single database containing all subjects together (both routine clinical subjects and trial subjects) or there are separate databases that can be set

up for each individual clinical trial. There are advantages and disadvantages to both methods, but the DXA QA centre will specify which method they require in the clinical trial protocol.

The format of the database and the details required in subjects' biographies must be established before any subjects attend for a scan. Considerable time and thought should be given to data-entry details and the database structure because, if set up incorrectly, it might be difficult to read or export data in the database files. Clinical trials in separate databases are the clearest to identify because only subjects in the trial should be included in the database. Because subjects in the database will be anonymous, a separate trial-subject record book, containing details of the subject's full name, initials, and subject number, will provide further identification should the subject subsequently return for a scan once the trial is completed.

If all trials are included in the same database, great care must be taken to ensure that subjects' biography information is entered correctly for each trial. Additionally, a protocol for standardizing routine clinical subjects' biographies should be established to ensure that their data can be readily identified in the database files.

7.2.5. Subjects' Biography Information

Check with the subject that their biographical information recorded on the request card is correct before entering their details into the scanner biography page. There is a significant error rate for incorrect data entry (e.g. misspelt name or incorrect date of birth). An incorrect name entry could be difficult to change once entered into the computer. If the incorrect year of birth is entered, the Z-score results will be incorrect for all scans, because Z-score calculation relies on a comparison with an age-matched peer group. If the subject has attended for a previous scan, the original biography details for that subject must be used. Creating a duplicate biography will prevent computer calculation of rate of change for the scan results and will cause confusion at subsequent follow-up visits.

For the majority of clinical trials, the subjects' biographies must be anonymous to conform to subject-confidentiality regulations; therefore, ensure that the correct clinical trial identifiers and subject randomization numbers are entered in each subject's biography. If the scan QA centre has requested a specific format for the biography details, this must be adhered to because the same format will be used at all scan sites for a particular clinical trial. This ensures that when the scan data is sent in electronic format to the QA centre it can be readily merged into the trial database for data review.

The QA centre should provide details of the biography entry in a study-specific manual. Ensure that you have received and read a copy of the manual before scanning any subjects, and that you understand how to keep scans anonymous. It is helpful to condense the biography and scanning information from the study manual onto a single side of an A4-size sheet of paper for a quick reference guide (Figure 7.1), to avoid having to shuffle through a cumbersome manual each time a subject is scanned.

Figure 7.1 is an example of a subject biography on the Hologic QDR 2000. It demonstrates that despite the subject's biography being anonymous the operator can still identify it. In this example, the subject's full name is replaced by their initials (first name, middle name, and surname). The "Pat ID" box contains the subject's specific trial number. The "Scan Code" box contains the technologist's identification number. The "Zip" box contains the subject's visit number. The "Ref MD" box contains the study site number. "Date of Birth", "Sex", "Weight", "Height", and "Ethnic" boxes should always contain subject-specific information. Inclusion of the correct information ensures that both subject and scan data can be

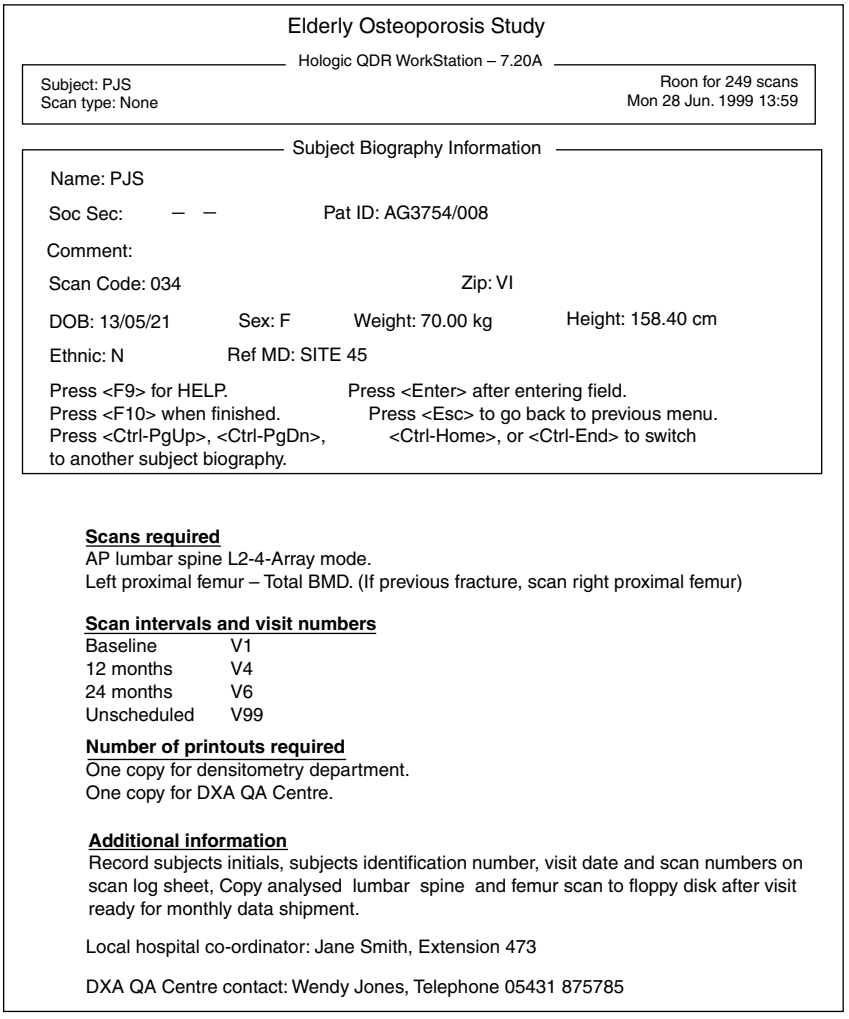


Figure 7.1. Quick reference guide to subject biography information required for a clinical trial on the Hologic QDR 2000 bone densitometer.

readily extracted from the scan computer database and transferred to another database (as required for regular data shipments) for subsequent analysis. Included below the biography in Figure 7.1 is further information relevant to the clinical trial and useful contact telephone numbers.

7.3. DXA Scan Acquisition and Analysis: A General Review

This section reviews the principles of reproducible subject positioning that are applicable to all types of DXA scanning. It should not replace manufacturers' procedures but should serve as an additional method of improving techniques. Specific regions will be discussed fully in subsequent sections.

7.3.1. Subject Positioning

Before commencing positioning for each scan, explain to the subject what is required of them and how long the scan will take. If they are required to hold their breath in a specific manner during a scan, practise the procedure with the subject until they feel confident. A nervous, fidgeting subject will cause blurring on the image, necessitating a repeat scan.

Scan technologists should be familiar with positioning techniques for the different types of scans, so that the subject is positioned correctly for the first scan. This helps in keeping the examination time as short as possible and avoiding unnecessary X-ray exposure to the subject (even if the scanner is a low-radiation-dose pencil-beam scanner). Previous scan images and clinical information on the referral form (e.g. scoliosis) will aid positioning. It is preferable to spend time positioning the subject correctly, following a standard set of guidelines, before starting to scan rather than to move the subject into position once the image appears on the screen. Positioning of the subject at follow-up visits should mimic that of the initial visit, particularly if modifications were made to the positioning at that initial visit, because this will aid comparative analysis. If modification of positioning protocols is required (e.g. the subject is unable to raise legs onto a foam positioning block) record this information on the scan printout for future reference and in the study documentation for the attention of the DXA QA centre.

Technologists with a background in diagnostic radiography will usually be suitably experienced in positioning subjects, but it is advisable for those from other backgrounds to gain further experience in addition to the manufacturer's training. The ideal training for nonradiographic technologists would include several days initially spent in a Radiology Department, reviewing the positioning of both subjects without significant positioning problems and those with severe positioning problems, such as considerable pain, scoliosis, arthritis, etc. Discussion of radiographs would aid appreciation of the variety of normal anatomical appearances and how to improve an incorrectly acquired image. Such background experience ensures that a rapid evaluation of positioning can be made in the early stages of scan acquisition, enabling the scan to be aborted and the subject repositioned with the minimum of additional X-ray exposure.

Centres with several technologists should ensure that all staff adhere to the same protocols. Protocols of positioning, scanning, and analysis should be readily available, and time should be dedicated to regularly reviewing techniques as a group, with a member of staff acting as the subject so that actual positioning techniques can be practised and discussed. Actually scanning members of staff, however, is not recommended and is illegal in many countries because of the radiation dose involved. However, positioning other members of staff, up to the point of starting the X-rays, is a good first training exercise. An experienced technologist can also mimic the "difficult subject".

Following suitable training on positioning subjects, the recommendations in the manufacturer's operating manual should be followed for the majority of subjects. These are the positioning protocols used to scan the manufacturer's reference populations and thereby derive reference ranges, so consistency with positioning will ensure reference data can be applied to each subject.

7.3.2. Image Review During Scan Acquisition

DXA scans should be viewed throughout the entire acquisition process so that problems can be quickly identified and rectified. At the initiation of the scan, the points to check are that the start position is correct, there are no obvious artefacts, and the subject's positioning is correct. As the scan progresses, the image should be constantly reviewed to ensure correct positioning, adequate soft tissue inclusion, and to prevent blurring and artefacts. The scan should be quickly stopped if faults are identified so that they can be rectified immediately and to avoid unnecessary irradiation of the subject.

7.3.3. Scan Analysis

Before proceeding with the analysis, confirm that this is required as part of the trial. Many studies now require the scan to be submitted to the QA centre unanalyzed for the following reasons:

1. To improve precision, a limited number of technologists review and analyse the data. The DXA QA centre undertakes this analysis.
2. To assist in maintaining a blind, the analysis is not performed at the site. Subjects receiving (and responding to) treatment, if it is efficacious, will show an increase in BMD that will alert the technologist to the treatment. This is in conflict with the logic and requirements for a double-blind study.

However, it is still important to review the image carefully to ensure that the image can be analysed. The best method for this is to analyse the scan as normal but not to save the analysis.

Each manufacturer has a specific protocol for defining the region of interest (ROI), which, as with the subject's positioning, will also have been used to derive the normal reference ranges. Analysis of a scan must follow the manufacturer's guidelines so that comparisons with reference data are valid.

Scan analysis can be carried out by either analysing or comparing the scan image. The initial procedure should be the adjustment of the image density and contrast such that bony anatomy can be clearly identified.

7.3.3.1. Analysis

This method involves input from the technologist to define the ROIs on the recently acquired scan image according to the manufacturer's guidelines. Care should be taken to ensure that all relevant areas are included. The next step in the process is the automated detection of bone and soft tissue edges by the software programme. These outlines should always be reviewed. In cases of obvious error, however, the advice of different manufacturers varies as to the extent of changes that should be made. The final stage is the computer calculation of BMD and display of results on the screen.

7.3.3.2. Compare Analysis

This method is used at subsequent follow-up visits to ensure that the ROI is identical to that of the subject's first (baseline) scan. It requires the technologist to use the computer software to overlay the initial image, referred to as the baseline image, onto the most recent image. Always use the baseline ROI, even if the scan was acquired several years ago, to permit long-term bone mass changes to be calculated for a consistent ROI. Assuming there is a good match between baseline and follow-up scans, it should not be necessary to modify a comparison analysis. If modification is required (owing to subsequent deformity since the baseline scan), contact the clinical trial DXA QA centre for relevant guidance.

Each scan image should be analysed or compared and reviewed before proceeding to the next scan so that the scan can be repeated if necessary, either with the subject in the same position or with adjustment to the subject's position if this was inadequate. The features to check vary depending on the scan and will be discussed later within the relevant sections. For a clinical trial, you might be limited in the type of modifications you can make, for example exclusion of osteophytes. You might instead be required to note such problems on a scan log sheet or data action sheet (the name varies depending on the QA centre) for review by the QA centre.

Clinical trials might require inclusion or exclusion of specific anatomy that differs from a department's standard clinical protocol, that is inclusion of the first to fourth lumbar vertebrae (L1 to L4) rather than the second to fourth lumbar vertebrae (L2 to L4). Ensure that all technologists are aware of the trial-specific analysis criteria if they differ from routine clinical protocols. Record details on the "quick reference guide", as shown in Figure 7.1.

7.3.4. Printed Image Review

The printouts from each analysis will be the paper record of the subject's scan. Review the printouts, ensuring that the biographical information is correct and

that the correct analysis is printed for future reference. Any modifications during the positioning, scanning, or analysis should be noted on the printout that will stay within the department and the study documentation for the attention of the DXA QA centre. Not all studies require a paper record of the scan, an electronic copy might be sufficient; check with the specific study QA manual. This information can be recorded on the "quick reference guide", as shown in Figure 7.1. If required, the scan should now be copied to the appropriate back-up data medium (e.g. CD or optical disc) for the clinical trial data shipment.

7.3.5. Unanalysed Scans

If the QA centre requires unanalysed scan data, complete the image review described in each scan section but do not save any analysed scan data. Copy the unanalysed scan onto the back-up data medium and label it according to the study guidelines.

7.4. Lumbar Spine DXA Scan

The lumbar spine is scanned in either the postero-anterior (PA) or the antero-posterior (AP) direction, depending on the type of scanner. For all scanners, the subject lies supine on the mattress. Throughout this section, this type of scan will be referred to as the AP lumbar spine scan.

7.4.1. Subject Positioning

The subject should be positioned according to the manufacturer's guidelines. Time spent initially ensuring that the subject is lying straight (Figure 7.2) will produce a lumbar spine image that is straight and lies in the middle of the scan field. On an array-beam DXA densitometer, it is essential for subjects to be in the centre of the scan image so that the central ray of the beam passes through the midline of the lumbar spine at the baseline and all subsequent visits. Owing to the divergent X-ray beam, the projected area of the lumbar spine will vary depending on whether the central ray passes through the centre of the lumbar spine, transverse processes, soft tissue adjacent to the spine, etc. Thus, care with alignment at both the baseline visit and all subsequent visits will ensure a similar projected area of the lumbar vertebral bodies at all visits. The subject's spine should be straight in all planes:

- Not rotated around the cranio-caudal axis of the spine
- Not lying diagonally across the scan field.

7.4.1.1. Correction of Cranio-Caudal Rotation

The initial check is to correct rotation around the cranio-caudal axis of the spine, which is the harder of the two rotational problems to correct, particularly

in elderly subjects because it is inherent in scoliosis. It is not possible to correct rotation owing to scoliosis. To correct for this problem in subjects unaffected by scoliosis, ensure that the anterior superior iliac spines (ASISs) are equidistant from the mattress surface. Simultaneously place each thumb on the ASIS and your finger tips on the table top, and then judge whether either thumb (i.e. the ASIS) is higher/lower on one side of the subject than the other. If one ASIS is further from the mattress than the other, this indicates the pelvis, and hence also the lumbar spine, is rotated. This can be corrected by asking the subject to raise their pelvis up a couple of centimetres from the mattress and then immediately guiding the pelvis back down onto the mattress, keeping your thumbs on the ASIS to ensure they are now equidistant from the mattress. An additional guide is to ensure that both shoulders are equidistant from the mattress. These two positioning guides can be readily practised on work colleagues, without having to scan anyone.

7.4.1.2. Correction of Diagonal Rotation

To ensure the subject is not lying diagonally across the scan field, align the subject with the longitudinal centre line on the mattress such that it "passes" up the centre of the subject's body (Figure 7.2). Standing at the top or bottom of the table should give you the best position for judging whether the subject is lying straight. To correct the subject's position, align the laser positioning light so that it is equidistant between the ASISs, which can be readily palpated on the anterior aspect of the pelvis on all but the largest of subjects. Tracking the laser up the subject's abdomen, it should be seen to lie over the lower end of the subject's sternum, the xiphisternum. It is worth noting that, if the subject's shoulders are equidistant from the centre and their feet are placed slightly apart either side of the centre line, the subject should "feel" that they are lying straight on the mattress. If, on questioning the subject, they respond that they do not feel straight, you will probably find when you start scanning that the subject has some degree of scoliosis.

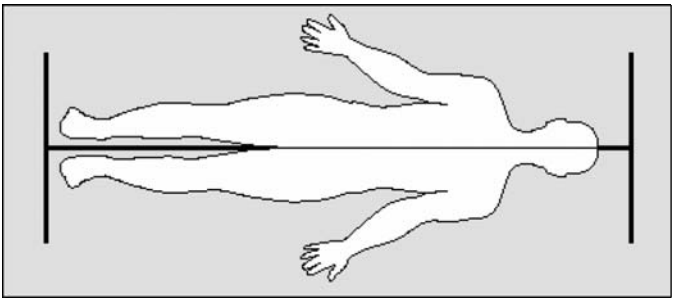


Figure 7.2. Initial subject positioning with the midline of the subject aligned to the centre line of the mattress.

7.4.1.3. Leg Support Block

The final part of positioning the subject should always be placement of the leg support block to reduce the lumbar lordosis (Figure 7.3). Once this block is in position, it is difficult to do any of the positioning described above. Placement of the leg support block varies between manufacturers. There are three heights at which the block can potentially be used; therefore, check the manufacturer's operating manual because one block height will have been used to acquire reference data.[5–7] If the subject is uncomfortable with the leg support block (e.g. very short legs, previous hip or knee surgery, or painful hip or knee pathology), lower the block height until they feel comfortable and record the height used on the scan printout.

Ensure that the subject's hands are outside the scan field and that the subject is comfortable with the pillow provided. Generally, one pillow should be adequate; however, elderly subjects with a thoracic kyphosis might require two, or even three, pillows. The same number of pillows should be used at all visits and recorded on the printouts if it is different to the standard one pillow.

7.4.2. Scanner Preparation

Once the subject is positioned satisfactorily, bring the laser positioning light to the correct start level approximately 2–4 cm below the pelvic crest, in the subject's

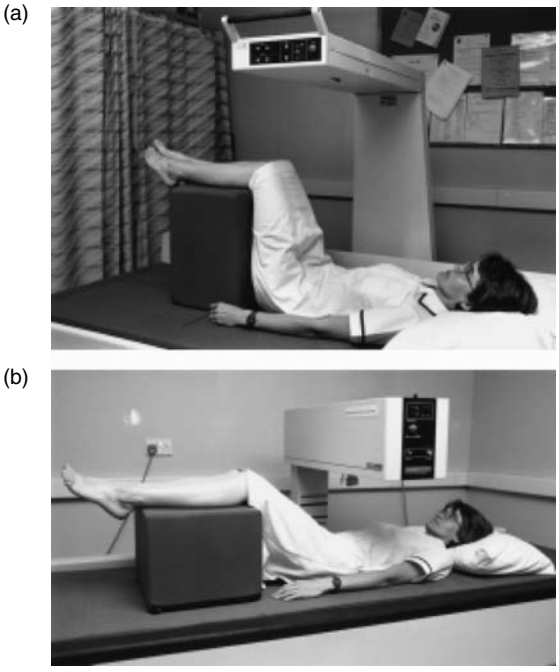


Figure 7.3. Subject positioned with the leg support block under the lower half of the legs. (a) On the Hologic bone densitometers the knees must be bent to 90°. (b) On the Lunar bone densitometers the knees are bent to 45°.

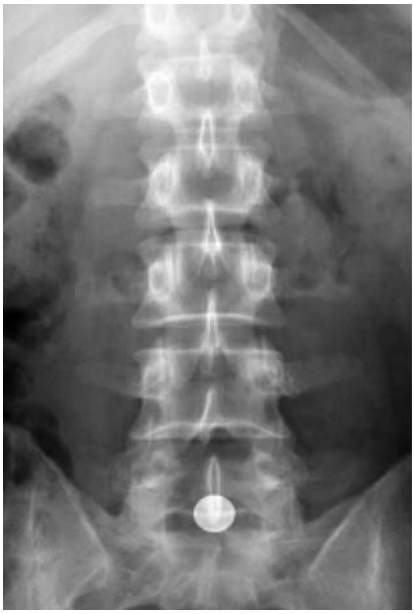


Figure 7.4. Start point for a lumbar spine DXA scan (indicated by the white dot), aligned with the subject's midline, approximately 2–4 cm below the pelvic crest, so that the scan commences in the lower portion of L5.

midline, so that the scan commences in the lower portion of the fifth lumbar vertebra (L5) (Figure 7.4). Explain the duration of the scan and any specific breathing requirements during the scan to the subject, and that they are required to remain still for the entire procedure. If a compression band is required, its purpose and duration of use should be explained to the subject, at the same time inquiring whether the subject has undergone recent abdominal surgery, in which case it should not be used. Reassure the subject that, although the band will feel very tight, it should not be painful. To ensure the compression band covers the entire region of the lumbar spine, it should be positioned at the lower border level with the laser positioning light. As the band is tightened, ensure the subject is not in any pain.

7.4.3. Scan Mode Selection

Select the appropriate scan mode and parameters according to the manufacturer's operating manual or the clinical trial specifications. For subjects returning for follow-up scans, it is essential that the same scan mode and parameters are selected as for the baseline visit.

A scout scan mode is available on some DXA scanners, which is useful for reviewing a subject's position in cases when a scoliosis is suspected or it is uncertain that the subject is in the centre of the scanning field. Such a scan has a lower radiation dose than the routine scan and, typically, takes only a couple of seconds.

However, the scout scan should not be used routinely in place of good positioning technique.

7.4.4. Image Review During Scanning

As the lumbar spine image begins to appear on the screen, check the following points:

1. Is the lumbar spine in the middle of the image?
2. Is the start level correct, that is middle or base of L5, with a small amount of pelvic crest in the lower corners of the image?
3. Is there any subject or lumbar spinal rotation?
4. Are there any overlying artefacts that could be removed (e.g. buttons or safety pins)?

Quickly stop the scan if necessary and either reposition the subject or the scanner arm, or investigate any artefacts.

As the scan progresses, check the following points:

1. Are there equal amounts of soft tissue on each side of the lumbar spine?
2. Is there any subject or lumbar spine rotation?
3. Has the subject moved during the scan?
4. Is the breathing routine adequate?
5. Is the compression adequate?
6. Are there any overlying artefacts that could be removed (e.g. buttons or safety pins)?
7. Are there any overlying artefacts that cannot be removed (e.g. bowel gas shadows or calcified aorta). These artefacts might require modification of the scan analysis or recording on the printouts and study documentation.

Stop the scan and make any necessary changes. Continue scanning until the lowest pair of ribs is clearly visualized. Figure 7.5 illustrates correctly acquired DXA scans of the lumbar spine.

7.4.5. Image Analysis/Comparison

On completion of the lumbar spine scan, immediately review the image and analyse the scan before moving the subject, so that if a repeat is required in the same position, this can be readily acquired. Examples of incorrectly acquired lumbar spine scans and abnormal appearances are shown in Figure 7.6. The review should include checking the points listed in Section 7.4.4, and the following additional points:

1. Optimize the image density and contrast to ensure that all the bony anatomy is clearly identifiable. Figure 7.7 illustrates a subject with very small ribs attached to the 12th tharacic vertebrae (T12). In a poorly optimized image, the T12 ribs cannot be seen and this could be mistaken for a subject with six lumbar vertebrae. Adjusting the density and contrast reveals small ribs either side of T12.

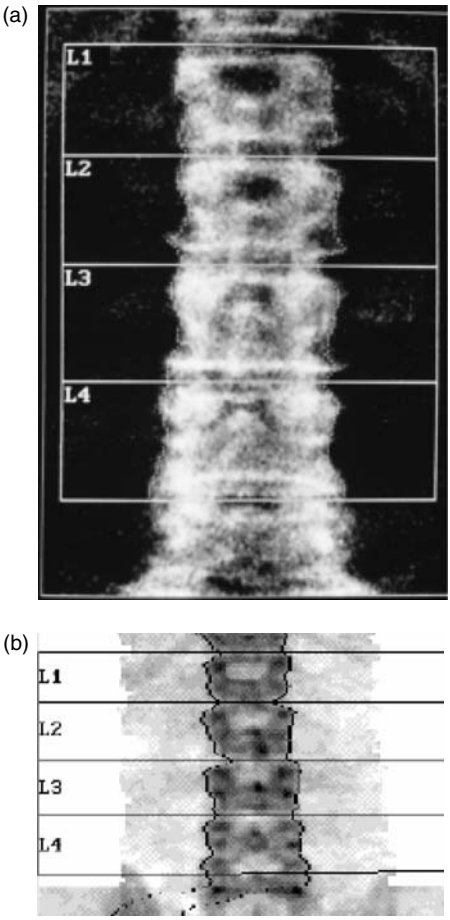


Figure 7.5. Correctly acquired DXA scans of the lumbar spine. (a) Hologic QDR 2000. (b) Lunar DPX-L. The lumbar spine is straight and lies in the middle of the image. Five lumbar vertebrae are clearly seen with no overlying artefacts. A small amount of pelvic crest is seen in both lower corners of the images and the lower part of T12 plus the T12 ribs are clearly identifiable at the top of the image.

2. Is subject positioning similar to the baseline scan?

3. Are there sufficient landmarks to identify L5 and L1? A small percentage of subjects have an abnormal number of lumbar vertebrae (six or four). Inclusion of bony landmarks (T12 ribs and the pelvic crests) aids identification, in addition to review of previous scans, radiographs, or reports.

4. Can each vertebral level be correctly identified, that is L1 to L5? Subjects with poor spinal anatomy owing to scoliosis, vertebral collapse, or advanced osteoarthritis (OA) might have poorly visualized or nonexistent intervertebral spaces. Review previous scans, radiographs, or reports (as above) and consider whether it is appropriate to delete any vertebrae from the analysis. Vertebral

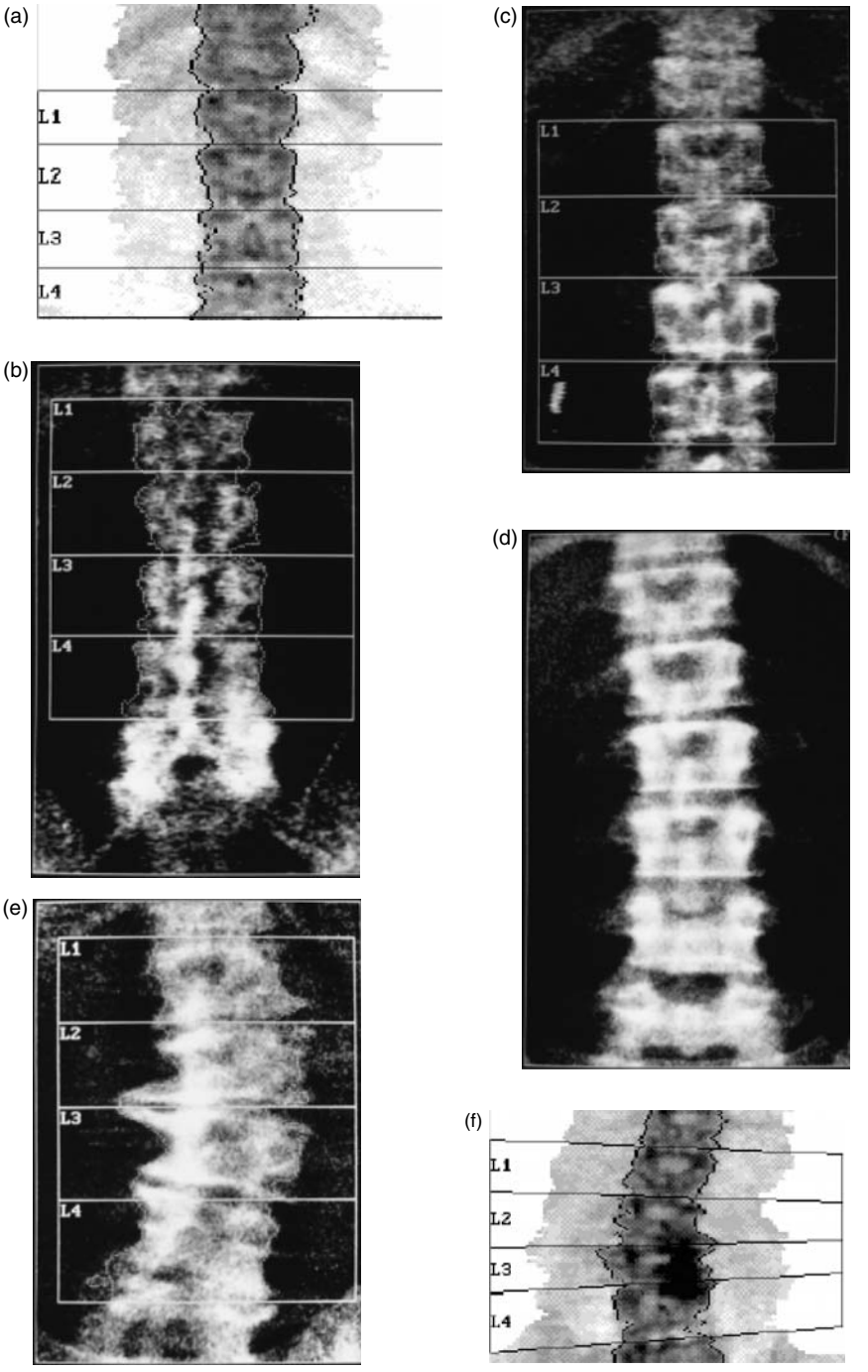


Figure 7.6. (Continued)

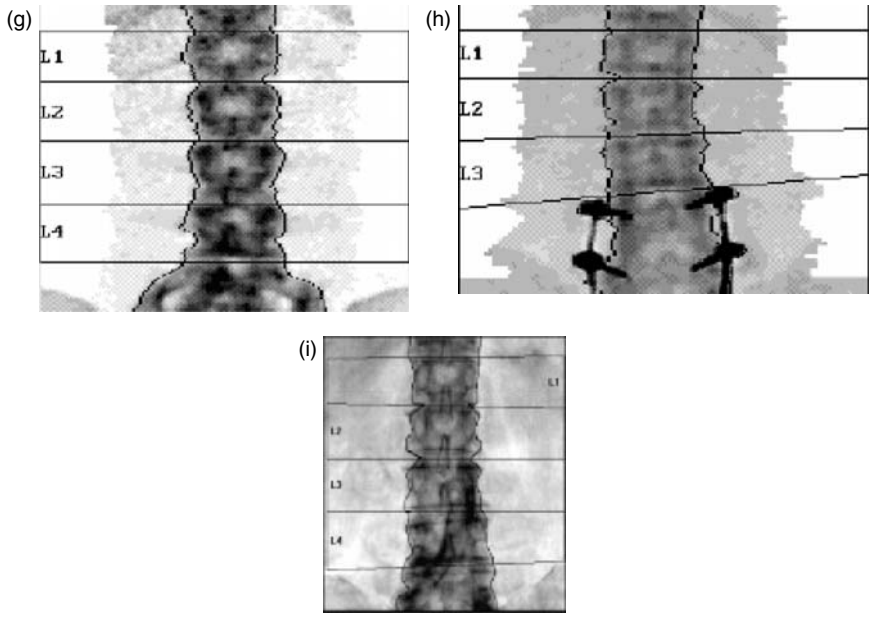


Figure 7.6. Examples of incorrectly acquired lumbar spine DXA scans and abnormal appearances. (a) Scan started too high. The lower edge of L4 has not been included. (b) Scan started too low. A significant portion of the sacrum has also been included. (c) Scan started too high. The entire L4 vertebra has been included but it is not possible to confirm whether the subject has only four lumbar vertebrae. The scan has been stopped too high at T11. The spine in not centred in the image and there is an artifact adjacent to L4. (d) Subject has six lumbar vertebrae. (e) Subject has osteophytes at all vertebral levels, with L2 and L3 particularly prominent. (f) Subject has a collapsed L3 vertebra. (g) Subject has bone bridging from L5 to the sacrum. (h) Subject has had previous surgery at L4. (i) A calcified aorta can be seen overlying L3 and L4.

collapse leads to an increase in BMD of the individual vertebrae because the area of the vertebra has decreased, whereas the bone mineral content (BMC) remains the same. OA changes will also produce an increase in BMD.[8,9] The clinical trial might have specific guidelines for the inclusion or exclusion of vertebrae.

Once a satisfactory scan acquisition is achieved continue with the analysis or comparison according to the guidelines in the manufacturer's operating manual and clinical trial specifications. The following further points should be noted:

1. When comparison analysis is carried out, ensure that the baseline ROI is overlaid onto the current scan such that the vertebral levels are identical to those at baseline.

2. Scoliosis analysis should be used where appropriate (Figure 7.8). Ensure this is also used at follow-up visits.

3. Review the individual vertebral levels. There should be a trend of increasing area and BMD from L1 to L4. A falsely elevated BMD could be owing to an

artefact, which could be removed (e.g. a button or navel piercing; Figure 7.9) and thus the scan should be repeated, or to vertebral collapse. A falsely elevated area could be owing to inclusion of the transverse process.

Print out the final analysis and record any modifications to the procedure on the printout for future reference. A copy should be retained with the subject's records and a further copy sent to the DXA QA centre, if required. All scans must be regularly saved to an electronic medium to ensure integrity of subject data. This procedure will be discussed in section 7.8.

7.5. Proximal Femur DXA Scan

The proximal femur is scanned in either the PA or the AP direction, depending on the type of scanner used. For all scanners, the subject lies supine on the mattress. Throughout this section this type of scan will be referred to as "the proximal femur".

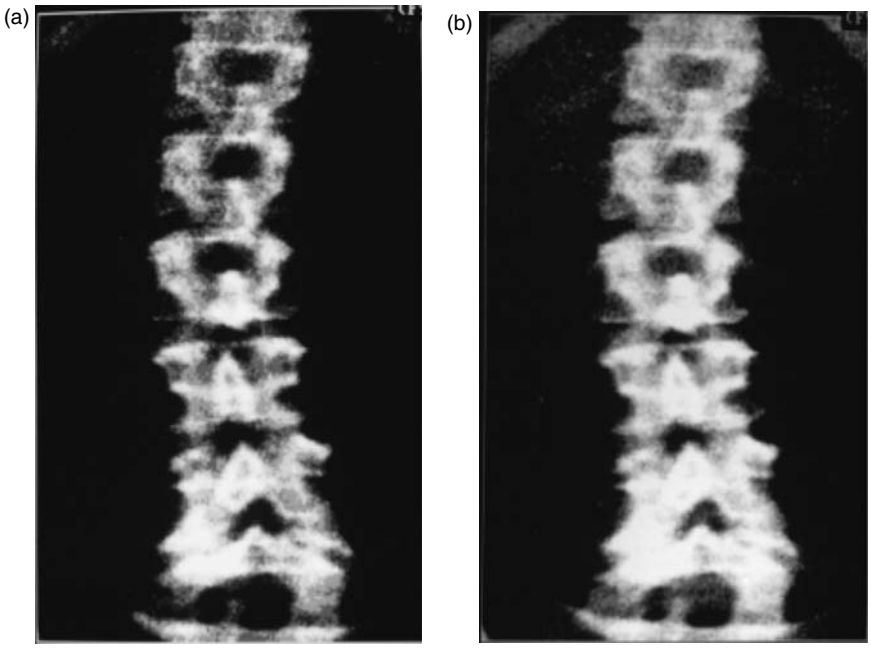


Figure 7.7. Density and contrast should be optimized before analysing a scan. (a) Without image enhancement the subject seems to have six lumbar vertebrae. (b) With optimal image enhancement a small pair of T12 ribs is shown.

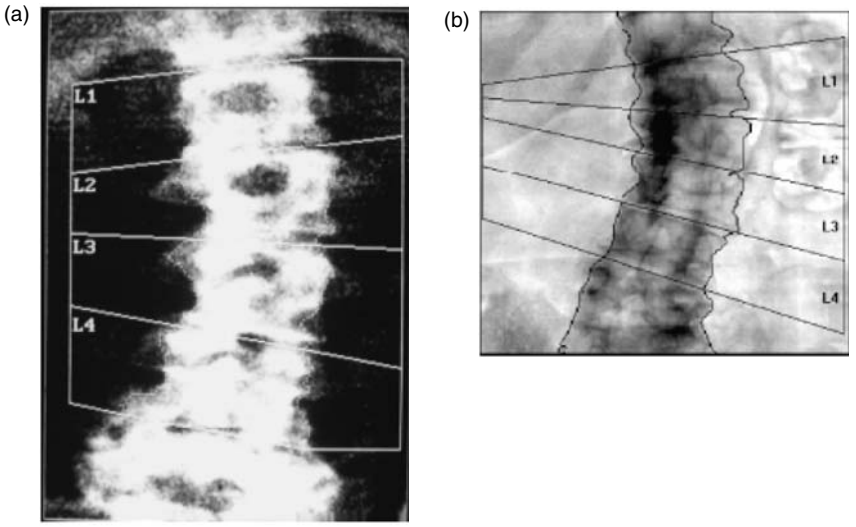


Figure 7.8. Scoliosis analysis should be used when necessary. (a) Hologic QDR 2000. (b) Lunar Expert-XL.

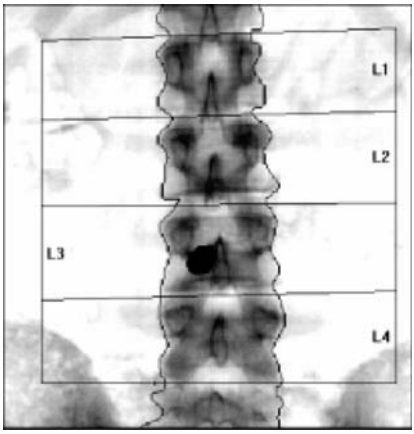


Figure 7.9. Subject with a navel piercing overlying the third lumbar vertebra. The subject was unable to remove the jewellery.

7.5.1. Subject Positioning

The left proximal femur is usually scanned, unless the clinical trial specifies otherwise. Inquire whether the subject has had a previous hip fracture or hip surgery, in which case the contralateral femur should be scanned. Similar to lumbar spine

positioning, the time spent initially ensuring that the subject is straight and correctly positioned will ensure a proximal femur scan that can be readily analysed. The proximal femur is very sensitive to correct positioning not just of the hip, but also of the entire lower limb. It is essential to review this positioning technique regularly, especially with new or part-time technologists.

Because the proximal femur scan is generally acquired after the lumbar spine scan, the subject should still be lying straight in the centre of the table (Figure 7.2). Each DXA operating manual has its own recommendations on positioning methods and devices. Before commencing positioning, explain to the subject that they are required to have their foot strapped in place for this scan. The subject might inquire why their foot has to be strapped when it is their hip that is being investigated! Reassure the subject, explaining what is required so that they will be relaxed and fully compliant with positioning. Incorrect leg positioning is the major contributor to poor precision at the proximal femur.

Ensure that the subject is lying straight on the mattress using the techniques described in Section 7.4.1. Identification of the greater trochanter should be carried out before any further positioning of the subject. The greater trochanter serves as the bony landmark for identifying the start position of the DXA scan, despite the manufacturers specifying different start points. Encourage the subject to relax both legs as much as possible at the hip joints because this will aid positioning.

7.5.1.1. Identification of the Greater Trochanter

Assuming the left proximal femur is being scanned, ask the subject to internally and externally rotate their left lower limb, turning from the hip, while keeping the knee in full extension. Guide the limb by gently holding the knee and ankle but ensure that it is the subject who is rotating the limb, because overrotation by the technologist could cause pain to the subject. (This technique also encourages the subject to relax their limb so that adequate internal rotation can be obtained before strapping the foot.) While the limb is rotated, gently palpate the greater trochanteric region, with your hand, to identify the position of the trochanter while it rotates with the limb. Identify the position by marking the white gown with a pen or, if the subject is wearing his or her own clothes, place a pen tip on the mattress, level with the greater trochanter.

7.5.1.2. Positioning of the Lower Limb

Positioning of the limb is the most important stage of the procedure because the entire limb must be correctly rotated, abducted, and secured. The aim is to position the femur at a constant degree of internal rotation that, in subjects with average femoral anteversion, enables the maximum length of the femoral neck to be visualized and lie clear of adjacent bone. This is achieved as follows:

1. Holding the subject's leg firmly at the knee and ankle, ask the subject to rotate their leg internally, externally, and internally again so that the medial border of the foot lies against a foot positioner (Figure 7.10).

2. All rotation must be at the hip; movement at the knee and ankle can be detected and prevented during this manoeuvre by guiding the subject's lower limb with a hand firmly holding each joint.

3. Following the final rotation, the foot should be secured to the foot holder with a Velcro® strap, with the medial edge of the foot following the angle of the holder.

4. A small pad placed under the knee will help maintain this position for subjects with painful hips or knees.

Additionally, you might be required to steady the holder with a sand bag on its base. Foot holders vary slightly between manufacturers; however, they all have an angle of around 25°, and therefore these principles of positioning can be applied to all holders.

The degree of lower limb abduction varies between manufacturers. Ensure the guidelines in the operating manual are carefully followed because the reference data will have been acquired using the methodology described in the manual. The Hologic manual instructs the operator to align the apex of the foot holder with the midline of the subject, thereby ensuring a constant degree of abduction for all subsequent visits. The Lunar manual instructs the operator to position the subject's leg parallel to the midline of the table. It is, therefore, not recommended to change foot holders between different manufacturers.

7.5.2. Scanner Preparation

Once the limb is positioned satisfactorily, ensure that the subject is comfortable and that their hands are outside the scan field (Figure 7.11). Bring the laser positioning light to the correct start point over the proximal femur, according to the manufacturer's instructions. Explain the duration of the scan to the subject, and that they should remain still during the procedure. If soft tissue equivalent material is required around the hip to exclude air from the scan (Lunar DPX range), it should be positioned at this stage, following the manufacturer's guidelines.

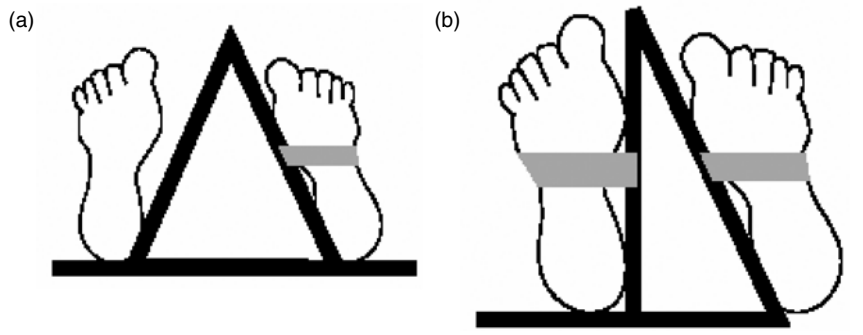


Figure 7.10. Correct positioning and internal rotation of the lower limb is aided by strapping the foot to the manufacturer's foot positioning guide. (a) Hologic. (b) Lunar.

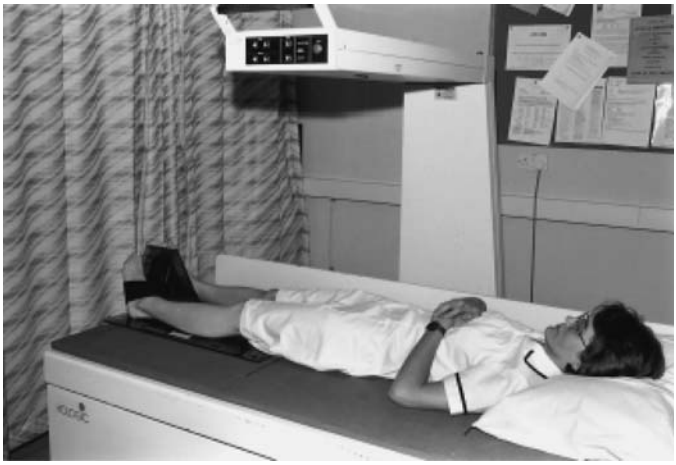


Figure 7.11. Positioning for a proximal femur scan with the Hologic foot positioner. The positioner is under the lower limbs with the left foot strapped in position. The hands are clear of the scan field.

7.5.3. Scan Mode Selection

Select the appropriate scan mode and parameters according to the manufacturer's operating manual or the clinical trial specifications. For subjects returning for follow-up scans, it is essential that the same scan mode and parameters are selected as those used at the baseline visit.

The scout scan mode, which was used for the lumbar spine scan (see Section 7.4.3), can be used if anatomical identification or positioning is difficult. However, the scout scan should not be used routinely in place of a good positioning technique.

7.5.4. Image Review During Scanning

As the proximal femoral image begins to appear on the screen, check the following points:

1. Is the start level correct, that is a sufficient distance below the lesser trochanter, according to the guidelines in the manufacturer's operating manual?
2. Is the femoral shaft sufficiently abducted, according to the guidelines in the manufacturer's operating manual?
3. Is the femoral shaft lying in the outer one-third of the image?
4. Are there any overlying artefacts that could be removed?

If necessary, quickly stop the scan and reposition either the subject or the scanner arm, or investigate any artefacts.

As the scan progresses, check the following points:

1. Is the femoral shaft sufficiently abducted, as required by each manufacturer?
2. Is there sufficient soft-tissue-equivalent material at the side of the hip to permit the ROI to be correctly positioned?

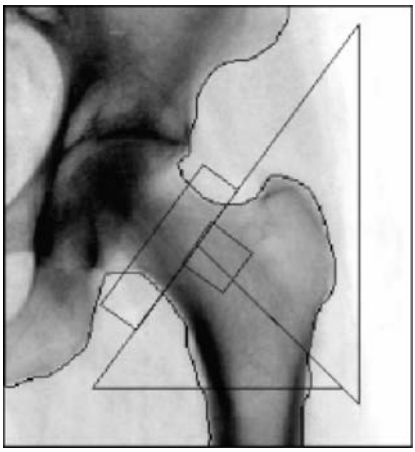


Figure 7.12. Proximal femur scan image correctly positioned with the shaft of the femur parallel to the centre line of the table. There is adequate soft tissue both above and below the neck of the femur and at the lateral margin of the greater trochanter. The femoral neck is not overlying the ischium. There are no artefacts.

3. Has the subject moved during the scan?
4. Are there any overlying artefacts that could be removed?
5. Are there any overlying artefacts that cannot be removed (e.g. ischium). These artefacts might require modification of the scan analysis or recording on the printouts and study documentation.

Stop the scan and make any necessary changes. Continue scanning until a sufficient distance above the femoral head, as specified by the manufacturer's operating manual. Examples of a correctly acquired proximal femur scan are shown in Figure 7.12.

7.5.5. Image Analysis/Comparison

On completion of the proximal femur scan, immediately review the image and analyse the scan before moving the subject, so that if a repeat scan is required in the same position, this can be readily acquired. Figure 7.13 illustrates examples of incorrectly acquired proximal femur scans and abnormal appearances. The review should include the points listed in the Section 7.5.3 and the following additional points:

1. Optimize the image density.
2. Is subject positioning similar to the baseline scan?
3. Is the degree of internal rotation similar to the baseline scan?
4. Is the entire ROI included, as defined in the operating manual?

Once a satisfactory scan acquisition is achieved, continue with the analysis or comparison according to the guidelines in the manufacturer's operating manual

and the clinical trial specifications. The following further points should be noted:

1. When comparison analysis is carried out, ensure that the baseline ROI is overlaid onto the current scan such that the neck and trochanteric regions are closely aligned.
2. Position the femora neck box ROI as instructed in the operating manual.
3. Review Ward's triangle and the trochanteric region for inconsistency of ROI placement.

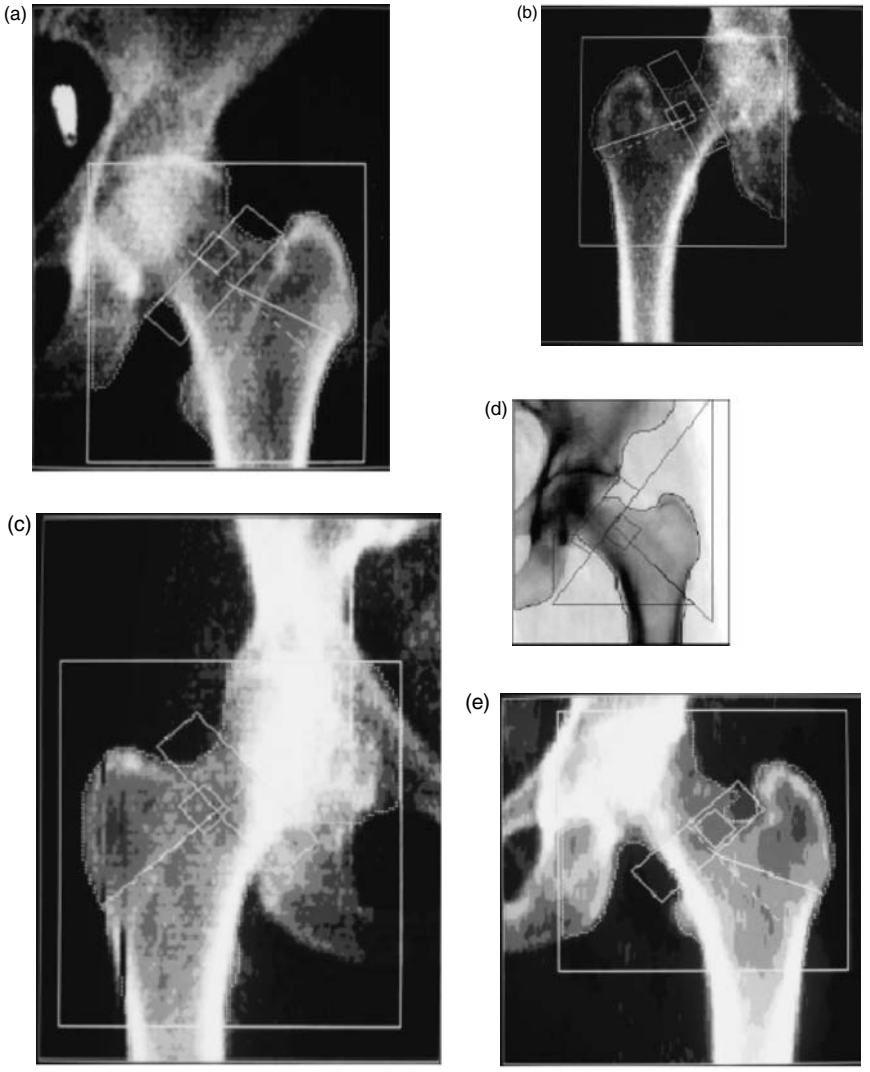


Figure 7.13. (Continued)

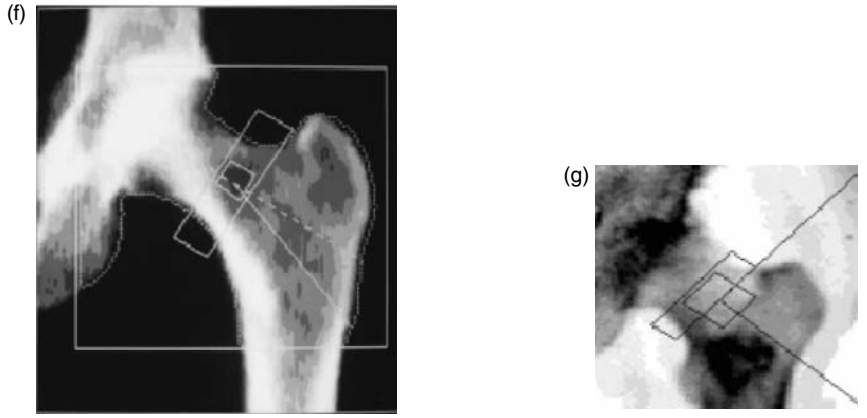


Figure 7.13. Examples of incorrectly acquired proximal femur DXA scans and abnormal appearances. (a) Scan started too high with a considerable area of the ilium also included. (b) Scan started too low with a considerable amount of the femoral shaft included. The femoral neck box is poorly placed with only two corners lying in soft tissue. (c) Movement artefact. (d) Poor bone outline at the femoral head. The automated software has placed the neck box too high. (e) Poor bone outline at the femoral neck. The automated software has placed the neck box too low. (f) Automated software has placed trochanteric line too low due to a bump on the side of the trochanter immediately below the trochanter. (g) Artefact in the femoral shaft. The subject had an old bullet injury.

Print out the final analysis and record any modifications to the procedure on the printout for future reference. A copy should be retained with the subject's records and a further copy sent to the DXA QA centre if required. All scans must be regularly saved to an electronic medium to ensure integrity of subject data. This procedure will be discussed in a later section.

7.6. Distal Forearm DXA Scan

The distal forearm is scanned in either the PA or the AP direction, depending on the type of scanner. Throughout this section this type of scan will be referred to as "a forearm scan".

The distal forearm can be scanned on either a dedicated forearm unit or an axial bone densitometer. The multitude of peripheral densitometer manufacturers and techniques means that a detailed description of individual machines is outside the scope of this book. However, techniques and principles of scanning are broadly similar for all machines.

7.6.1. Subject Positioning

The nondominant forearm is usually scanned, unless the clinical trial specifies otherwise. Inquire whether the subject has had a previous forearm

fracture, in which case the contralateral forearm should be scanned. As for the previous scan techniques, the time spent initially ensuring the subject's arm is correctly positioned will ensure a forearm scan image that can be readily analysed.

The subject should sit on a comfortable chair (with a firm supporting back and without wheels or arm rests) used specifically for forearm scanning, to ensure consistency of positioning. Measure and record the length of the subject's ulna, from the olecranon to the ulnar styloid process, for use during analysis. Position the subject's forearm according to individual manufacturers' instructions, encouraging the subject to relax their shoulder and elbow to assist with positioning. Ensuring that the subject is comfortable will improve compliance with positioning and maintenance of the position during the procedure.

For axial bone densitometers, the forearm is positioned pronated on the mattress and the elbow flexed to between 90° and 110°. Although a water bath is not required, a Perspex® positioning board is required under the subject's forearm with Lunar equipment.[5,6] The Perspex® serves to provide a constant background level for soft-tissue calculations and has a positioning grid marked on its surface to ensure the forearm is positioned parallel to the centre line of the table. Peripheral bone densitometers might require the subject's forearm to be submerged in a water bath, while they hold a positioning guide. When straps are provided, they should be used to secure the forearm during the scan.

7.6.2. Scan Mode Selection

Select the appropriate scan mode and parameters according to the guidelines in the manufacturer's operating manual or the clinical trial specifications. For subjects returning for follow-up scans, it is essential to select the same scan mode and parameters as for the baseline visit. The scanning time could be several minutes (8–10 minutes) for older pencil-beam systems; therefore, inform the subject of the importance of remaining still throughout the scan.

7.6.3. Image Review During Scanning

As the forearm image begins to appear on the screen, check the following points:

1. Is the subject's forearm in the middle of the image?
2. Is the start position correct?
3. Is there any rotation of the forearm?
4. Are there any overlying artefacts that could be removed (e.g. jewellery, watch, or buttons)?

If necessary, quickly stop the scan and either reposition the subject or the scanner arm, or investigate any artefacts. Figure 7.14 demonstrates a correctly acquired

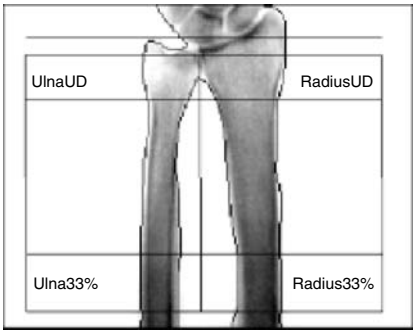


Figure 7.14. Correctly acquired forearm scan image including the distal radius and ulna plus some of the carpal bones. Adequate soft tissue has been included either side of the forearm. There is no rotation of the forearm, which would be demonstrated by overlapping of the distal radius and ulna. There are no artefacts.

forearm scan. As the scan progresses, check the following points:

1. Are there equal amounts of soft tissue on each side of the forearm?
2. Is there any forearm rotation?
3. Has the subject moved during the scan?
4. Are there any overlying artefacts that could be removed?
5. Are there any overlying artefacts that cannot be removed? These artefacts might require modification of the scan analysis or recording on the printouts and study documentation.

7.6.4. Image Analysis/Comparison

On completion of the forearm scan, review the image and analyse the scan before moving the subject, so that, if a repeat is required, this can be readily achieved. The review should include checking the points listed in the previous section, plus the following additional points:

1. Optimize the image density and contrast.
2. Is the entire ROI included, as defined in the operating manual?
3. Is the subject's positioning similar to the baseline scan?

Once a satisfactory image is acquired, continue with the analysis or comparison according to the guidelines in the manufacturer's operating manual and the clinical trial specifications. Figure 7.15 illustrates a forearm scan that has an incorrectly defined bone outline on the distal ulna. This effect is relatively common on both the distal radius and the ulna and should be corrected according to the manufacturers' guidelines.

Print out the final analysis and record any modifications on the printouts for future reference. A copy should be retained with the subject's records and a further

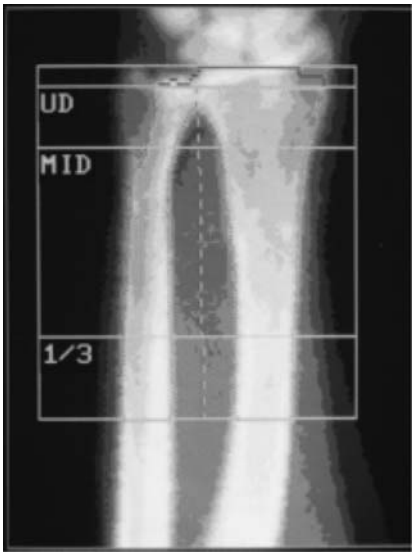


Figure 7.15. Bone outline of the distal ulna is poorly defined.

copy sent to the DXA QA centre, if required. All scans must be regularly saved to an electronic medium to ensure integrity of subject data. This procedure will be discussed in a later section.

7.7. Morphometric X-ray Absorptiometry (MXA)

Prior fragility fracture is becoming an integral part of the assessment of fracture risk in the osteoporotic subject. However, fractures of the thoracic vertebrae of the spine often do not present for clinical diagnosis. It is well known that subjects with one prevalent vertebral fracture are two or three times more likely to suffer a further fragility fracture, and if two or more prevalent fractures are present there is a ninefold increase in fracture risk, independent of BMD measurements[10–12]. Identification of a prevalent vertebral fracture should influence clinical management. It is an essential part of the pre-treatment work-up for parathyroid hormone (PTH) administration in the UK national guidelines on the secondary prevention of osteoporosis.[13] MXA is one technique used to quantify the extent of vertebral deformities from the fourth thoracic vertebrae to the fourth lumbar vertebrae (T4 to L4) and is an alternative technique to digitizing radiographs. Although MXA has some limitations, its comfortable subject positioning, good reproducibility, and semiautomated analysis will ensure that it has a role in vertebral height assessment.[14–16] With the latest generation of instruments, the technique has developed and is now referred to as lateral vertebral assessment (LVA) or vertebral fracture assessment (VFA), enabling automatic grading of vertebral fractures using either a semiquantitative visual method or

vertebral height measurements. Both single-energy and dual-energy lateral images are used for LVA/VFA, with different manufacturers using different image-acquisition modes.

Although LVA has good sensitivity for the identification of moderate-to-severe radiographic vertebral fractures (91.9%) and excellent negative-predictive value (98%), the International Society for Clinical Densitometry suggest that indications for and clinical use of LVA are not yet established.[17] Current UK guidelines, however, recommend that, if LVA is available, it should be used for assessment of vertebral fractures.[18] LVA is particularly useful when confounding factors are present on a PA DXA scan of the spine (e.g. the presence of osteophytes). It is, therefore, important to ensure that the assessment of vertebral morphometry by LVA is reproducible in terms of subject positioning, acquisition, and image analysis.

7.7.1. Subject Positioning

Many instruments now have a rotating C-arm that enables the subject's positioning for an MXA or LVA scan to be the same as for the AP lumbar spine scan. Care must be taken to ensure the thoracic spine is also parallel to the midline of the table. The subject must be instructed on any breathing procedures required during scan acquisition. The latest Lunar instruments, however, do not have the C-arm facility, and subjects have to be repositioned in the decubitus lateral position for scanning. This can present problems to the untrained operator, particularly if the subject has spinal curvature or scoliosis. Care must be taken to ensure that the vertebral end plates are parallel to the X-ray beam.

7.7.2. Scan Acquisition

Scan acquisition varies between manufacturers, with Hologic machines acquiring an AP scan initially, followed by a lateral morphometry scan.

The advantages of the AP thoracolumabar spine scan are as follows:

1. It enables the operator to ensure the subject's spine is parallel to the centre of the table
2. If the AP spine scan is acquired first, the start point for the lateral morphometry scan can be accurately identified
3. The number and level of the vertebrae can be identified.

On Lunar instruments, the scans are carried out in the same order, enabling the operator to identify any spinal curvature or scoliosis before moving the subject into the decubitus lateral position.

7.7.3. Scan Analysis

MXA scan analysis usually relies on a six-point semiautomated analysis procedure. The points are placed on the vertebral body as shown in Figure 7.16 so that

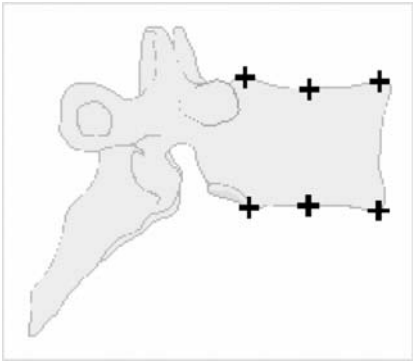


Figure 7.16. Six points are placed on the superior and inferior borders of the vertebral body so that the software algorithm can calculate the anterior, mid and posterior heights of the vertebrae.

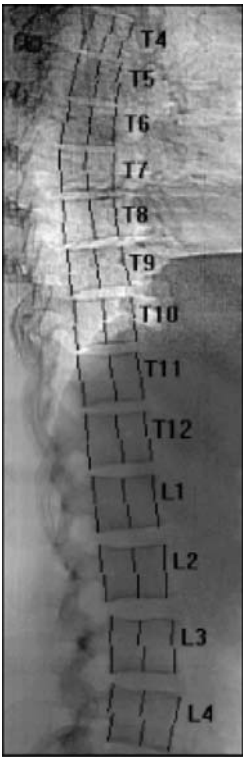


Figure 7.17. Correctly acquired MXA scan with T4 to L4 included.

the anterior, mid, and posterior heights of the vertebrae are recorded. Both Lunar and Hologic have developed different techniques for point placement so analysis must be carried out according to the instructions in the manufacturer's operating manual. Despite the procedure being termed "semiautomated", a considerable input is still required by the operator because each of the six points on all verte-

brae from T4 to L4 must be reviewed and its position adjusted accordingly. To ensure that point placement is standardized, the procedure should be reviewed by all operators. A correctly acquired and analysed MXA scan is shown in Figure 7.17.

LVA/VFA can be used to undertake a visual assessment of vertebral fractures using a semiquantitative grading method.[19] Fractures can also be identified using vertebral height measurements as follows:

1. A wedge fracture is defined by a ratio of the anterior to posterior vertebral heights (A/P ratio) Z-score of <-3
2. A biconcave fracture is defined by a ratio of the mid to posterior vertebral heights (M/P ratio) Z-score of <-3
3. A crush fracture is defined by anterior, mid, and posterior vertebral heights Z-scores of <-3

Fractures with a Z-score of <-4 are classified as severe fractures.

7.8. Data Archive

A regular procedure for data archive (also called "data back-up") should be established following installation of a bone densitometer. There is normally a standard archiving programme incorporated within the DXA software. Data archiving procedures vary greatly between manufactures and are discussed at length in all operating manuals. They will only be briefly discussed here. The aims of data archiving are as follows:

1. To free space on the scanner hard disk
2. To ensure long-term integrity of subject and QC data
3. To maintain two copies of all scans (both subject and QC data) on a permanent electronic medium, for reference at a later date.

To ensure data security and integrity is maintained, the following procedures should be written into a department protocol, and all technologists should be trained to undertake regular archiving of data:

1. Archiving (back-up) of subject and scan data must be done at the end of each day, to ensure data are not lost if there is a computer failure
2. Two copies of the archive should always be maintained
3. It is preferable to analyse scans before archiving the data so that both archives contain the most current record of the analysed scan
4. If changes are made to an archived scan, ensure the changes are saved on both archives
5. Do not store both copies of the archive together in the same cupboard. If data are lost because of fire or theft, you will lose all record of the subjects' visits. Store the main archive copy in a fireproof cabinet near the scanner that can be readily accessed. The second copy should be stored in a separate fireproof

cabinet in another part of the department where it can be readily accessed. Lock both cabinets at the end of the day to avoid theft.

A regular procedure for the "system back-up" should also be established. Its purpose is to copy the database system files. These files store subjects' biographies, DXA scan dates, BMD data, image archive information, and a range of information specific to each manufacturer. The system back-up is typically carried out once weekly and immediately before an equipment service. Use a different set of back-up data media. (e.g. compact disc or optical disc) once weekly for 4 weeks. At week 5, re-use the first set. This will ensure that there are several current copies of the database should any of the media become unknowingly corrupted. If your data are split into several databases (e.g. separate QC or trial databases) it will be necessary to repeat the system back-up for each database. Store these disks in a locked, fireproof cabinet in a separate room from the scanner.

7.9. Study Paperwork

This section has been included to review the paperwork supplied for clinical trials. Vast amounts of paperwork are designed and sent out by the DXA QA centres, but QA centres have very different ideas concerning the information required. Receiving an entirely different set of log sheets, QC requests, and data shipment details each time a new study starts is confusing, and mistakes are easily made. Reading through the DXA trial manual might be helpful, but until the first data shipment is compiled, the relevance of some paperwork could be unclear.

The reason for the vast amount of paperwork is to provide an "audit trail" for all data acquired and its subsequent transfers as evidence that the subject attended on a specific date and all queries relating to the subject were resolved. DXA QA centres should supply "no carbon paper" (which is better known as "NCR paper") to avoid endless photocopying, because this is when paperwork is lost.

The following general paperwork guidelines in Sections 7.9.1 to 7.9.5 can be applied to most clinical trials.

7.9.1. Scan Log Sheets

Scan log sheets are a record of the following:

1. The subject's identification (include subject's initials and identification number). The subject's full name should not be required, to maintain subject confidentiality
2. The dates of scan acquisition
3. The scan specific number. Each scan has a unique identifying number so scans cannot be confused between different visits and subjects
4. Scan-specific problems, which are recorded with the scan number for the attention of the QA centre.

Two formats of log sheet are commonly used: a single log sheet for all subjects scanned on one day (the preferred method for DXA technologists), and one log sheet for each subject (this generates an enormous amount of paper). Log sheets should be sent with the regular data shipment and a copy should be stored at the scan site.

7.9.2. Query Log Sheet

The query log sheet (also known as the "data action sheet") is used for scan-specific questions raised by either the technologist or the DXA QA centre. It could have several copy sheets attached for question-and-answer communications to and from the QA centre, so that, ultimately, both the DXA technologist and the QA centre have a copy of the initial question, analysis advice, and final outcome.

7.9.3. Service Report Log Sheet

A service report log sheet should be sent with each data shipment. This could be a single sheet with all the service visit details recorded since the previous data shipment (the preferred method for DXA technologists) or a single sheet for each service visit (this generates an enormous amount of paper). Routine service visits should be recorded, in addition to intermittent scanner faults.

7.9.4. Data Shipment Log Sheet

The data shipment log sheet should be compiled and sent with each data shipment. It lists the number and content of any data back-up media (either subject or QC scans) or other archive media sent with data shipments and the paperwork required by the DXA QA centre for each shipment, with tick boxes for completion. It serves as a useful prompt to include all the relevant forms in the data shipment. If the study does not provide data shipment log sheets, it is useful to compile a log sheet for reference when preparing a data shipment, to avoid omitting relevant documentation.

7.10. Shipment of Clinical Trial Data

Clinical trial BMD data should be sent at the specified regular intervals to the DXA QA centre. Regular shipment ensures that subject and QC data are regularly reviewed and problems are promptly addressed. For each shipment, follow the individual clinical trial guidelines and shipment log sheet to prepare the shipment, ensuring that copies of all paperwork are retained for the department's records.

Data shipments are usually sent directly to the DXA QA centre by courier. The clinical trial monitor will set up the courier service details. Do not send a data shipment in the post because confidential subject data are included in the shipment, and if the parcel is lost, it would not be insured. Each trial will be identified separately by the courier office, for the purpose of charging shipment costs to each trial, by a "third-party billing number", which should be quoted for all study-specific inquiries with the courier.

Each courier package sent by air requires a courier-specific airway bill, with details of the sender, recipient, account details, and goods being sent. If these are supplied already typed, check that the details are correct. When goods with value (electronic data media) are to be sent, a commercial invoice might also be required. This should be on the company's/hospital's headed notepaper and should include the number and value of goods in the package. Write the sender's and recipient's addresses clearly on the parcel itself, so that it can still be identified should the paperwork become separated from the parcel.

7.11. Record Keeping and Auditors

Participation in a clinical trial requires organized logical record keeping, in compliance with good clinical practice (GCP) guidelines and for the auditors.[20] The best advice is to be organized from the start of the trial; it can save a lot of problems later, especially if the site is subject to an audit.[21] If a DXA centre is participating in several clinical trials, records must be maintained separately for each trial. Subject source data (e.g. electronic data, printouts, or reports) should be kept for not less than 15 years following completion of the clinical trial. This requirement should be clearly identified on all relevant files and suitable secure storage should be arranged.[20] The following section reviews issues for which auditors could require to see documentary evidence.

7.11.1. Protocols

The following protocols should be readily available in the vicinity of the scanner:

1. Clinical trial protocol—the principal investigator should supply a copy for the DXA technologist, and forward any protocol updates
2. DXA protocol—the DXA QA centre should supply a manual, containing details of scanning protocols and trial-specific information
3. Bone densitometer operator's manual—an up-to-date copy should be supplied on installation of the machine. It should be kept with the densitometer so that it is readily available
4. Site-specific protocols—this includes protocols developed in addition to the manufacturer's procedures, for example a protocol for the regular archiving of data.

7.11.2. Training

The auditor will require documentary evidence that technologists are suitably qualified in both the use of equipment capable of generating ionizing radiation and the techniques of BMD assessment. Appropriate documentation includes the following:

1. A signed *curriculum vitae* (CV) for each technologist involved with the clinical trial.
2. Documentary evidence of DXA training. Although many technologists receive the initial manufacturer training, this is early in the training process and might not cover the whole variety of subject situations that can arise. After this initial training, following a period of broad DXA experience, it would be appropriate for further training to be undertaken. This could be dedicated training for a clinical trial often in the form of a study day. Additionally, there are several organizations that offer DXA training. The International Society of Clinical Densitometry (ISCD) in the USA offers a 2-day training and certification course. In the UK and Europe, training is available as a study day from organizations, such as the UK National Osteoporosis Society (NOS), or a training course with certification from universities.
3. Many clinical trials have DXA training days before commencement of the trial. This is an opportunity to review the scanning and analysis requirements of the trial. Technologists attending the training are expected to have some experience of scanning, because detailed training in DXA principles is not normally included. This is an ideal opportunity to meet other technologists and the DXA QA centre representatives to discuss the trial and scanning informally.

7.11.3. QC Procedures

As discussed earlier, daily QC of a bone densitometer is an essential prerequisite before scanning subjects. The QC results should be reviewed and graphed on a monthly basis, to look for trends or step changes in the data. Print and store the graphs in chronological order. Printing the graphs on a monthly basis provides documentary evidence that the QC has been regularly performed and reviewed.

7.11.4. Bone Densitometer Maintenance Procedures

Bone densitometers should have a service contract to ensure regular 6-monthly machine serving and technical support by the authorized supplier. Records of service reports for both routine service visits and scanner faults should be maintained with the densitometer, whereas copies should be sent with the regular data shipment to the QA centre.

Additionally, the technologist should complete a fault log for all problems that occur, including those that do not require an engineer. It should include date of fault, a description of the fault, and subsequent action taken. This provides a useful record for the engineer's service visit and might highlight recurrent problems.

7.11.5. Subject Files

Subject files (DXA printouts and copies of log sheets) should be stored in a dry, secure place in a clear logical order, either alphabetically or by subject trial number. Auditors could ask to review a selection of subject records, or, if only a small number of subjects have been recruited at the site, they might want to review all the subjects' records. They will inform the centre, before arrival, which subjects' records they wish to review.

7.11.6. Correspondence

All correspondence pertaining to the study should be kept and stored in chronological order. Always keep copies of correspondence sent from the DXA centre. This serves as documentary evidence that queries were answered. Following a visit by the auditors, the DXA centre should receive a written report from the auditors listing the problem areas, if any. The auditors will require written confirmation that problems have been addressed. Keep a record of this correspondence for future reference because the clinical trial could be audited again at a later date. Earlier problems will be an obvious area for investigation by the auditors.

The same auditing principles are applicable to all clinical trials; therefore, problems identified by one auditor can be rectified in all trials at a centre. Maintain clinical trial documentation separate from other trial information at the same site.

7.11.7. Contact Information

A list of relevant contact names should be included in the "quick reference guide" (Figure 7.1). Each clinical trial will have a local study coordinator, employed by the hospital, who is responsible for the day-to-day organization of the study and reports directly to the principal investigator.

The pharmaceutical company might employ a clinical research organization (CRO) to oversee the set-up and running of the trial. The CRO contracts a DXA QC centre to oversee the DXA. A clinical research associate (CRA), employed by the CRO, will visit the hospital on a regular basis, the frequency depending on the number of subjects recruited to the site. The CRA is responsible for the integrity of the data at each site and should meet regularly with the DXA technologist to ensure everything is running smoothly and there are no problems. However, many pharmaceutical companies run their own studies, only employing a CRA and DXA QA centre.

Establishing links with all the above personnel will aid the smooth running of the study. At the start of the clinical trial a meeting with the local study coordinator and CRA will establish the logistics of subjects attending for their clinic, DXA, and other essential appointments. A regular clinical trial team meeting will ensure that any problems can be readily addressed.

7.12. Summary

The most important aspect of DXA is consistency at all stages of the technique, from the preparation of the subject and the operator to reproducible subject positioning and analysis so that the BMD results produced are accurate and precise. Despite the fact that DXA has been available for more than 14 years, it is only in more recent years that it has becoming increasingly available. However, the extent of consistent formal training is still very limited, with many technologists only becoming familiar with the procedures during their working day. While ensuring all clinical trial staff are appropriately trained, the DXA technologists should also receive training in GCP and clinical trial procedures that are beyond routine clinical DXA scanning. Implementation of the procedures discussed in this chapter will ensure integrity of the BMD data for clinical trials.

References

1. Jergas, M. and Genant, H.K. (1997). Lateral dual X-ray absorptiometry of the lumbar spine: Current status. *Bone* **20:**4311–14.
2. Blake, G.M., Herd, R.J.M. and Fogelman, I. (1996). A longitudinal study of supine lateral DXA of the lumbar spine: A comparison with posterior spine, hip and total-body DXA. *Osteoporos Int* **6:**462–70.
3. Formica, C.A. (1999). Total body bone mineral and body composition by absorptiometry. In: Blake, G.M., Wahner, H.W. and Fogelman, I. (eds). *The Evaluation of Osteoporosis: Dual Energy X-ray Absorptiometry and Ultrasound in Clinical Practice*, 2nd edn, p. 313–46. London: Martin Dunitz.
4. Wahner, H.W. and Fogelman, I. (1994). *The Evaluation of Osteoporosis: Dual Energy X-ray Absorptiometry in Clinical Practice*, p. 49–74. London: Martin Dunitz.
5. Lunar DPX-L Operators Manual Documentation Version 8/93. Madison: Lunar Corporation.
6. Lunar Expert-XL User's Guide Documentation Version 2/98 1. Madison: Lunar Corporation.
7. Hologic QDR 2000 Operators Manual and User's Guide. Hologic System Software Version 7.10. Document Number 080–0384 Revision A. Bedford: Hologic Inc.
8. Masud, T., Langley, S., Wiltshire, P. *et al*. (1993). Effect of spinal osteophytosis on bone mineral density measurements in vertebral osteoporosis. *BMJ* **307:**172–3.
9. Liu, G., Peacock, M., Eilam, O. *et al*. (1997). Effect of osteoarthritis in the lumbar spine and hip on bone mineral density and diagnosis of osteoporosis in elderly men and women. *Osteoporos Int* **7(6):**564–9.
10. Ross, P.D., Davis, J.W., Epstein, R.S. *et al*. (1992). Ability of vertebral dimensions from a single radiograph to identify fractures. *Calcif Tissue Int* **51:**95–9.
11. Ross, P.D., Yhee, Y.K., He, Y.F. *et al*. (1993). A new method for vertebral fracture diagnosis. *J Bone Miner Res* **8:**167–74.
12. Klotzbuecher, C.M., Ross, P.D., Landsman, P.B. *et al*. (2000). Patients with prior fractures have an increased risk of future fractures: a summary of the literature and statistical synthesis. *J Bone Miner Res* **15:**721–39.
13. Bisphosphonates (alendronate, etidronate, risedronate), selective oestrogen receptor modulators (raloxifene) and parathyroid hormone (teriparatide) for the secondary

prevention of osteoporotic fragility fractures in postmenopausal women. Technology Appraisal 87 (2005). London: National Institute for Clinical Excellence. http://www.nice.org.uk/pdf/TA087guidance.pdf accessed 20/11/06
14. Steiger, P., Cummings, S.R., Genant, H.K. *et al*. and the Study of Osteoporotic Fractures Research Group (1994). Morphometric X-ray absorptiometry of the spine: Correlation *in vivo* with morphometric radiography. *Osteoporos Int* **4:**238–44.
15. Rea, J.A., Steiger, P., Blake, G. *et al*. (1998). Morphometric X-ray absorptiometry: reference data for vertebral dimensions. *J Bone Miner Res* **13(3):**464–74.
16. Cawte, S.A. (1999). Morphometric X-ray absorptiometry using the Hologic QDR 2000, MPhil thesis. Nottingham: University of Nottingham.
17. Writing Group for the ISCD Position Development Conference (2004). Technical standardization for dual-energy x-ray absorptiometry. *J Clin Densitom* **7:**27–36.
18. Management of Osteoporosis (2003). Scottish Intercollegiate Guidelines Network. http://www.sign.ac.uk/guidelines/fulltext/71/index.html accessed 20/11/06
19. Genant, H.K., Wu, C.Y., van Kuijk, C. *et al*. (1993). Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* **8:**1137–48.
20. Blunt, B.A., Jones, D., Svenson, R.K. *et al*. (1998). Good clinical practice and audits for dual X-ray absorptiometry and X-ray imaging laboratories and quality assurance centres involved in drug trials, private practice, and research. *J Clin Densitometry* **1(4):**323–37.
21. McDonald, V. (1994). Surviving an FDA audit. *Clin Res Focus* **5(2):**12.

8

Laboratory and Instrument Quality Control

Derek Pearson and Nigel Lawson

8.1. Introduction

Clinical trials now run for many years. Sponsors need assurance that the instruments they use remain stable for that length of time and report consistent results. It is important, however, to clarify the difference between quality assurance (QA) and quality control (QC). During any process, there should be ongoing process control. This is usually an automatic feedback situation. QC is the ongoing sampling of the process being evaluated. This enables change in the process, but not normally immediately, unless sampling and analysis is rapid. QA is the evaluation of the QC process or audit. It is normally conducted periodically and is a sampling process of the QC process to ensure the described checks are being conducted, such as in a manufacturing line where, for example, washers are being made, the mean weight of the washers being produced can be evaluated. If the average weight of the washers varies beyond the predetermined limits, the mix can be adjusted to keep the washers within the basic limits. For every 100 washers made, one washer will be taken and carefully measured. If the washer is outside the predefined limits, the manufacturing process can be adjusted and, if necessary, stopped and restarted, to ensure the washers are made to specification (QC). Periodically, an audit will be made of the whole process to ensure everything was conducted to all the written specifications (QA). QC is the remit of the local investigator and QA could form part of the audit carried out by the clinical research organization (CRO), QA centre, or trial sponsor.

The aim of this chapter is to discuss the QC necessary at a centre carrying out a clinical trial. The chapter will cover dual-energy X-ray absorptiometry (DXA), ultrasonometry, and biochemistry QC, if bone markers are being used in a trial.

8.2. DXA and Ultrasonometry

A number of types of failure can occur, as follows:

1. A step change in bone mineral density (BMD) or quantitative ultrasound (QUS) measurement
2. A gradual trend in results
3. An increase in scatter about a mean result, which is often an indication for operator error rather than a change in instrument performance
4. Changes owing to the environment, for example a temperature change affecting a QUS instrument moved from one location to another.

Experience with the instruments in common use would suggest that technical QC is not that useful in predicting instrument failure—this is usually acute and severe. It works and then it fails. Step changes are more likely to be induced by component changes or servicing visits than equipment failure. Investigators must be aware of the size of any change so that an informed decision can be taken on the need to correct results from a particular instrument.

The examples in this section mainly discuss DXA QC. The principles and techniques, however, apply to other modalities.

8.2.1. What QC is Necessary?

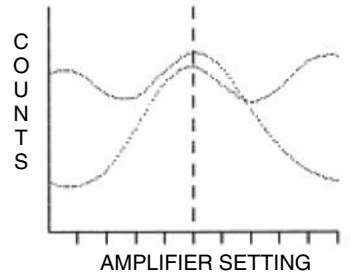
It is vital that the QC procedures required by the manufacturer are followed on a daily basis. QC should be carried out even on days when no subjects are booked to attend in order to apply some of the analytical techniques described in this chapter. Weekends and public holidays must be ignored! Over and above the recommended QC, sponsors must ensure that a phantom is measured on a daily basis that will provide linearity information because this is essential if there is to be any attempt to correct the data on the basis of phantom results. To this end, the European spine phantom (ESP), the Bona Fide phantom (BFP) or Lunar (GE Healthcare, Madison WI, USA) aluminium spine phantom can be used for BMD measurements (see Chapter 4), although hydroxyapatite-based phantoms are preferred by the US Food and Drug Administration (FDA).

There is a significant difference in QC philosophy between the main DXA manufacturers, Lunar and Hologic (Bedford MA, USA). Lunar instruments report detailed QC results for a series of radiation, electrical, and mechanical checks, in addition to bone mineral content for bone standards in a calibration standard. The results from the bone standards are stored and used as a daily calibration by the equipment in the calculation of BMD. A "PASS" or "FAIL" is reported for each of the radiation, electrical, and mechanical checks (Figure 8.1). At the time of writing, the latest version of software is being released, which enables a second QC check to be performed at start-up that mimics a subject measurement.

By contrast, Hologic devices use a daily phantom measurement for QC. The philosophy makes the assumption that the equipment is working if the correct BMD result is reported. The phantom includes only one BMD value (see Chapter 4),

QUALITY ASSURANCE RESULTS

X-RAY Voltage (kVp)	76.0	Date	09.03.1
X-RAY Current (uA)	150.0	System	7060



DETECTOR

AMPLIFIER SETTING	COUNTS LOW KeV	COUNTS HIGH Kev
280	64432	4896
330	65242	2298
380	53386	10794
430	51613	33837
480	65942	59114
530	76483	68461
580	67866	58454
630	44877	48885
680	24122	56653
730	18006	71458
780	4954	74742

PROCEDURE	VALUE	EVALUATION
Lights	–	Pass
Peak Setting	530 units	Pass
Background (Low keV)	0 cps	
(High keV)	1 cps	Pass
Beam Stop Action	–	Pass
Percent Spillover	7.98 %	Pass
Chi square	7	Pass
Air counts (Low keV)	785176 cps	Pass
(High keV)	473440 cps	Pass
Air Ratio	0.60	Pass
Transverse Mechanics	12444/12444 steps	Pass
Longitudinal Mechanics	19840/19845 steps	Pass
Tissue Value	1.311	Pass
Collimation Ratio	3.998	Pass

	1	2	3	4	5	MEAN	SD	%CV
LARGE BM	279.6	277.0	277.2	276.4	276.5	277.3	1.16	0.42
WIDTH	499	492	495	497	496	496	2.32	0.47
MEDIUM BM	210.0	206.2	209.1	206.5	209.2	202.0	1.47	0.70
WIDTH	434	434	430	431	431	432	1.67	0.39
SMALL BM	147.8	145.3	147.4	145.7	148.9	147.6	1.31	0.89
WIDTH	263	362	366	361	366	364	2.06	0.57

Figure 8.1. Lunar DPX-L QC printout.

which is not adequate to monitor clinical trials.[1] Because there is no check of the mechanics of the scanning couch, it is also important to monitor the area of regions of interest (ROIs), in addition to BMD, because this could reveal problems with couch movement.

Many of the ultrasonometry instruments rely on a single phantom measurement as the basis of QC. There are no linearity phantoms that can be easily obtained commercially. The Leeds phantoms described in Chapter 4 are not robust or easily obtained. All phantoms seem to be temperature-sensitive. Water-based equipment can compensate for this by using a temperature-controlled water bath and allowing time for the phantom temperature to stabilize before a measurement is taken.

The advantage with QUS is that an *in-vivo* measurement can be carried out and used instead, particularly if low-density and high-density heels are available among the staff. They could still be temperature-dependent, however.[2,3] Daily measurements might not be possible and different analysis methods have to be used to cope with an in-vivo change in QUS results over the period of years, but good consistency can be obtained.

8.2.2. Baseline Phantom Measurements

To make effective use of daily phantom measurements, a baseline measurement must be made. Multiple measurements made on the same day are not adequate because they do not take into account the daily variation of the equipment. Data must be acquired according to the interval for which the equipment is to be monitored (on a daily basis in most circumstances). This is demonstrated in Figure 8.2, where the multiple measurements made by the manufacturer during commissioning and servicing using the Hologic spine phantom can clearly be seen to be below the average for the phantom. The reference value on the commissioning visit was 1.0372 ± 0.0047 g/cm^2 ($n=20$), whereas the average for the following 6 months was 1.0396 ± 0.0045 g/cm^2. The reference value is significantly different to the long-term average at the $p < 0.1$ level. In Figure 8.2, the manufacturer has

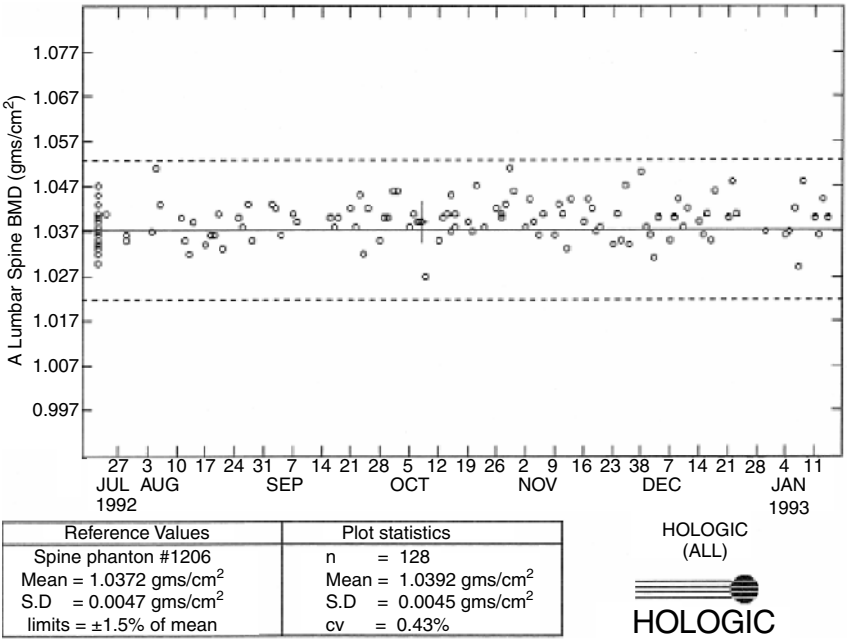


Figure 8.2. Hologic quality control chart. Multiple measurements made by the manufacturer during commissioning using the Hologic spine phantom can clearly be seen to be below the average for the phantom.

included the reference value measurements in the mean plot statistics, which is why it is different to the long-term average above. Had the reference value been established for the first 20 days, it would have been 1.0382 ± 0.0045 g/cm^2, much closer to the long-term average.

To ensure a stable baseline is established, it is recommended that measurements are carried out for at least 15 days. If there is a failure in the QC data and the problem has been identified and corrected, the baseline must be reestablished. Because of the length of time this takes, there are periods when it is not possible to monitor QC apart from using the manufacturer's accepted methods.

8.2.3. Monitoring Techniques

QC results can be monitored using a number of statistical techniques. Trending tools are available in the manufacturers' software. The longitudinal plot of phantom data available within the Lunar software is of little value, and the data should be extracted to another database or spreadsheet package for analysis and plotting. On Hologic equipment, the daily phantom results are plotted on a control chart according to the mean expected BMD ± 1.5% (Figure 8.2). If the measured BMD falls within these limits, the equipment is deemed to have passed. Trending tools are provided within the software (moving average and linear regression). These will be discussed in detail below, but are not ideal for identifying the different kinds of failure identified above. What is required are monitoring techniques that prospectively identify failures rather than allow a retrospective review of the data. To this end, the Shewhart rules and Cumulative Sum (CUSUM) charts will be discussed, which are methods that enable failures to be identified as they happen.

8.2.4. Linear Regression

Linear regression is only appropriate when one particular type of fault is evident in the QC data, i.e. when there is a linear trend in the data. It has been applied by some authors to identify the rate of change within a linear trend,[4] but they recognised that what might appear as a trend could also be interpreted as a step change in the baseline. Figure 8.3 shows some longitudinal DXA data from a Hologic QDR 2000 in which there has been a step change following equipment repair in May 1993. The Hologic reference value and ± 1.5% limits are shown. At first glance, the linear regression model seems to adequately describe the data (Figure 8.3a), whereas the step change is more evident in Figure 8.3b. The step change model minimizes the variance of the phantom data around the fitted model. It is possible to develop software within a spreadsheet that will automatically identify the breakpoint, assuming that step change has occurred somewhere within the data. The variance about the breakpoint model is also shown in Figure 8.3b, demonstrating that the breakpoint occurs at the point of minimum variance on 30 April 1993.

Thus, the linear regression model must be used with care. It is most likely that there has been some sort of step change in the phantom data, and, mathematically,

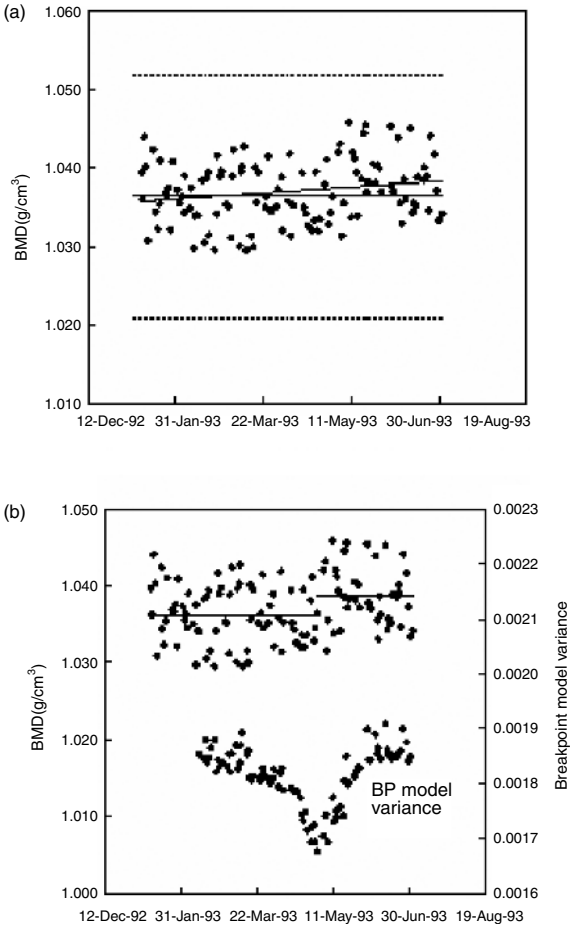


Figure 8.3. Linear regression and breakpoint analysis of QC data (a) Linear regression through QC data with 6 1.5% limits shown. (b) Breakpoint analysis of the QC data in Figure 8.3a. The residual variance of the breakpoint model is shown. The breakpoint is chosen where the variance is at a minimum.

it is best to use the simplest model that best describes the data (i.e. the breakpoint model). The breakpoint model is difficult to apply prospectively.

8.2.5. Moving Average

The moving average is best used for aiding the eye when examining data. It is not easy to use for identifying failures, but has been used as the basis of correction methods. The problem is to identify how many points to include in the average to make the average useful. Too few, and the graph looks no better than the original data. Too many, and significant random errors are smoothed into the average.

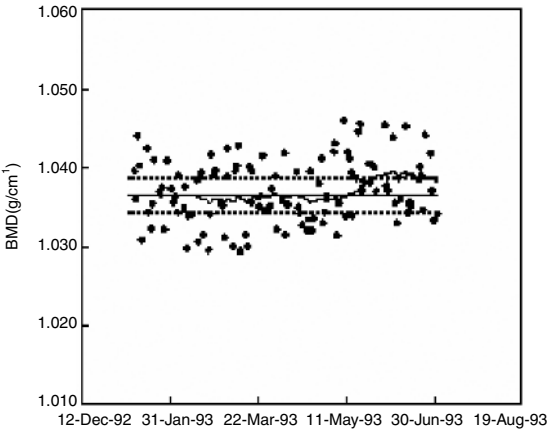


Figure 8.4. Moving average of the QC data in Figure 8.3a.

Figure 8.4 shows the same QC data as Figure 8.3 using a 25-point moving average. The reference value of the mean and action limits are also shown.[5] The action limits (in this case ± 0.6 standard deviations [SD] of the reference value) depend on the number of points within the moving average. The upper action limit is exceeded on 27 May 1993. The moving average will identify the trend. An action limit can also be set on the SD of the moving average, which will help to identify random fluctuations and an increase in the variance of the data.[5]

8.2.6. Shewhart Rules

These rules were developed within Clinical Chemistry community to assist the QC of biochemical assays.[6] They have since been widely applied to BMD data.[7–9] Phantom BMD results are plotted on a simple control chart after a baseline value and the SD of the baseline value have been established, as above. The target value of BMD should be plotted as zero and the individual measurements plotted in SD units, or dual axes can be used to simplify plotting new data (Figure 8.5). The total BMD for a particular phantom can be plotted, but individual ROIs should also be plotted to check that there are not linearity problems with the equipment. For example, on the Lunar DPX-L instrument it might be appropriate to plot the BMD of the second to fourth lumbar vertebrae (L2 to L4) of the Lunar aluminium spine phantom, in addition to each individual vertebra. As each new BMD result is plotted on the control chart, the Shewhart rules are applied (Table 8.1). If the first rule is not broken, no further action is required. If the first rule is broken, the other rules are applied. If one of the subsequent rules is broken, further investigation is necessary. Some investigators use a fixed percentage of the baseline value (0.5%) to represent 1 SD.[10] Thus, rule 2 is as follows: "Do two successive points fall outside ± 1% of the baseline value?" This is a reasonable approximation to the

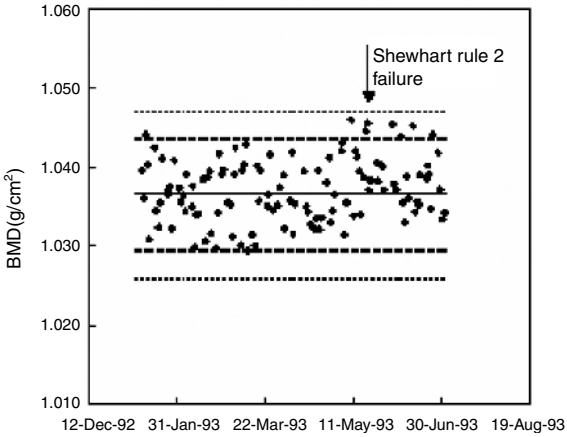


Figure 8.5. Shewhart rule plot of the QC data in Figure 8.3a. The ± 2 SD and ± 3 SD limits are shown.

Table 8.1. The Shewhart rules.

1.	If a point falls outside 2 SD of the target value, apply the following rules.
2.	Do two successive points fall outside 2 SD of the target value?
3.	Does one point fall outside 3 SD of the target value?
4.	Do four successive points fall outside 1 SD of the target value on one side of the target value?
5.	Is there a difference of 4 SD between successive points?
6.	Are 10 successive points on one side of the target value?

measured SD of the Hologic spine phantom, but cannot be applied to other phantoms or individual vertebrae of one of the other phantoms in common use because the SD is likely to be higher than 0.5%. It is relatively simple to extract the QC information from the databases on the densitometers and develop a simple spreadsheet to apply the Shewhart rules.

Rules 3 and 5 are designed to identify random errors, represented by an increased scatter around the baseline value. These could be truly random as, by definition, 4.5% of data points will lie outside ± 2 SD of the baseline value. However, if there is a general increase in scatter about the baseline value, sponsors should consider reviewing operators' training to ensure that positioning and analysis is correctly carried out. Are operators scanning with the same acquisition parameters and using the compare function during analysis?

Rules 2, 4, and 6 are designed to identify systematic failures, which are more characteristic of instrument failure. They will identify a gradual trend of the data away from the baseline value, or a step change in baseline.

Figure 8.5 gives an example of Shewhart rule failure on data collected using the Hologic spine phantom on a Hologic QDR 2000 instrument. The failure is not demonstrated on the standard Hologic QC control chart, but was related to a fault on the scanning arm's motion.

8.2.7. CUSUM Charts

CUSUM charts were developed by industry,[11] used for gamma camera QC,[12,13] and then adapted for use in bone densitometry.[5,8,9] The baseline phantom measurement or target value must be established using the methods discussed above. For each subsequent data point, the difference between the measured value and the target value is calculated. The cumulative sum (CUSUM) of these differences is then calculated as follows:

$$C_n = \sum_{i=1}^{n} (d_i - T). \tag{8.1}$$

where C_n is the cumulative sum of n data points, d_i is the measured BMD data for the ith data point, and T is the target value. C_n is then plotted on a chart, in which the vertical axis is plotted in SD units of the target value. This enables a direct comparison of CUSUM charts irrespective of the absolute value of the target value and the SD. Where the DXA remains stable, the CUSUM chart will be horizontal and scattered about zero. A rising or falling CUSUM chart suggests that there is a trend away from the target value. A sample CUSUM chart is shown in Figure 8.6.

There are two methods that can be used to see if there has been a significant change in phantom measurements. The first is to calculate control limits. The upper control limit is given as follows:

$$U_i = ((d_i - d_{\text{mean}})/\sigma) - 0.5 + U_{i-1}, \tag{8.2}$$

where U_i is the upper control limit for the ith data point, d_{mean} is the average of the phantom values up to and including the ith data point, σ is the SD of the target value,

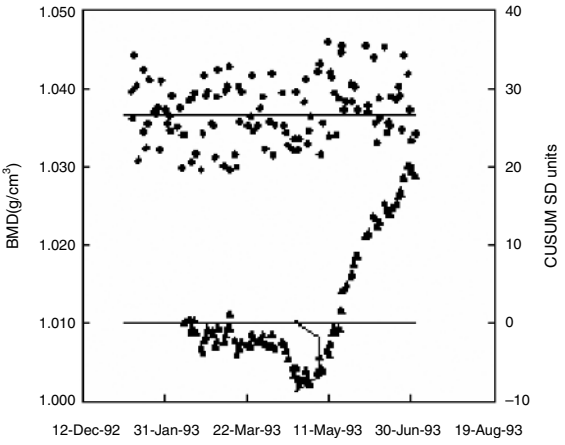


Figure 8.6. CUSUM plot of the QC data from Figure 8.3a. The lower curve is the CUSUM data. The mask identifies a QC failure in early May 1993.

and U_{i-1} the previous value of the upper control limit. The lower control limit is as follows:

$$L_i = ((d_{\text{mean}} - d_i)/\sigma) - 0.5 + L_{i-1}. \tag{8.3}$$

If U_i or L_i fall below zero, they are reset to zero. If either value exceeds a value of 5, a QC failure is said to have occurred.

The alternative method is to place a V-shaped mask on the most recent data point on the CUSUM graph (Figure 8.6). The height of the mask and slope of the arms are related to the standard error of the target value and determine the stringency of the chart in detecting QC failures. If any of the points before the most recent data point fall outside the mask, a QC failure is said to have occurred. The larger the mask, the fewer the QC failures that will be detected. The size of the mask must be determined by the data from the equipment and phantom being used because it will depend on the inherent precision of the equipment measured using that phantom. If there is greater scatter around the target value, a less stringent test must be applied with a larger mask. If the scatter around the target value is small, a smaller mask can be used because a smaller change could represent a QC failure.

The size of the mask can be set using receiver operating characteristic (ROC) curves.[9] The size of the mask is varied and the number of true-positive and false-positive QC failures is identified. The optimal value of the mask can be determined from the ROC curves as the point at which the ROC curve intercepts the diagonal connecting 100% of true-positives to 100% of false-positives (Figure 8.7). Using

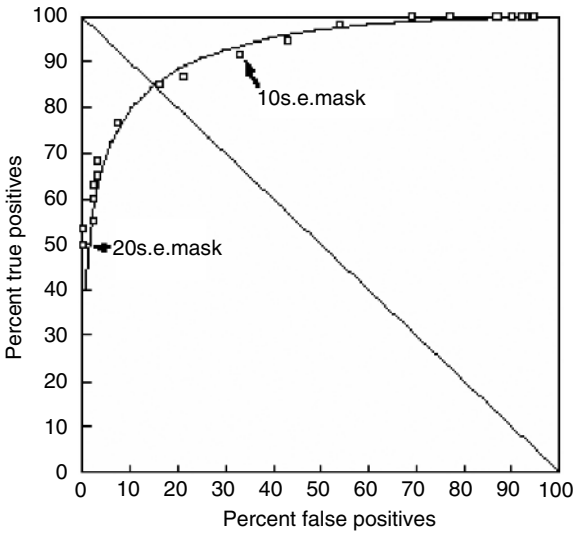


Figure 8.7. ROC curve varying the mask size on the CUSUM plot. The percentages of true-positive and false-positive QC failures are recorded for each mask size. The optimal mask size is chosen where the diagonal line crosses the ROC curve.

the Hologic spine phantom on a Hologic QDR 2000 instrument, one author has used a mask height of 2.6 SD of the target value[9] and a slope of 0.26 SD of the target value per data point. This gave a true-positive rate of 77% and a false-positive rate of 7%. Alternatively, a decision can be taken on the acceptable percentage of false-positive failures and the mask can be set to give that level of false-positives.[8] In this case, the height of the mask was 3.6 SD and the slope was 0.9 SD per data point. Using the Lunar aluminium spine phantom on Lunar DPX equipment gave a true-positive rate of 53% and a false-positive rate of 8%. The true-positive rate is lower because the slope of the arms is greater than in the first example. The mask size has to be determined for each data set, and the complexity of establishing the CUSUM technique has deterred many investigators.

8.2.8. Managing QC Failures

At well-organized clinical trial sites, QC will be reviewed regularly by the investigators and the trial's QA centre, to ensure that failures are identified quickly. It is vital that failure is detected early and that there is a good relationship between the investigators and the QA centres. The first indication of a QC failure can often be picked up by eye. The data should be reviewed in an ongoing fashion, by both the investigator-site staff and the QA centre. The error might seem random, and it is reasonable to wait and see what happens over the next few days. However, once a trend or step change has been established, it is unethical to continue monitoring subjects if a known failure has occurred.

Appointments will have to be cancelled while the cause of the problem is identified and fixed. At this stage, it is usual to contact the manufacturer or distributor of the instrument to organize a service call and have the problem corrected. Once a solution is found, it is wise to remeasure the phantom data for a few days to demonstrate that the problem is cured. The question then arises, should data be corrected? There is no right or wrong answer to this question, but there are a number of principles that can be used as a guide, as follows:

1. If it was a short-term failure involving a few subjects, it might well be simpler to call them back for a further measurement.

2. Avoid correcting data if at all possible, because every data correction adds more error to the results.

3. If the change in measured BMD is small ($<5\%$ has been suggested) and the data are part of a larger multicentre trial, avoid correcting the data. The variation will be lost in the wider between-centre and between-subject variations.

4. The error in applying the correction has to be considered in the context of the probable change in BMD and size of the correction. The error in correction will be between 0.5% and 1%, depending on the phantom used, and the number of points included in determining the correction factor. If a trend is identified a linear regression, a correction factor might be necessary, depending on the data of the subject measurement.

5. If, for example, the change occurs after all study subjects have had a 1-year follow-up measurement and before any of them have the next measurement, do

not correct if you only want to demonstrate a treatment effect. Both treatment and control groups will be affected in the same way and the treatment effect will still be present in the data. Problems arise if you want to quantify the magnitude of the treatment effect, however. If the fault occurs during one group of subject measurements (halfway through the 1-year measurement), there is a greater need to correct the data.

6. Applying the breakpoint model (Figure 8.3b) is a good method of identifying the size of the correction factor needed.

7. The change should be evaluated at a range of densities. A phantom measuring only one density (e.g. 1 g/cm^2) might not adequately represent the changes in the calibration over the range of densities of the subject population being evaluated (e.g. 0.5–0.75 g/cm^2). In other words, there could be nonlinear changes. In the experience of one of the authors, a change was noted at high density, but for the subject population under evaluation, the change in BMD was low enough to avoid having to correct the data.

8.2.9. Equipment Replacement

Investigators should avoid equipment replacement at all costs. The reliability of most DXA equipment is such that this should be possible even in long-term trials. Some instruments have been in use for 10 years without a significant shift in calibration or any need to apply corrections to subject data, because of calibration shift.

The ideal for planned replacement is as follows:

1. Develop an *in-vivo* crosscalibration using subjects who cover the whole range of BMD values. Include reasonable subject numbers, selecting from elderly osteoporotic subjects, normal postmenopausal, and young normal subjects to ensure the range of BMD values. Between 60 and 100 subjects will give a standard error of estimate for the *in-vivo* crosscalibration of approximately 3%.

2. Run the instruments side by side for as long as possible and measure trial subjects twice on both instruments during periods some time apart, to demonstrate that any trend in BMD is monitored across both systems. This is not unethical because it ensures data quality within the trial.

If this is not possible, obtain a crosscalibration with an appropriate phantom (see Chapter 4) that has a range of BMD values. The ESP or BFP will give adequate crosscalibrations that can be applied to group data, but they are harder to apply in the individual clinical setting. Figure 4.5 shows an *in-vivo* crosscalibration between a Lunar DPX-L device and a Hologic QDR 2000 instrument, with the *in-vivo* regression and phantom data points superimposed. The crosscalibration from the BFP is coincident with the *in-vivo* regression in this group of subjects.

If the equipment has been replaced without notification to the QA centre, recall as many subjects as possible who were measured in the 1 month preceding the changeover and measure them again as soon as possible after the change. Use this data as the basis of an *in-vivo* crosscalibration, making the assumption that there

has been no significant change in BMD during this time. The error associated with the crosscalibration will be larger because of this assumption, but should be adequate to provide a crosscalibration that can be supplied to group data.

It is not adequate to rely on standardized BMD in the context of instrument replacement because inter-instrument variability is not taken into account in standardization.

8.3. Biochemistry QC

QC has a crucial role in any laboratory, and in clinical laboratories, the way in which it is performed can have a bearing on the quality of treatment a subject can expect to receive.

8.3.1. Accreditation

All laboratories undertaking clinical trials should have some form of accreditation. All the various components, which are important in assuring the quality of data produced, can come under an accreditation umbrella. For example, documentation that the refrigerators and freezers were maintained at the correct temperatures for the storage of specimens and reagents, reagents were made up correctly, or equipment was correctly maintained are all crucial pieces of information when trying to assess the quality of results obtained. It is possible that, for each trial, independent auditors can assess each of these categories. However, it is simpler to ensure that participating laboratories have some form of accreditation, which should assure that these procedures are undertaken.

Various bodies are available to ensure that the laboratory is performing effectively. For example, in the USA and UK it would be expected that all clinical laboratories measuring markers of bone turnover and calcium homeostasis would be accredited by the Clinical Laboratory Improvement Amendments of 1988 (CLIA'88, http://www.fda.gov/cdrh/clia/index.html accessed 20/11/06) and Clinical Pathology Accreditation (CPA, http://www.cpa-uk.co.uk/accessed 20/11/06), respectively. Indeed, it would be surprising to employ any hospital laboratory that did not have this type of accreditation. Although the CLIA'88 and CPA are very rigorous and try to ensure that the right result and clinical advice reaches the right subject at the right time, they are not as exacting as some of the other schemes with respect to, for example, archiving of raw data. Therefore, audit trails on the production of test results are sometimes more difficult to undertake than in, for example, an ISO 9000 or good laboratory practice (GLP)-approved laboratory, for which there would be less emphasis on the clinical usefulness of the data generated but more time spent ensuring that any data were generated exactly as according to the standing operating procedure (SOP) stated and a full audit trail is possible. Because most clinical laboratories do not usually wish to incur the extra expense of joining extra accreditation schemes, these extra audit trails could be set up by the investigators if required.

8.3.2. Internal QC

Whatever the test, whether it generates quantitative or qualitative results, there has to be some form of internal QC (IQC). For quantitative results, IQC should take into account both the precision and the accuracy of the methods employed. For qualitative results, there should be some form of assessment of variability. For example, where results are just reported as positive or negative, a positive and negative control should always be included. Most markers of bone turnover and related clinical chemistry analyses are quantitative; therefore, most of what follows refers to quantitative analyses.

8.3.2.1. Precision

The precision of any analysis can usually be expressed as a coefficient of variation (CV), which is simply expressed as a percentage:

$$\mathrm{CV} = \frac{\mathrm{SD} \times 100\%}{\mu}, \tag{8.4}$$

where SD is the standard deviation and μ is the mean of repeated measurements of the assay. The precision of the assay can then be described in terms of intrabatch or interbatch CVs, where the intrabatch CV would have been obtained by determining the mean and SD for an analyte measured on the same sample a given number of times (n) *within* the same batch. The interbatch CV would have been obtained by determining the mean and SD for an analyte measured on the same sample a given number of times (n) *between* batches. In both cases, n should be a value of 10 or more.

All assays should have had their intrabatch and interbatch CVs determined, and laboratories should regularly check the precision of assays to ensure they are performing correctly. Furthermore, the precision of an assay is required if correct QC procedures are to be put in place. The interbatch CVs for a given analyte should be greater than the intrabatch CVs because the former should have a component of the latter within them.

CVs can be also be determined by the so-called "two-up" method of comparing a number of duplicate samples, either within the same batch or between batches. This method can give rise to CVs that are lower than those determined as described above and should be avoided because they could give a false impression of how well an assay is performing.

The concentrations used to determine CVs are also very important. Ideally, they should be determined at a number of relevant concentrations for that particular analyte. For example, there is very little point in determining the CV of a plasma calcium method at 5.0 $\mathrm{mmolL^{-1}}$ or 6.0 $\mathrm{mmolL^{-1}}$ if values of 2.2–2.6 $\mathrm{mmolL^{-1}}$ would be expected in a normal population. Because low and high calcium results can be expected, it would, therefore, be useful to determine CVs at low, normal, and high concentrations, for example 1.8 $\mathrm{mmolL^{-1}}$, 2.4 $\mathrm{mmolL^{-1}}$, and 3.2 $\mathrm{mmolL^{-1}}$.

Interbatch CVs will vary from laboratory to laboratory and from analyte to analyte. For routine analyses undertaken on modern automated analysers, the interbatch CV for calcium, for example, should be <2.0% on concentrations in the range of 1.8–3.2 mmolL^{-1}. For more manual techniques and for less robust techniques, performance can fall off dramatically, especially at lower concentrations of the analyte. For example, for 25-hydroxycholecalciferol, as measured in duplicate by a commercial radioimmunoassay kit (INCSTAR Corporation, Stillwater, MN, USA), the quoted interbatch CV is 12.2% for a concentration of 46.6 μgL^{-1}. This concentration is at the top of the expected reference range for 25-hydroxycholecalciferol. At concentrations that might be expected in deficient states (<8.9 μgL^{-1}), the quoted interbatch CV is 18.3% for 25-hydroxycholecalciferol at a concentration of 7.8 μgL^{-1}.

It is important for both clinical and laboratory investigators to be aware of the significance of such precision figures. For example, the quoted interbatch CV for a commercial radioimmunoassay (Immunotopics Inc., San Clemente, CA, USA) of the bone resorption marker osteocalcin is 6.7% at a concentration of 1.5 μgL^{-1} and 5.5% at a concentration of 14.2 μgL^{-1}. The quoted precision uses osteocalcin concentrations in the range of expected values found in a normal population (2.4–11.7 μgL^{-1}), and such CVs can be expected using a commercial assay in duplicate. What this means in routine clinical practice, if the usual QC procedures are employed (see below), is that any value ± 2 SD of the mean would be accepted for a subject with an expected osteocalcin of 14.4 μgL^{-1}, that is for this subject's sample, an osteocalcin result between 12.8 μgL^{-1} and 16.0 μgL^{-1} could be produced.

For assays with higher interbatch CVs, for example the 25-hydroxycholecalciferol assay discussed earlier, the differences are even greater. Using the company's own figures for a sample with an expected 25-hydroxycholecalciferol concentration of 7.8 μgL^{-1}, any result ± 2 SD of this value could be expected, that is any value between 5.0 μgL^{-1} and 10.6 μgL^{-1}. Therefore, at this concentration of 25-hydroxycholecalciferol, a doubling in concentration could have occurred by chance.

This situation is further compounded if various assays are used in routine practice and "quoted CVs" of manufacturers are very rarely achieved. This does not mean that the laboratory is a poor performer, it is just a recognition of the inherent analytical variability of any analysis. Clearly, researchers should consider this both when designing experiments and when trying to interpret the significance of results. Furthermore, suspicion should be raised when laboratories quote impossibly low interbatch CVs for complex analyses. An interbatch CV of <2% for a radioimmunoassay of 25-hydroxycholecalciferol at a concentration of 8.0 μgL^{-1} would be excellent; however, it would have little, if any, relationship with reality. Laboratories should always determine their own CVs if they are going to use their results meaningfully and be able to construct reliable QC charts.

8.3.2.2. Accuracy

The accuracy of any measurement can be defined by how close the concentration you obtain for an analyte is to its true concentration in that sample. For simple analytes,

calibration material is usually available that has accurately defined concentrations. For example, the concentration of calcium in serum-based calibration material can be determined by a reference method such as inductively coupled plasma mass spectrometry (ICP-MS). Therefore, a very accurate result can be ascribed to that material, and your routine method can be calibrated against this result.

Reference methods are available for most commonly measured analytes, and reference material with accurately determined concentrations are available from the National Institute of Standards and Technology (NIST, 100 Bureau Drive, Stop 3460, Gaithenburg, MD20899–3460, USA. http://www.nist.gov/accessed 20/11/06) and the National Institute for Biological Standards and Control (NIBSC, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3QG, UK. http://www.nibsc.ac.uk/accessed 20/11/06). Ideally, laboratories or the manufacturers of the reagents should be able to show the traceability of results back to these or similar materials.

The situation with most markers of bone turnover is, unfortunately, not so straightforward. Most of these markers do not have readily available calibrators that can be traced back to a reference material. This has obviously led to variability between methods, and in the case of multicentre trials, the same method should be used on all sites.

8.3.2.3. Monitoring IQC

For modern, random-access analysers, QC material should be run at frequent intervals, but how frequently is a matter of debate. Most laboratories try to follow manufacturers' recommendations about the frequency of running QC material. Because calibration on modern analysers is usually very stable, the minimum recommendations are to run three levels of control material once daily.

For batch analysis, QC material should be placed at the beginning and end of each batch, and always run with each batch. In large batches, QC material should be run at different points throughout the batch, for example every 10 specimens.

In either case, the QC results should be recorded accurately and charted using a typical QC chart. Most modern analysers have these charts built into the PC-based software that operates the instruments. For manual assays, these data should be charted, using either pen and paper or one of the many commercial QC packages available. It is important that whichever method of assessing internal QC is used, the charts are regularly examined and readily available for inspection.

Most laboratories adopt the Westgard rules for IQC, which basically state that the batch should not be accepted if the following occur during the analysis of a QC sample:

1. The result is $> \pm 3$ SD
2. More than one result is $> \pm 2$ SD
3. More than 10 consecutive results are $> +1$ SD
4. More than 10 consecutive results are > -1 SD.

Because these rules are dependent on the method of determination of the inter-batch CV, the importance of accurately determining your own CVs becomes self-evident.

IQC data are crucially important and should be available for inspection at all times. The use of computer-held data is a distinctive advantage. However, steps must be taken to ensure that it is regularly reviewed and notes, etc, are clearly logged. A computer full of IQC data that is never looked at will hide a multitude of sins.

8.3.2.4. Choice of IQC Material

Commercial QC material is readily available for most routine analyses. However, such material is not available for many bone markers. Some kit manufacturers provide QC material with the kit. Ideally, extra QC material should also be employed, prepared, or collected from an independent source to the kit manufacturer. In-house material can be used, but great care must be taken in its storage, and checks on sample stability must be established. In addition, care should be taken with these potentially high-risk samples when using in-house material.

8.3.2.5. Subject Means

Although IQC will demonstrate many analytical variances, other, mainly preanalytical variables will not be detected. There are numerous preanalytical variables that can affect the eventual result and should be taken into account. GLP should ensure these are kept to a minimum, for example always taking the samples at the same time of day, minimizing the delay between taking samples and storage, minimizing changes in storage conditions, always using the same type of collection tubes, etc. If sufficient analyses are performed, monitoring subject means is a useful adjunct to IQC, and can highlight changes in some of the preanalytical variables.

8.3.3. External QA (EQA)

There is a simple rule for EQA: if there is a scheme available for the analytes being investigated, join it. To be accredited with CPA, laboratories have to be registered with the appropriate EQA schemes, if they are available. There are several schemes in the UK that operate under the umbrella of UK National EQA Schemes (NEQAS). Further details can be obtained from http://www.ukneqas.org.uk/accessed 20/11/06. Until the various markers of bone turnover analytes become used more widely, it is unlikely that many EQA schemes will become available for their measurement. At present, there is a pilot scheme being operated by UK NEQAS from the Department of Immunology, Northern General Hospital, Sheffield, UK (http://www.immqas.org.uk/accessed 20/11/06). The scheme includes urine deoxypyridinoline, urine *N*-telopeptide, serum bone-specific alkaline phosphatase, serum osteocalcin, and serum procollagen 1 C-terminal propeptide. The laboratory's performance on these various schemes should be available for scrutiny.

Other Clinical Chemistry schemes are available in the UK, including the Wales External Assessment Scheme (WEQAS, Medical Biochemistry Department, University Hospital of Wales, Heath Park, Cardiff, UK. http://www.weqas.co.uk/ accessed 20/11/06) and the Randox International Quality Assessment Scheme (RIQAS, Randox Laboratories Ltd, Ardmore, Diamond Road, Crumlin, Co. Antrim, UK, http://www.randox.com/English/products.cfm?ccs=630 accessed 20/11/06). Most countries with developed pathology services normally have their own EQA schemes in place. The CLIA'88 website lists several proficiency testing organizations for North America.

If no scheme is available, it is advisable for the investigators and/or the trial organizers to establish comparability between the laboratory methods used. For multicentre trials, it is useful if samples with known concentrations of the analytes under question are measured at all sites. This will act as an independent check and should ensure comparability between sites. Such a process should ideally be carried out before the trial commences, and at several times throughout the trial.

8.3.4. Alternative Methods

Even if all the above QC and QA processes are put into place, further checks can be used when required. For example, there might be a concern about a relatively unstable peptide, such as osteocalcin, where there should be minimum delay between taking the sample, separating the serum, and freezing the sample. A sample taken in a fluoride oxalate tube (inhibits glycolysis) can be taken at the same time as the clotted sample for osteocalcin measurement. Glucose will not be metabolised in the fluoride oxalate sample, but will be metabolised at $>10\%$ per hour in an unseparated clotted sample. Measurement of glucose in both samples will reveal any delays in sample separation.

8.4 Summary

This chapter has discussed the laboratory and DXA QC required to ensure that a CRO will be assured that the data they receive is of the quality that they require. It has summarised mathematical techniques that can be used with any quantitative QC data. QC often seems a time consuming, expensive luxury, but is well worth the effort as it will allow the early identification of faults on DXA equipment, provide methods of correcting BMD data and give confidence in laboratory results.

References

1. Cawte, S.A., Pearson, D., Green, D.J. *et al*. (1999). Cross-calibration, precision and patient dose measurements in preparation for clinical trials using dual energy X-ray absorptiometry of the lumbar spine. *Br J Radiol* **72:**354–62.
2. Pocock, N.A., Babichev, A., Culton, N. *et al*. (2000). Temperature dependency of quantitative ultrasound. *Osteoporos Int* **11:**316–20.

3. Iki, M., Kajita, E., Mitamura, S. *et al*. (1999). Precision of quantitative ultrasound measurement of the heel bone and effects of ambient temperature on the parameters. *Osteoporos Int* **10:**462–7.
4. Gluer, C-C., Faulkener, K.G., Estilo, M.J. *et al*. (1993). Quality assurance for bone densitometry research studies: Concept and impact. *Osteoporosis Int* **3:**227–35.
5. Lu, Y., Matthew, A.K., Blunt, B.A. *et al*. (1996). Dual X-ray absorptiometry quality control: Comparison of visual examination and process-control charts. *J Bone Miner Res* **11:**626–37.
6. Westgard, J.O., Barry, P.L., Hunt, M.R. *et al*. (1981). A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* **27:**493–501.
7. Orwoll, E.S., Oviatt, S.K. and Biddle, J.A. (1993). Precision of dual-energy X-ray Absorptiometry: Development of quality control rules and their application in longitudinal studies. *J Bone Miner Res* **8:**693–9.
8. Garland, S.W., Lees, B. and Stevenson, J.C. (1997). DXA Longitudinal quality control: A comparison of inbuilt quality assurance, visual inspection: multi-rule Shewhart charts and Cusum analysis. *Osteoporos Int* **7:**231–7.
9. Pearson, D. and Cawte, S.A. (1997). Long-term quality control of DXA: A comparison of Shewhart rules and Cusum charts. *Osteoporosis Int* **7:**338–43.
10. Faulkner, K.G. and McClung, M.R. (1995). Quality control of DXA instruments in multicentre trials. *Osteoporos Int* **5:**218–27.
11. British Standards Institute (1980–1981). *BS5703 Guide to data analysis and quality control using Cusum techniques*. London: British Standards Institute.
12. Pearson, D. and Kapasi, T. (1988). A comparison of two methods of measuring gamma camera uniformity. *Nucl Med Commun* **9:**69–73.
13. Knight, A.C. and Williams, E.D. (1992). An evaluation of Cusum analysis and control charts applied to quantitative gamma-camera uniformity parameters for automated quality control. *Eur J Nucl Med* **19:**125–30.

Section 3

Data Analysis and Presentation

9

Data Analysis and Presentation: Writing a Paper for Publication

Derek Pearson

9.1. Introduction

The extent to which your clinical trial will contribute to the greater scientific good will depend, to a large degree, on the quality of the presentation and dissemination of the results. Your trial is likely to be one of many that address the research question you have posed. In some cases, the treatment effect will be overestimated and results, particularly from small trials, will be contradictory. The results from a number of trials will probably have to be combined to get a true picture of the effectiveness of a new molecular entity (NME). Ideally, the report of your trial will be of sufficient quality to be included in a metaanalysis and demonstrate the effectiveness of your intervention in the treatment of osteoporosis. There are, unfortunately, a number of limitations that are common when writing up trials that lead to bias and the exclusion of studies from subsequent metaanalysis, including the following:[1]

1. Use of multiple endpoints (if 20 items are measured on a subject, one is bound to be significant—result: a publication)
2. Use of surrogate endpoints [e.g. bone mineral density (BMD) as a surrogate marker of fracture risk]
3. Too many subgroup analyses
4. Incorrect analysis of repeated measures
5. Too many treatment groups in one study
6. Small study numbers
7. Underreporting of nonsignificant results.

The standard of reporting of clinical trials has improved significantly over the years, but a review of the journals will reveal that many of these inadequacies are still present. This is an ethical problem for investigators. For trials to provide a sound basis for effective treatment of osteoporosis, they must be well designed, well executed, and well reported. Badly executed and badly reported studies are of little benefit to subjects.

The aim of this chapter is to present the results from a small study in a way that is adequate for publication. A warning, however: the sample data are test data and are provided so that the reader can check his or her sums when implementing an analysis. These data do not stand up to close scrutiny against the standards laid out in this chapter, but enable calculations to be simply implemented and checked in many of the common spreadsheet and statistics packages. The chapter will also take as an example only a randomized, double-blind, controlled trial. The principles will apply to other designs of trial but the detailed statistics might not (e.g. cross-over trials). A detailed description of common statistical tests (e.g. paired *t*-tests) is not included, but analysis of variance applied to longitudinal data is covered in some detail.

9.2. The CONSORT Statement (Consolidated Standards of Reporting Trials)

The CONSORT statement was published in 1996[2] and revised in 2001[3] as a response to the "wide chasm between what a trial should report and what is actually published in the literature." It provides a checklist and flow chart that enable authors and reviewers to check that a trial is adequately reported. It provides six headings and five subheadings that can be used within a publication to enable readers to make a judgement about the trial in a standardized format, as follows:

- *Title*. Make sure the title describes the type of trial (e.g. randomized, double-blind, cross-over, etc).
- *Abstract*. Make use of a structured format in the abstract (see below)
- *Method*:
  - *Protocol*. Describe the study population, in addition to the inclusion and exclusion criteria, interventions and their timing, primary and secondary outcome measures, and minimum important differences in those measures, and indicate how the proposed sample size was calculated. Describe the methods for statistical analyses and whether an intention-to-treat analysis was undertaken. If appropriate, describe any stopping rules
  - *Assignment*. Describe the method used to assign subjects to the different treatment arms of the trial
  - *Blinding*. Describe the methods used to blind the study, including the appearance and taste of capsules (if appropriate)
- *Results*:
  - *Subject flow.* Use a flow chart to show the subject flow through the trial (see Figure 9.1)
  - *Analysis.* State the effect of intervention on primary and secondary outcome measures. Remember to include confidence intervals. Always give results in absolute numbers if possible (e.g. 17 out of 34 subjects rather than 50%). Present the summary data and statistical analysis in such a way that your results can be duplicated by someone else and the results can be used usefully in, for example, a metaanalysis

- *Comment*. Present an interpretation of the study findings that is supported by the evidence—that is do not try to overinterpret your data. Identify any limitations and bias within your study. Put your conclusions within the context of the evidence available in the wider literature.

There is a useful bibliography in the statement that supports the inclusion of most of the descriptors. The flow chart provides information about the progress of subjects through a randomized, controlled trial (RCT) with two groups. An example is given in Figure 9.1. This is the most common type of trial, but the guidance in

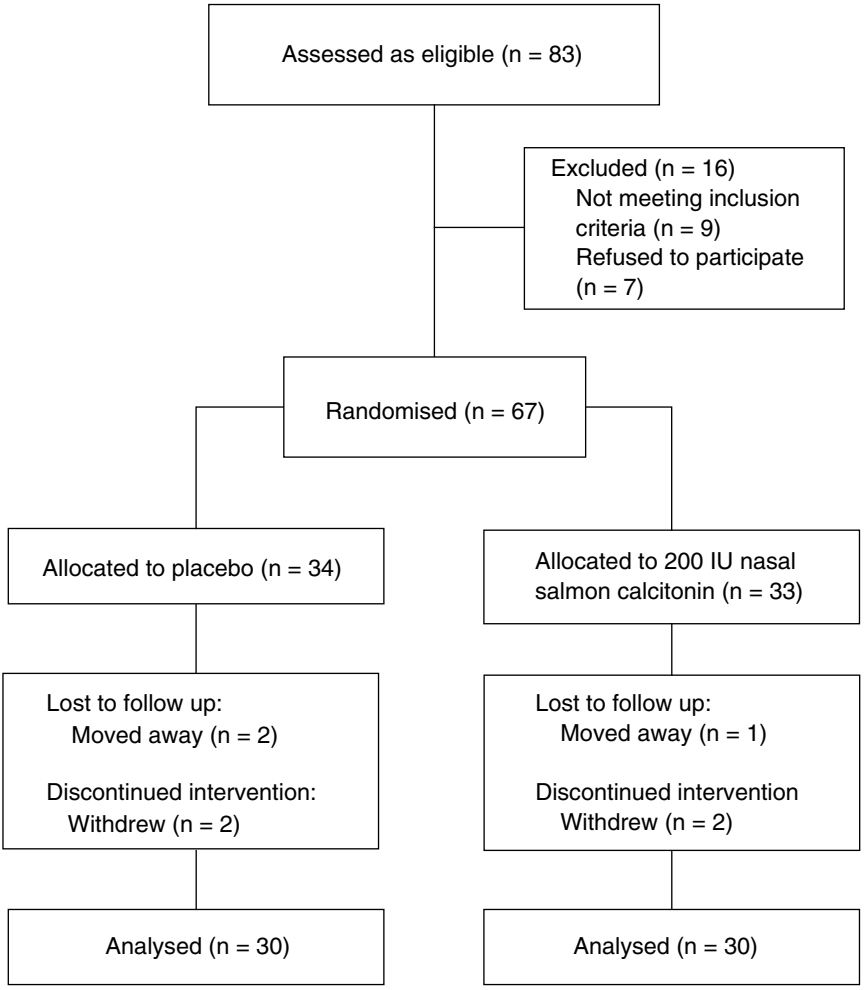


Figure 9.1. Sample flow chart describing the progress of subjects through a randomized trial.

the chart can be applied to more complex trials with appropriate modification. The CONSORT statement will help you review the quality of published clinical trials and in writing the report of your trial.

9.3. The Title

The title of your report should factually describe the nature of the trial. To get a snappy title and grab the attention of readers, the title itself often introduces bias by overselling the interpretation, or power, of the study (known as "flashy title" bias[4]). An example of a "flashy title" might be "Calcitonin reduces fracture risk." This makes the assumption that BMD is an adequate surrogate measurement for fracture risk. It does not mention the study population (in this case postmenopausal women) and so implies that it is generally applicable. The title ignores the fact that a short-term trial in postmenopausal women close to the menopause is unlikely to have any impact on the long-term fracture risk in the elderly. The title should include the facts (if appropriate) that the trial is randomized, blinded, whether it is placebo-controlled or an active comparator trial, and a description of the subject group. In the case of the example data, the title of the paper could be as follows: "A randomized, double-blind, placebo-controlled trial of nasal salmon calcitonin in the prevention of bone loss in perimenopausal women." Readers will immediately be able to assess the intervention, outcome, and study group and have some assurance that the trial was conducted in a proper manner to a proper study design. First impressions count!

9.4. Abstract

First impressions are so important that the title and abstract of the paper are often the only parts that readers ever read thoroughly. There will be a quick glance at the pictures and a scan of the conclusion. If the paper looks interesting, it will get photocopied and be put in the reading pile, only to be moved deeper into the filing system at a later stage. It is vital, therefore, that the abstract is structured in such a way as to get across the main facts of the paper, including the magnitude of the treatment differences. This will reduce the length of time that a reader requires to make a critical appraisal of your paper and enable accurate searching of published abstracts when carrying out a structured review.

The Ad Hoc Working Group for Critical Appraisal of the Medical Literature proposed guidelines for structuring abstracts.[5] The guidelines propose dividing the abstract into seven sections, as follows:

1. *Objective*. What is the objective or question addressed in the paper?
2. *Design*. Describe the study design. Is it randomized, blinded, or controlled? Is it a cross-over trial? Is it case-controlled, a survey, a cost–benefit or cost-effectiveness analysis?

3. *Setting*. It is important to describe the context of the study so that the reader can assess whether it is applicable in their own circumstances. Was it conducted in primary or secondary care?
4. *Subjects/other participants*. Describe the subject group studied, the number of participants, including how many were eligible and refused to take part, the number of withdrawals and the number completing the study. Include the number of subjects withdrawn because of adverse events and summarize the nature of those events. Summarize the selection procedures (e.g. random, consecutive cases, or volunteers) and major inclusion and exclusion criteria.
5. *Intervention*. Describe the duration and method of administration of the main intervention, using the generic and brand names of drugs used.
6. *Measurements and main results*. Describe the main measurements used in the study and provide an explanation of the measurement if a novel or unusual measurement is made. Describe the results. Report nonsignificant findings in the abstract as well, to avoid bias. In a survey of three reputable journals, it was found that 70% of significant findings were reported in the abstract compared with only 25% of nonsignificant findings.[1] Report the statistical significance of the results, quoting the actual significance level rather than an arbitrary cut-off.
7. *Conclusion*. The study conclusion should be supported by the main results quoted in the abstract. State if further trials are required before the NME is used in routine clinical practice for the clinical indication described in the paper.

The nonstructured and structured abstracts for the example study are shown in Table 9.1 and Table 9.2. The structured abstract is longer than the unstructured abstract, but describes the content of the paper in sufficient detail to enable the reader to determine that this paper will be of interest. It avoids "flashy title" bias and does not make unsubstantiated claims about the benefits of calcitonin.

Table 9.1. Nonstructured abstract.

Calcitonin reduces fracture risk

Seventy-nine postmenopausal women with no history of bone disease were recruited using adverts in the local press, local radio and GP practices onto a randomized, double-blind, placebo-controlled trial of Calcitonin. Subjects were treated for 4 years and BMD of the lumbar spine and femur was measured using DXA at baseline, 1, 2, 3 and 4 years. Quantitative ultrasound of the calcaneus was also measured using a Lunar Achilles Plus. BMD and QUS rose significantly in the treatment group when compared with the control group ($p < 0.01$), who lost bone during the study. The increase in BMD represents a 30% decrease in fracture risk in the treatment group. Calcitonin has a significant role to play in reducing fracture risk in postmenopausal women.

Table 9.2. Structured abstract.

A randomized, double-blind, placebo-controlled trial of nosal salmon calcitonin in the prevention of bone loss in postmenopausal women.
Study Objective: To determine the efficacy of nasal salmon calcitonin in preventing bone loss in post-menopausal women.
Design: Randomized, double-blinded, placebo-controlled trial with a 4-year treatment period.
Setting: A population based study recruiting subjects by adverts in the press, local radio and primary care clinics.
Subjects: A population based sample of 67 postmenopausal women aged 46 to 59 on entry to the trial. Women with osteoporosis at the screening visit and diseases affecting bone metabolism were excluded from the trial. 60 subjects completed the study. No subjects were withdrawn due to adverse events.
Interventions: 200 IU nasal salmon calcitonin (tradename) over 4 years. All subjects were offered calcium and vitamin D supplementation.
Measurements and Main Results: Bone mineral density (BMD) of the lumbar spine and femur were measured using dual energy X-ray absorptiometry (DXA). Broadband ultrasound attenuation (BUA), speed of sound (SOS) and stiffness were measured in the calcaneus using quantitative ultrasound (QUS). There was a significant increase of BMD in favor of Calcitonin over placebo at the lumbar spine (11.4%, $p < 0.01$), total hip (7.4%, $p < 0.01$) and of QUS at the calcaneus (10.1%, $p < 0.01$).
Conclusions: Calcitonin taken over 4 years results in an improvement in BMD and QUS over placebo. QUS can be used in clinical trials to measure treatment effects. Larger trials that monitor fracture rates in postmenopausal women are required to assess the impact of Calcitonin on fracture prevention.

9.5. What Should Be Included in the Introduction?

That first sentence! How we struggle over the wording of that first sentence. It is usually a warm, comforting phrase, designed to capture the readers' attention and draw them into the rest of the paper. It often introduces bias by overstating the size of the research question. "Osteoporosis-related fracture is a growing cause of mortality and morbidity, affecting 40% of women over the age of 70 years and costing the National Health Service in the UK an estimated £940 million per annum." True, but your small trial is not going to solve that problem overnight. The introduction should, again, be factual and state the *prospectively* defined research hypothesis.

9.6. Method

The CONSORT statement is particularly helpful in structuring the method section. The planned study population and way in which subjects were approached should be described in outlining the protocol. Was it by random selection from a general practice list? Was it by invitation through an advertisement in the clinic? Were consecutive subjects who met the inclusion criteria of the study approached in clinic? The method of recruitment can introduce bias and reviewers will find it easier if there is a clear description of the recruitment process. In the sample study, normal postmenopausal women were recruited through advertisements placed in

surgeries, clinics, the local press, and on the radio. They were invited to attend an information session about osteoporosis and the proposed research. Those who expressed an interest at that stage were given detailed subject information and consented into a screening stage for the trial. The screening stage included clinical work-up, BMD measurement, and biochemical tests. If subjects met the inclusion criteria, they were then randomized to treatment with calcitonin or placebo. Describe also the inclusion and exclusion criteria in detail.

The primary outcome measure was BMD of the lumbar spine. Depending on the audience for your publication, the method for measuring BMD might need to be described in some detail. If the paper is to be published in a journal of which the readers are familiar with the measurement techniques, less detail need be given.

BMD was measured to have a standard deviation (SD) of 0.12 g/cm^2 and the mean difference between treatment and control groups was 0.108 g/cm^2, which was assumed to be of clinical relevance. Assuming a probability of detecting a significant difference, if calcitonin is no more effective than placebo, of 0.05 (an α or *type I* error) and a probability of not detecting a significant difference, if one exists (a β or *type II* error), of 0.10, 26 subjects were required in each arm of the study. Allowing for withdrawals, the study aimed to recruit 34 subjects to each arm of the study. Results are given in Appendix B for the 30 subjects who completed the study.

The method should also describe the proposed statistical analysis. This should include more than just the name of the statistics package used to carry out the analysis, for example a description of the statistical methods used and why they were chosen. In this case, as described in Section 9.7 below, analysis planned to examine the normality of study data using the Shapiro–Francia W′ test, with the use of analysis of variance (ANOVA) to compare the data at baseline for differences between treatment groups and centres. One of two possible ANOVA models was to be used to analyse the longitudinal outcome data. Where data were not normally distributed, a Kruskal–Wallis test was used to compare variables at baseline for group and centre effects and outcome at the end of the study.

9.6.1. Intention-to-Treat Analysis and Missing Values

A discussion of the treatment of missing values and whether the analysis was carried out on an intention-to-treat basis should also be included in the method. Subjects leave clinical trials at all stages and for many reasons (or no reason at all), even in a well-designed and well-run trial. They might leave after randomization and before treatment if it is found that they do not meet the inclusion criteria. They might drop out of the trial for valid reasons or treatment could be stopped because of adverse events. It is important to document the number of subjects recruited, the number of subjects randomized, and the number of subjects who can be evaluated. This last category should be decided during study design and might be only subjects who complete the study, those who have an acceptable number of missing values, or those who have completed a minimum number of observations. An intention-to-treat analysis includes all subjects that the investigator intended to treat but who

might have dropped out of the study for a variety of reasons. The argument is that this will give a more realistic view of the treatment effect in real life, because subjects will fail to comply with treatment in the clinical setting. There must be sufficient outcome data that can be evaluated to proceed with the analysis. Outcome data must be used carefully, because the handling of missing values can bias the outcome and lead to an overestimate or underestimate of the treatment effect.

Missing values can be handled in a number of ways. If it is the case, for example, that a subject was measured at 1 year posttreatment and 3 years posttreatment, but missed the 2-year visit, some form of interpolation is acceptable. Linear interpolation is the simplest model. If the missing values are at the end of the study owing to subject drop-out, a measurement of the primary outcome variable must be made, if possible, or the method of carrying forward the last observation can be used. The latter method is biased if the subjects dropped out because of side effects or adverse events, because the treatment effect will be overestimated. If a subject drops out of the study early, carrying forward the last observation might be invalid without supporting follow-up information. For example, in a bisphosphonate-based study, in which much of the gain in BMD is during the first 9 months of treatment, it could be valid to carry forward the 1-year observation to the end of the study. If the gain in BMD is slower and occurs over a longer period of time, the 1-year measurement can give an inadequate estimate of the treatment effect. Again, judgement must be used, according to knowledge of similar NMEs and their effect on BMD.

9.6.2. Randomization and Blinding

The aim of randomization is to ensure that there is a similar distribution of baseline variables in the treatment and control groups and that unknown factors affecting the outcome of the trial are evenly spread. Various methods of randomization have been described in Chapter 2 and should be reported in the method section of the study. How has randomization been carried out? In a multicentre trial, randomization is often performed centrally, but subjects can be stratified within the randomization on the basis of centre or baseline variables that might affect outcome (e.g. age or BMD). The details of the method used to generate the allocation to each group and the method by which the investigators are blinded to that allocation should be described. For example, in our sample study, block-randomization was used. The randomization codes were generated and held by the pharmacy department on behalf of the investigators. Subjects were randomized when they attended the pharmacy to collect their medication for the first time. This enabled the randomization to be blinded to both the investigator and the subject.

The mechanism of delivery of the NME and control preparations should be described (e.g. capsule, tablet, or patch), in addition to the similarity in appearance between the control and the NME preparation. This could include the taste and packaging of both preparations.

Any evidence that demonstrates the quality of the blinding should be included, including evidence from the subjects themselves, the investigators, and those

assessing outcome [e.g. the dual-energy X-ray absorptiometry (DXA) technologists]. Although many sponsors insist on DXA scans being analysed by an independent clinical research organization (CRO), it is an important part of DXA quality control (QC) to carry out an analysis of the scan result while the subject is still present to ensure there are no technical difficulties with the scan. For repeat scans, this often includes using a comparison facility within the DXA software. The technologist can then be unblinded if there is a consistent improvement in BMD in treated subjects.

9.6.3. Other Methodological Issues

It is important to include other issues in the method section that could bias the outcome of the study. This could include a discussion of the appropriate choice of equipment used to measure BMD. Was there a rationale for measuring BMD at the chosen site using the chosen technology? Are there any crosscalibration issues between centres in a multicentre trial? How was the crosscalibration carried out (see Chapter 4)? How were equipment failures and QC failures handled? Guidance is given in Chapter 8 (see Section 8.2.8) on how to handle such failures. It is important for investigators to recognise possible bias in study design and execution and include this in the paper. There are a large number of sources of bias,[4,6] and to acknowledge them in the paper provides evidence to other investigators that the trial has been thoughtfully designed.

Many journals will not publish without a reference to approval by the Institutional Review Board or Independent Ethics Committee (IRB/IEC) and description of the informed consent process. This, and any other ethical considerations that arise from the trial, should be included in the method section.

9.7. Results

The main problem with the presentation of results is that there are two standards, as follows:

1. The results, as the statistician insists, are correct, but no one understands the words between the pictures.
2. The results as generally presented in a form understood by clinicians as the lingua franca of osteoporosis trials, but not statistically correct. An example would be multiple testing of repeated measures data, comparing the percentage change in BMD at 1 year and 2 years to baseline using multiple single sample *t*-tests.

It is important that the results of the study are reported in a manner that is both statistically robust and clearly understood by the readers. The aim of this section is to attempt this using the sample data provided.

At the start of the results section, summarize the progress of subjects through the study. Use the CONSORT flow chart to help you (Figure 9.1). The inability to recruit eligible subjects, for example, might indicate a problem with the complexity of the trial design.

On entry to the trial, it is important to ensure that the control and treatment groups are the same and there are no centre effects in a multicentre trial (e.g. the age of subjects recruited at one centre is very different to those recruited at the other centres). This can be achieved with an analysis of the baseline variables that describe the demographics (e.g. age or body-mass index [BMI]) and severity of disease (e.g. baseline BMD or biochemistry). This includes critical variables that are likely to affect the response to treatment or will be used as primary or secondary outcome measures (e.g. fracturehistory, BMD, or bone biochemistry) and other risk factors that could affect outcome (e.g. smoking or alcohol intake). There is always a temptation to collect too much information rather than too little. Take care in the selection of baseline comparisons, because multiple significance testing can confound the interpretation. Use only variables that have some rationale in relation to the study, as outlined above.

Categorical variables can be compared using simple Chi-squared or Fisher's exact tests. There is, for example, no significant difference in smoking between treatment and control groups in our study sample (Table 9.3).

When considering continuous variables, such as baseline BMD, begin with a test of the normality of the data. If there are sufficient subject numbers, test for the normality of the data within each centre. There are a number of tests that can be used. Skewness (a measure of the asymmetry of the distribution) and kurtosis (a measure of how pointed a distribution is) can be used, although both are susceptible to bias because of outliers. The skewness (g) divided by the standard error of skewness is distributed according to the t distribution with $n - 1$ degrees of freedom, where n is the number of subjects within the group tested. These are calculated as follows:

$$g = \sum_i \frac{(x_i - \bar{x})^3}{(n-1)\sigma^3} \quad \text{and} \tag{9.1}$$

$$SE_g = \sqrt{6/n}. \tag{9.2}$$

If the t statistic is significant, the distribution is not normal and nonparametric statistics should be used. Kurtosis and the standard error of kurtosis are calculated as follows:

$$k = \sum_i \frac{(x_i - \bar{x})^4}{(n-1)\sigma^4} \quad \text{and} \tag{9.3}$$

$$SE_k = \sqrt{24/n}. \tag{9.4}$$

Table 9.3. Breakdown of subjects by smoking status.

	Control group	Treatment group
Smokers	12	15
Nonsmokers	18	15

$\chi^2 = 0.606$; $p = 0.44$. There is no significant difference between control and treatment groups.

Another method is to use the Shapiro–Francia W′ test,[7] which uses a plot of the normal scores against the observed data. The normal score for each data point is calculated as the standardized normal deviate for each data point. First, assemble the data in ascending order. Then, calculate the expected cumulative frequency of each data point, as follows:

$$P_i = \frac{(i - 3/8)}{n + 1/4}. \tag{9.5}$$

The normal score N_i, for the ith data point in a series, is the number of SDs above or below the mean at which the data point would be expected to lie given the cumulative frequency of P_i. N_i can be found in tables or calculated in Microsoft Excel® using the NORMSINV function. The correlation coefficient of N_i against x_i is calculated. W′ is the square of the correlation coefficient. The closer the value of W′ is to 1, the more normal the distribution. W′ has been tabulated to give the probability of the null hypothesis that W′ is equal to 1,[7] with small values of W′ indicating that the distribution is not normal. Figure 9.2 is the normal score plot for broadband ultrasound attenuation (BUA) from the sample data. The W′ is 0.914. From tables, the probability of the null hypothesis that W′ is 1.0 is $p < 0.01$ and, therefore, the distribution is nonnormal. This is confirmed by the histogram plot (Figure 9.3) and the t statistic calculated from skewness, 3.36 ($p < 0.002$), which is clearly highly significant. A plot of lumbar spine BMD is shown in

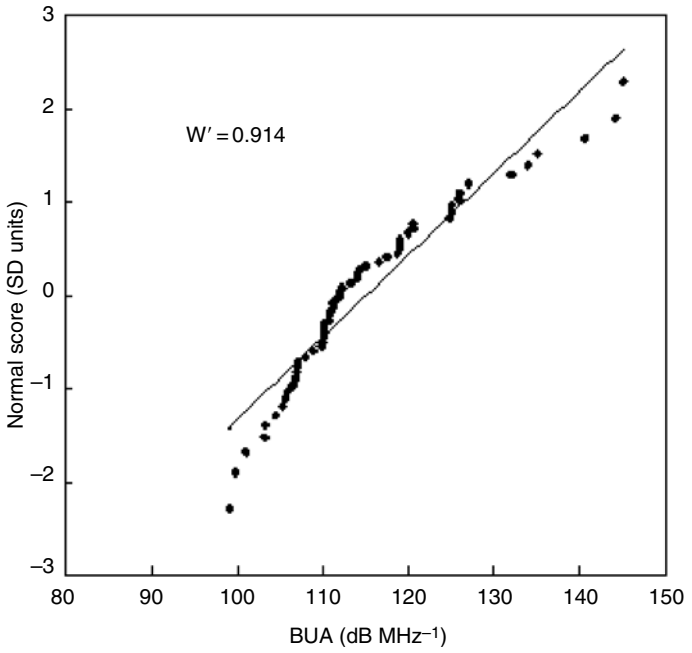


Figure 9.2. Normal score plot of BUA from the sample data.

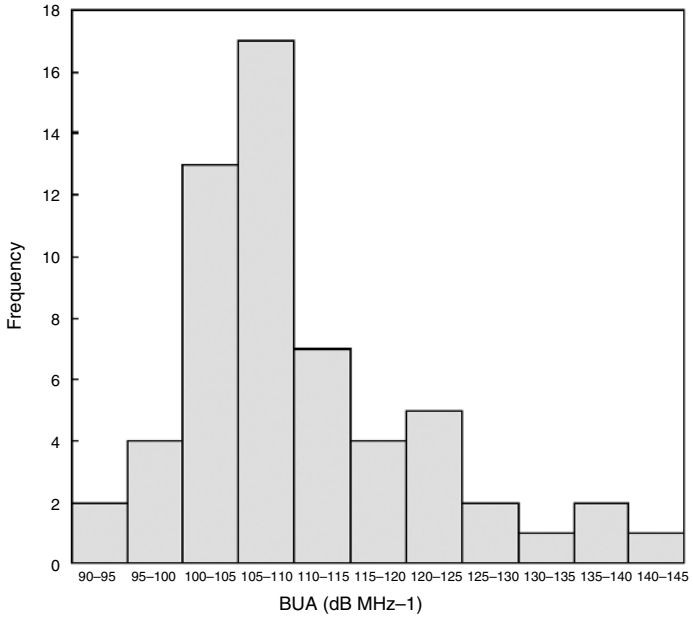


Figure 9.3. Histogram of BUA data at baseline.

Figure 9.4, where W′ is 0.967 ($p > 0.05$) and, therefore, the data is normal. This is confirmed by the histogram plot (Figure 9.5) and by the t statistic calculated from the skewness, 1.85 ($p = 0.07$), which is not significant.

If the data do not follow a normal distribution, nonparametric statistics must be used. These are detailed in Section 9.7.2.

ANOVA of baseline variables, with the subject group and centre as factors, can be used to test for baseline difference. The model used to test this is as follows:

$$Y_{ijk} = m + G_1 + C_k + e_{ijk} \tag{9.6}$$

where Y_{ijk} is the response of the jth subject in the ith subject group at the kth centre, μ is the overall mean, G_i is the effect of ith subject group, C_k the effect of the kth centre, and ε_{ijk} is the random error associated with measuring Y. Let there be p subject groups, m subjects in total, and n centres. The ANOVA is given in Table 9.4.

For the sample data, there is no significant effect of the group or centre on either age or BMD at baseline (Table 9.5). If there is a significant effect of either centre or age, a post-hoc range test will identify where the differences lie. There are a number of range tests that can be used, including the Bonferroni method, the Scheffe method, the Tukey method, and Duncan's multiple range test.[7] They are all variations on the theme of the multiple t-test, which is corrected for the number of

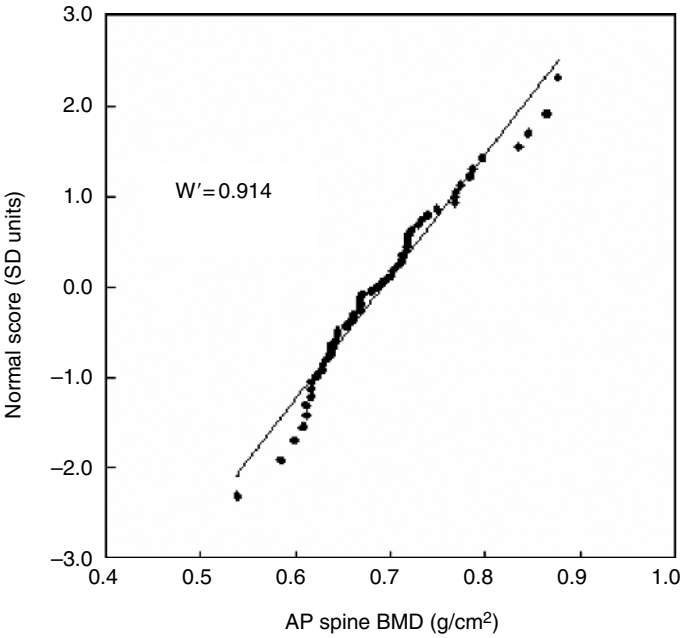


Figure 9.4. Normal score plot of lumbar spine BMD data from the example study.

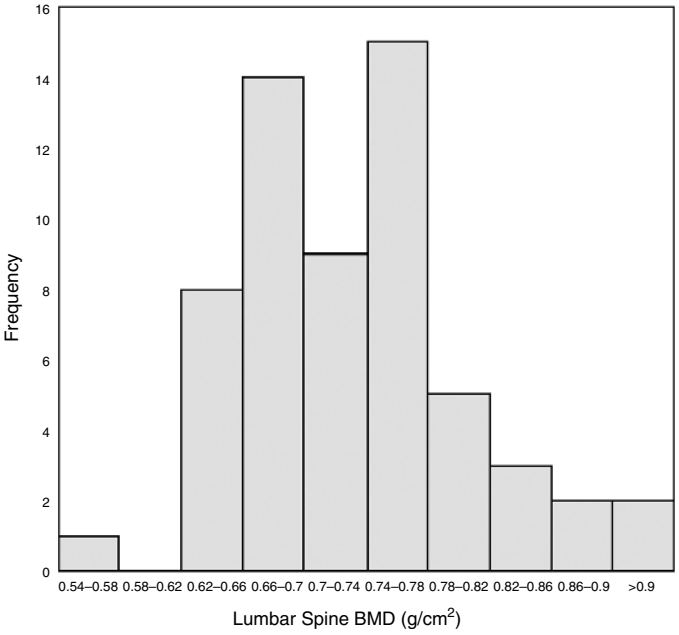


Figure 9.5. Histogram of BMD at the AP lumbar spine at baseline.

Table 9.4. The ANOVA model used to test baseline differences.

Source of variance	Sum of squares	*df*	Mean square	*F*
Group	$SSG = nm\sum_i(\overline{Y}_i - \overline{Y})^2$	$p - 1$	$MSG = \frac{SSG}{p - 1}$	$\frac{MSG}{MSE}$
Centre	$SSC = pm\sum_k(\overline{Y}_k - \overline{Y})^2$	$n - 1$	$MSC = \frac{SSC}{n - 1}$	$\frac{MSC}{MSE}$
Residual	$SSE = SST - SSC - SSG$	$m - p - n + 1$	$MSE = \frac{SSE}{m - p - n + 1}$	
Total	$SST = \sum_i\sum_j\sum_k(Y_{ijk} - \overline{Y})^2$	$m - 1$		

Table 9.5. The ANOVA model used to test differences in BMD at baseline between groups and between centres.

Source of variation	Sum of squares	*df*	Mean square	*F*	Significance of *F*
Group	0.00019	1	0.00019	0.015	$p = 0.904$
Centre	0.00963	1	0.00963	0.753	$p = 0.389$
Residual	0.72874	57	0.01278		
Total	0.73855	59			

comparisons made. The simplest method is to calculate a *t* value for the comparison as follows:

$$t = t_{p,df}\sqrt{SSE^2\left(\frac{1}{m_1} + \frac{1}{m_2}\right)}, \tag{9.7}$$

where *SSE* is the residual sum of squares from the ANOVA and $t_{p,df}$ is the *t* value for the desired level of significance, *p*, and the residual degrees of freedom, *df*, in the ANOVA. The level of significance is corrected for the number of possible comparisons that can be made. In this case, with two groups and two centres, there are six possible comparisons. Thus, instead of using a significance level of $p = 0.05$, a significance level of $p = 0.0083$ is used. If the difference in mean BMD between the control group at centre A and the treatment group at centre B, for example, is greater than the *t* value for comparison, there is a significant difference between those two groups.

The investigator then has to consider the reason that any significant difference has occurred and they must make a judgement about the clinical significance of the difference. Is there a difference in more than one baseline variable and is it consistent for a particular centre or group? This would be a cause of concern and the validity of the trial must be questioned. If the difference is in only one or two of the baseline variables and is not consistent between groups and centres, a judgement has to be made regarding whether the analysis can continue. Most trials are robust enough to cope with small differences between the groups at baseline,

unless the difference is in one of the primary outcome variables or any variable that is likely to affect the treatment response. There is no clear-cut answer to this, but there should be a debate by the Data Monitoring Committee to consider the impact on the trial.

9.7.1. Analysis of Longitudinal Outcome Data

The common method of analysing and reporting longitudinal outcome data is to report the percentage change from baseline and carry out multiple *t*-tests to compare both the outcome with baseline and treatment with control. Figure 9.6 shows a typical graph reporting the percentage change in BMD and the significant differences demonstrated by using multiple *t*-tests. The graph shows the mean ±1 SD. The advantages of using percentage change are that it seems to get around the problem of small differences in equipment calibration between centres and copes with the large differences in BMD between subjects. This enables data from multiple sites to be pooled easily. There are three problems with this approach. First, the percentage change in BUA is not normally distributed. The range of percentage change is from −100% (if the final measurement is much smaller than baseline) to infinity (if the final measurement is much larger than baseline). It

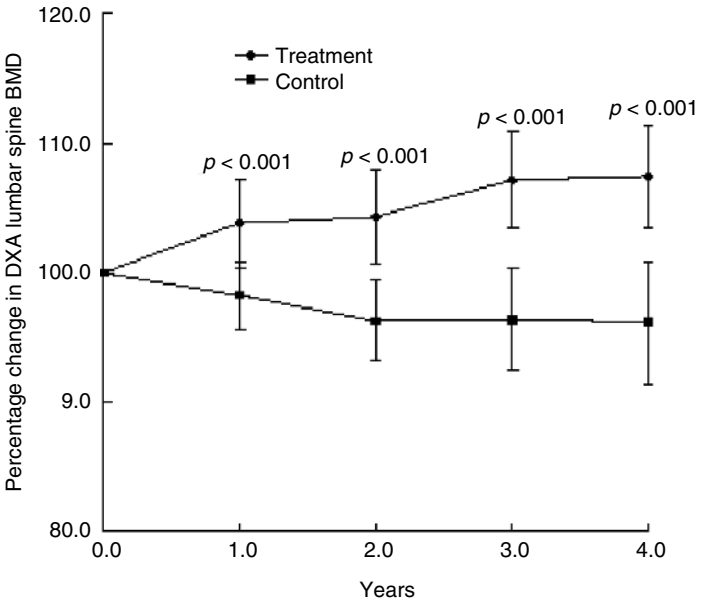


Figure 9.6. Percentage change in BMD from baseline (mean ± 1 SD).

could be argued that, over the range of changes noted in BMD of plus or minus a few percentage points, this does not matter, but it is better to describe the percentage change using medians and quartiles rather than means and SDs. Nonparametric statistics should be used. Second, it assumes that the change is linearly related to the baseline measurement. This should be checked if this is the chosen method of analysis. Third, and most importantly, multiple significance tests result in a greater risk of a significant difference being highlighted if a significant difference does not exist.

The statistician would suggest that only the final outcome measurement must be analysed to demonstrate a treatment effect. The clinician, however, then asks when the change in BMD occurred and when it became significant, because this will affect the monitoring period for BMD when the NME passes from the research stage into routine clinical use. An alternative is to summarize the data using the absolute change, peak change, or rate of change before analysis, but this suffers from the same problem as using the final outcome variable. There are two methods of ANOVA that come to our rescue. Both give the same answer and can be implemented in a simple spreadsheet or many of the common statistics packages available.

The main source of variation in BMD data is the variation between subjects. The changes in BMD with time or treatment are relatively small compared with the range of BMD values in the subject group. The hope with an NME is that BMD will rise in the treatment group and fall or remain stable in the placebo or active comparator group. This means that the groups behave in a different way as time progresses. This is known as an "interaction effect" and investigators are looking to see if this interaction effect is significant. The model that accounts for the variation between subjects and the interaction between treatment and time is as follows:

$$Y_{ijk} = \mu + T_I + P_{j(i)} + (TV)_{ik} + \varepsilon_{ijk}, \tag{9.8}$$

where Y_{ijk} is the response of the jth subject on the ith treatment at the kth visit, μ is the overall mean, T_i is the effect of the ith treatment, $P_{j(i)}$ is the effect of the jth subject within the ith treatment (the between-subject variation), V_k is the effect of the kth visit, $(TV)_{ik}$ is the effect of the interaction between time and treatment, and ε_{ijk} is the random error in measuring Y. The model assumes that ε_{ijk} is independent of $P_{j(i)}$, i.e. there is no relationship between the BMD and the error in measuring BMD. The between-subject variation will also include small differences in equipment calibration between centres. The ANOVA for this model is given in Table 9.6 and the ANOVA for lumbar spine BMD is given in Table 9.7. Note that the majority of the variation is explained by the between-subject variation, that is the variation in BMD between individual subjects. The other important statistic in the table is the interaction term between treatment and visit. This is significant, demonstrating that there is a significant treatment effect.

The alternative model is to use an analysis of variance and covariance. The analysis of covariance assumes that the BMD is linearly related to variables

Table 9.6. ANOVA for the model in Equation 9.8.

Source of variance	Sum of squares	*df*	Mean square	*F*
Treatment	$SST = nm\sum_i(\overline{Y}_i - \overline{Y})^2$	$p - 1$	$MST = \frac{SST}{p-1}$	$\frac{MST}{MSP(T)}$
Subject (treatment)	$SSP(T) = n\sum_i\sum_j(\overline{Y}_{ij} - \overline{Y}_i)^2$	$p\,(m-1)$	$MSP(T) = \frac{SSP(T)}{p(m-1)}$	
Visit	$SSV = pm\sum_k(\overline{Y}_k - \overline{Y})^2$	$n - 1$	$MSV = \frac{SSV}{n-1}$	$\frac{MSC}{MSE}$
Treatment by visit	$SS(TV) = \sum_i\sum_j\sum_k(\overline{Y}_{ik} - \overline{Y}_i - \overline{Y}_k + \overline{Y})^2$	$(p-1)\,(n-1)$	$MS(TV) = \frac{SS(TV)}{(p-1)(n-1)}$	$\frac{MS(TV)}{MSE}$
Residual	$SSE = \sum_i\sum_j\sum_k(\overline{Y}_{ijk} - \overline{Y}_{ij} - \overline{Y}_i)^2$	$p(m-1)\,(n-1)$	$MSE = \frac{SSE}{p(m-1)(n-1)}$	
Total	$SST = \sum_i\sum_j\sum_k(\overline{Y}_{ijk} - \overline{Y}_i)^2$	$pmn - 1$		

Table 9.7. The ANOVA from Table 9.7 using the lumbar spine BMD data from the example study.

Source of variance	Sum of squares	*df*	Mean square	*F*	Significance of *F*
Treatment	0.364	1	0.3637	5.82	$p = 0.019$
Subject (treatment)	3.625	58	0.0625		
Visit	0.0143	4	0.00358	7.94	$p < 0.001$
Treatment by visit	0.1107	4	0.0277	61.48	$p < 0.001$
Residual	0.1044	232	0.00045		
Total	4.218	299			

measured at baseline. These can either be baseline variables that affect BMD (e.g. age or BMI) or, more simply, use the baseline BMD as the covariate. The model for this analysis is as follows:

$$Y_{ijk} = \mu + T_i + \beta X_{ijk} + V_k + (TV)_{ik} + \varepsilon_{ijk}, \tag{9.9}$$

where the components of the model are as defined in Equation 9.8, the baseline measure of response. To calculate the ANOVA, it is necessary to calculate the following:

$$S_{xx} = \sum_i\sum_j\sum_k(X_{ijk} - \overline{X})^2,$$

$$S_{xy} = \sum_i\sum_j\sum_k(Y_{ijk} - \overline{Y})(X_{ijk} - \overline{X}),$$

$$
\begin{aligned}
S_{xy} &= \sum_i \sum_j \sum_k (Y_{ijk} - \overline{Y})^2, \\
T_{xx} &= m \sum_i \sum_k (X_{ik} - \overline{X})^2, \\
T_{xy} &= m \sum_i \sum_k (Y_{ik} - \overline{Y})(X_{jk} - \overline{X}), \\
T_{xy} &= m \sum_i \sum_k (Y_{ik} - \overline{Y})^2, \\
E_{xx} &= S_{xx} - T_{xx}, \\
E_{xy} &= S_{xy} - T_{xy}, \\
E_{yy} &= S_{yy} - T_{yy}.
\end{aligned} \tag{9.10}
$$

The slope of the regression is given by the following equation:

$$\beta = E_{xy}/E_{xx}. \tag{9.11}$$

The analysis of variance is given in Table 9.8. For BMD at the lumbar spine in the sample data, the ANOVA is given in Table 9.9. Many of the figures are the same as for the previous model. The variance associated with the treatment is reduced because most of this is included in the regression variance. If the regression is nonsignificant, the first ANOVA model should be used.

Table 9.8. ANOVA for the model in Equation 9.9.

Source of variance	Sum of squares	*df*	Mean square	*F*
Treatment	$SST = nm\sum_i (\overline{Y}_i - \overline{Y} - b(\overline{X}_i - \overline{X}))^2$	$p - 1$	$MST = \frac{SST}{p - 1}$	$\frac{MST}{MSP(T)}$
Visit	$SSV = pm\sum_k (\overline{Y}_k - \overline{Y})^2$	$n - 1$	$MSV = \frac{SSV}{n - 1}$	$\frac{MSC}{MSE}$
Treatment by visit	$SS(TV) = \sum_i \sum_j \sum_k (\overline{Y}_{ik} - \overline{Y}_i - \overline{Y}_k + \overline{Y})^2$	$(p - 1)(n - 1)$	$MS(TV) = \frac{SS(TV)}{(p - 1)(n - 1)}$	$\frac{MS(TV)}{MSE}$
Regression	$SSR = \frac{E_{xy}^2}{E_{xy}}$	1	$MSR = SSR$	$\frac{MSR}{MSE}$
Within plus residual	$SSE = \sum_i \sum_j \sum_k (\overline{Y}_{ijk} - \overline{Y}_{jk})^2 - SRR$	$pnm - pn - 1$	$MSE = \frac{SSE}{pnm - pn - 1}$	

Table 9.9. The ANOVA from Table 9.9 using the lumbar spine BMD data from the example study.

Source of variance	Sum of squares	*df*	Mean square	*F*	Significance of *F*
Treatment	0.3288	1	0.3288	373.75	$p < 0.001$
Visit	0.0143	4	0.0036	4.06	$p = 0.003$
Treatment by visit	0.1107	4	0.0277	31.45	$p < 0.001$
Regression	3.4746	1	3.4746	3949.63	$p < 0.001$
Within plus residual	0.2542	289	0.0009		

The mean BMD for each group at each visit can then be recalculated to take into account the regression. These are known as adjusted cell means and are calculated as follows:

$$\overline{Y}'_{ik} = \overline{Y}_{ik} - \beta(\overline{X} - \overline{X}_i). \tag{9.12}$$

Both these models assume an equal number of subjects in each group and at each visit. They can be generalized to cope with unequal group sizes. The sum of squares for the treatment in Table 9.7, for example, would become the following:

$$SST = n\sum_i m_i(\overline{Y}_i - \overline{Y})^2, \tag{9.13}$$

where m_i is the number of subjects in the ith group. The handling of degrees of freedom is a matter for some debate. The commonest method seems to be to use the harmonic mean of the numbers in each group. This is calculated as follows:

$$\frac{i}{\left(\sum_i \frac{1}{m_j}\right)} \tag{9.14}$$

These ANOVA models are available in most personal computer based statistics packages and can be developed within a spreadsheet if necessary. Statistics packages can generally deal with unequal numbers in each group.

Post-hoc methods, described earlier in Section 9.7, can then be applied to discover where the significant differences lie.

Figure 9.7 is the graphical representation of the outcome of using the second of these two ANOVA models. The baseline data are plotted as mean ± 1 SD and the other data are plotted as the adjusted cell means. There is a significant treatment effect ($p < 0.001$), a significant effect of visit ($p = 0.003$), and a significant interaction term ($p < 0.001$), showing that treatment and control groups respond differently with time. Applying the post-hoc methods shows that the difference between the groups is significant as early as 1 year posttreatment ($p < 0.001$).

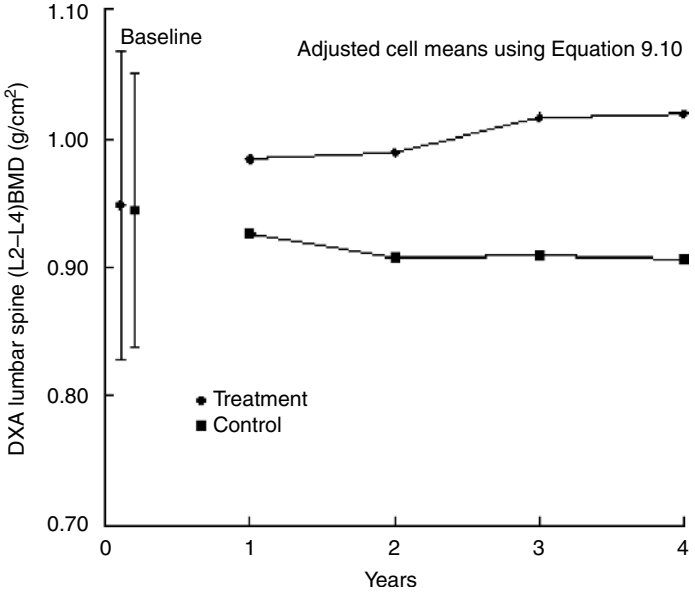


Figure 9.7. Absolute lumbar spine BMD plotted against time for the example study. The baseline values are plotted as mean ± 1 SD and the other values are plotted as the adjusted cell means calculated using Equation 9.12.

9.7.2. Nonparametric Statistics

The nonparametric equivalents of the ANOVAs used above can be applied to nonnormal data. It is acceptable to use a Wilcoxon rank sum test to compare the control and treatment groups. In this test, the data from the control and treatment groups are combined and ranked. The ranks for each group are then summed. If m and n are the number of subjects in the control and treatment groups, the sum of ranks in the treatment group is as follows:

$$r = \sum_{i=1}^{n} r_i. \tag{9.15}$$

The test statistic is then calculated as follows:

$$W = r - \frac{n(n+1)}{2}. \tag{9.16}$$

The acceptance region on the sum of ranks is tabulated,[8] in addition to methods of testing the significance of the test when the number of subjects is outside the tabulated values. The W statistic is also tabulated and the test statistic for large samples, corrected for ties, can also be calculated.[6] In our sample data, the baseline BUA has been ranked. The sum of ranks in the treatment group is 746 and

$W = 340$. The limits of acceptance of r are 691 to 933. Because r lies within these limits, there is no significant difference between control and treatment groups at baseline. The test statistic for W is as follows:

$$Z = \frac{W - nm/2}{\sqrt{\text{var}(W)}}, \tag{9.17}$$

where

$$\text{var}(W) = \frac{nm(n + m + 1)}{12}. \tag{9.18}$$

In this case, $Z = 1.054$; Z is a standardized normal deviate and is tabulated.[8] The probability of the two groups not being significantly different is 0.29. This test is available in most commercial statistics computer software packages. The extension of this test to more than two groups is known as the Kruskal–Wallis test.[6,9] In this test, the mean rank in each group is calculated and the test statistic H calculated as follows:

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^{k} n_i(\bar{r}_i - \bar{r})^2, \tag{9.19}$$

where n_i is the number of subjects in the ith group, k is the total number of treatment groups, N is the total number of subjects, $\bar{r}_i$ is the mean rank in the ith group, and $\bar{r}$ is the overall mean rank calculated as follows:

$$\bar{r} = \frac{N + 1}{2}. \tag{9.20}$$

If there is a significant difference between the groups, H will be greater than χ^2 for $k - 1$ degrees of freedom when there are a large number of subjects in the trial. Again, there are corrections to be made if there are a large number of tied ranks.

9.7.3. Treatment Effect

It is important to be able to quantify the effect of the treatment with the control group. The simplest way to calculate the treatment effect is to subtract the mean change from baseline for the control group from the mean change from baseline in the treatment group and calculate the 95% confidence interval on the difference. The significance of the treatment effect can then be tested. The difference is as follows:

$$d = \bar{Y}_t - \bar{Y}_c, \tag{9.21}$$

where $\overline{Y}_t$ and $\overline{Y}_c$ are the mean change from baseline in the treatment and control groups. The SD of the difference is as follows:

$$s = \sqrt{\frac{(n_t - 1)s_t^2 + (n_c - 1)s_c^2}{(n_t + n_c - 2)}}. \tag{9.22}$$

where s_c and s_t are the SDs of the control and treatment groups respectively and n_c and n_t the number of subjects in the control and treatment groups. The confidence interval is as follows:

$$CI = s \times \sqrt{\frac{1}{n_t} + \frac{1}{n_c}} \times t(0.025, n_t + n_c - 2) \tag{9.23}$$

and the significance of the treatment effect d is tested using the following statistic:

$$t = \frac{d}{s\sqrt{\frac{1}{n_t} + \frac{1}{n_c}}}, \tag{9.24}$$

with $n_t + n_c - 2$ degrees of freedom. In our example study, the mean change in BMD at the PA spine from baseline was 0.069 ± 0.035 g/cm^2 in the treatment group and -0.037 ± 0.045 g/cm^2 in the control group. The treatment effect was 0.108 g/cm^2, with a SD of 0.040 g/cm^2 and confidence interval of 0.024 g/cm^2. The t statistic is 10.2, with 58 degrees of freedom, which is highly significant.

In multicentre trials, it is important to calculate a treatment effect for each centre to ensure there are no centre differences.

9.7.4. Adverse Event Monitoring

The purpose of adverse event monitoring is to ensure that the adverse events associated with an NME are the same in the treatment and control groups, that is there are no significant adverse events that are owing to the study drug. Normally, investigators code adverse events into four categories, as follows:

1. Probably not related to the study drug
2. Possibly related to the study drug
3. Probably related to the study drug
4. Definitely related to the study drug.

It is simple, then, to compare the incidence of adverse events between treatment and control groups in these four categories using a chi-squared test. The type and severity of adverse events can also be coded and compared between treatment and control groups, because relying on the number of adverse events might be too crude a measure. It could be that, although the overall incidence of adverse events is the same in each group, the severity of the events differs. This can be vital evidence in an active comparator study that shows that the NME under investigation

has a lower rate of side effects than the current standard treatment, even if the effect on BMD or fracture risk is the same.

For example, in the Fracture Intervention Trial,[10] in which subjects were treated with alendronate, adverse events in the upper gastrointestinal tract were of particular interest because of the risk of oesophagitis when treating with an oral bisphosphonate. There were 1047 (47.2%) upper gastrointestinal tract events in the placebo group and 1052 (47.5%) events in the treatment group. Although there were 19 cases of oesophagitis in the treatment group, there were 10 cases in the control group. The relative hazard of oesophagitis was, therefore, 1.9 with 95% confidence limits of 0.9–4.26, implying there was no significant difference in the incidence of any upper gastrointestinal side effects.

9.8. Conclusions

Much of the interest surrounding the use of ultrasonography to measure BMD has come about because of overoptimistic statements about the technique included in the conclusions of publications. Statements that suggested ultrasound added additional information on bone quality were perpetuated because the variance in ultrasound results was not fully explained by the variance in BMD. Although there is evidence of a structural component in the ultrasound results from *in-vitro* studies, there was often little evidence to support these conclusions from the body of the publications. Over time, the position became well established. It is an example of poor discussion within publications. There is a temptation to include such statements—particularly when the results are not as good as hoped! They should be avoided.

The purpose of the concluding section of the publication is to put your study in the context of the available evidence surrounding its use in the treatment of osteoporosis and state the interpretation of the data based on the facts presented in the results section. The conclusions of the example trial used in this book are given in the structured abstract in Table 9.3. It is not necessary to expand these greatly to report the conclusions from the trial.

9.9. Summary

The structured reporting of clinical trials is an important part of disseminating the results from a study. The standards for structured abstracts and the CONSORT statement provide investigators with a template that is easy to follow and that is also easy to read. It enables other investigators easy access to the facts about your trial and will make your work of a standard that enables its inclusion in a future metaanalysis. The example used here is simple, and trials are often far more complex in their analysis if there are many centres involved and the study design is more complex. The statistics have been included, although in more detail compared with the rest of this text, to help investigators and others involved in clinical

trials work through the basics of testing data for normality, carry out baseline comparisons on the data, look for centre effects, and analyse longitudinal data in such a way as to answer the demands of clinical colleagues while maintaining the statistical moral high ground!

References

1. Pocock, S.J., Hughes, M.D. and Lee, R.J. (1987). Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* **317:**426–32.
2. Begg, C., Cho, M., Eastwood, S. *et al*. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* **276:**637–9.
3. Moher, D., Schulz, K.F., Altman, D.G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. **357:**1191–4.
4. Jadad, A.R. (1998). *Randomized Controlled Trials*. London: BMJ Books.
5. Ad Hoc Working Group for Critical Appraisal of the Medical Literature (1987). A proposal for more informative abstracts of clinical articles. *Ann Intern Med* **106:**598–604.
6. Chow, S.-C. and Liu, J.-P. (1998). *Design and Analysis of Clinical Trials*. New York: John Wiley Sons.
7. Altman, D.G. (1991). *Practical Statistics for Medical Research*. London: Chapman and Hall.
8. Lentner, C. (ed.) (1982). *Geigy Scientific Tables*, Volume 2, 8th edn. Basle: CIBA-GEIGY.
9. Siegel, S. and Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.
10. Cummings, S.R., Black, D.M., Thompson, D.E. *et al*. (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* **280:**2077–82.

Section 4

Background

10

Current Therapies for Osteoporosis

Jonathan R. Bayly

10.1. Introduction

This chapter concentrates on contemporary evidence for the effectiveness of pharmacological interventions to reduce fracture risk in subjects with osteoporosis. The management of such subjects is placed in the context of the pressures to contain the cost of healthcare while managing the personal and health economic consequences to subjects and society. It is argued that the financial consequences are underestimated in our current health economic models. Osteoporosis is a relatively new clinical area that is perhaps best viewed as a preventable and treatable risk factor for fragility fracture. Treatments with robust evidence from randomized, controlled trials (RCTs) have only become available in the last 20 years, and the most widely prescribed legacy treatment, in the form of hormone-replacement therapy (HRT), is rarely used nowadays. This chapter does not dwell on the public health issues associated with the prevention of osteoporosis and the attainment and maintenance of optimum bone health through exercise, diet, and avoidance of smoking and excess alcohol. These are important issues, but they are beyond the scope of this pharmacological view. The interaction between osteoporosis and falls is, however, briefly considered because both are risk factors for fracture.

10.2. Context

Osteoporosis is becoming a major healthcare problem because of its association with fragility fractures. A number of independent and semiindependent skeletal and nonskeletal risk factors enhance the ability of bone mineral density (BMD) to predict future fracture risk. Age is the greatest predictor of an individual's risk of fracture and we are set to see a rapid rise in incidence of fracture as the population of older people increases, with greater longevity. The personal and health economic burdens are huge. The clinical consequences of painful fracture cause

increased mortality, debility, dependence on social care, and a reduced quality of life. It is conservatively estimated that the health and social care consequences of osteoporotic fractures are as much as GB£1.8 billion annually (2.6 billion; US$3.3 billion).[1] This could be a significant underestimate because more recent research has suggested the in-patient costs of hip fracture might be more than double that used to arrive at the above figure and that hip fractures that involve nursing home care might also be far more costly than earlier estimates suggested.[2,3] In addition, fractures in subjects over 60 years old account for more than 2 million bed days each year in England alone.[4] Despite this, there is consistent evidence that, even the highest risk subjects, such as those with prior fragility fracture, are underidentified and undertreated in both primary and secondary care.[5,6]

Osteoporosis has been defined as "a systemic skeletal disease, characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture."[7] The diagnostic threshold has been based on measures of bone mass and the World Health Organization (WHO) defined osteoporosis as "a bone mass at the hip that is more than 2.5 standard deviations (SD) below the mean of a young woman at peak bone mass."

This age-independent measure of BMD came to be known as "the T-score."[8] It is important to remember that this diagnostic cut-off point was defined for epidemiological reasons[9] and is not, in itself, an intervention threshold. Indeed, over a reasonable timescale, the majority of fractures will occur in subjects who did not have osteoporosis at baseline.[10,11] Nevertheless, treatments will be discussed that have been shown to both improve BMD and reduce future fracture risk in subjects with osteoporosis both with and without a history of prior fragility fracture. Only a proportion of the fracture risk reduction is explained by improvement in BMD[12] and the pharmacological actions of these therapies are almost certainly more complex than just their ability to increase bone mass. A number of clinical risk factors seem to act as proxies of other characteristics of bone quality and help clinicians target therapies cost-effectively at those with the highest risk (Table 10.1). There are other considerations, such as bone geometry and biochemical markers of increased bone turnover, that are also associated with a higher fracture risk but they are not as yet sufficiently quantified to be useful in case finding. Probably the best way to make rational decisions about who to treat is look at their absolute fracture risk by site over a 5 to 10 year period of time and this is a current objective of the WHO. A fracture-risk assessment tool similar to those used to predict the risk of cardiovascular disease is under development.

10.3. Characteristics of Available Treatments

Therapies for osteoporosis are licensed for either prevention or treatment, or both. This distinction is somewhat artificial and whether a treatment is used in either

way will tend to depend more on the balance between risks and benefits and whether the treatment is acceptable to the subject and cost-effective. Osteoporosis itself is asymptomatic and its clinical significance is that it is an important modifiable risk factor for low-trauma fracture. When selecting a therapy, it is more relevant that the treatment has antifracture efficacy and to determine whether this efficacy is for both vertebral and nonvertebral fractures, particularly including hip fracture because this is the most costly fracture to the subject and society. Figure 10.1 and Figure 10.2 illustrate the different effects of four different osteoporosis treatments on vertebral and nonvertebral fractures in a metaanalysis of published trials.

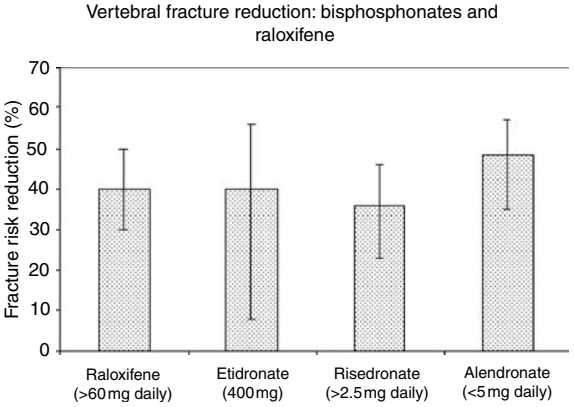


Figure 10.1. An illustration of the comparable effects of four osteoporosis treatments on vertebral fracture risk (95% confidence intervals marked). Adapted from Cranney, A. *et al.* (2002). *Endocr Rev* **23:**570–8.

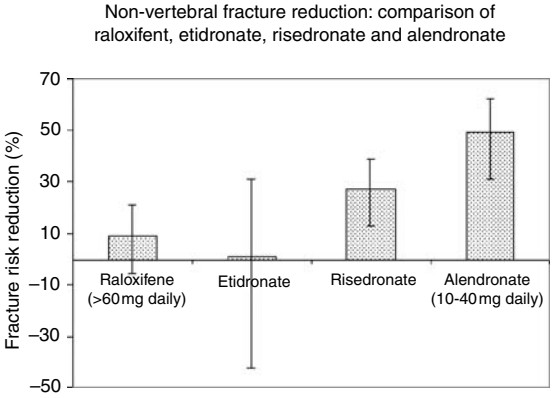


Figure 10.2. An illustration of the comparable effects of four osteoporosis treatments on nonvertebral fracture risk (95% confidence intervals marked). Adapted from Cranney, A. *et al.* (2002). *Endocr Rev* **23:**570–8.

Other desirable characteristics of a pharmacological intervention for osteoporosis are safety and tolerability. Ideally, a preparation should be easy to take because this will improve the chances of both compliance and persistence with treatment. Cost-effectiveness is increasingly determining which preparations healthcare organizations will permit clinicians to prescribe and in England; the National Institute for Health and Clinical Effectiveness (NICE) and the activities of prescribing advisers and formulary committees are very influential in this process.

It must be remembered that the majority of the RCTs that have been published are in Caucasian postmenopausal women. There are substantial variations in the prevalence of osteoporosis and osteoporotic fractures in different countries, even in this group of subjects. Although probably equally effective in men, the evidence base is limited. We have little knowledge of the efficacy of treatments in racial groups other than Caucasians, in which the absolute fracture risk could be much lower and, therefore, cost-effectiveness more difficult to demonstrate.

Bone is a dynamic tissue that is constantly remodelling through the activity of osteoclasts and osteoblasts, whose function is modulated by skeletal and extraskeletal signalling that is beyond the scope of this chapter. Treatments for osteoporosis aimed at reducing fractures can broadly be divided into three groups: those that reduce resorption by inhibiting osteoclastic activity, those that have anabolic functions that stimulate osteoblastic activity to lay down more bone, and one preparation that seems to have a dual action.

Finally, vitamin D and calcium are both integral to bone health and their role in the management of osteoporosis will also be discussed. Virtually all the RCTs described below that have fracture as an outcome attempted to ensure subjects were replete in both calcium and vitamin D.

10.3.1. HRT

HRT has historically been the mainstay of treatment for those with osteoporosis or who are at risk of osteoporosis. There was, for some time, good evidence for the prevention of postmenopausal bone loss and some limited evidence from observational studies that suggested HRT reduced fractures.[13] The major use of HRT was, however, in the perimenopausal and immediately postmenopausal woman and the criticism was that treating younger women in their late 50s and early 60s was not likely to have a great impact on the incidence of vertebral and hip fractures in their 70s and 80s because the benefits of HRT seemed to be rapidly lost on discontinuation.[14] Ironically, it was the same RCT, the Women's Health Initiative (WHI) study,[15] which finally produced convincing evidence for efficacy in reducing fractures while it demonstrated an unacceptable increased risk in thromboembolic side effects, stroke, coronary heart disease, and breast cancer. Although the study has had many critics and the absolute risk of adverse events was quite low, particularly in younger women, it has effectively signalled the end of HRT as a widely used preparation for osteoporosis. It still has a role in the treatment of women needing bone protection after suffering a menopause well before the modal age, but other

uses require careful risk evaluation in partnership with the subject, especially if a combined preparation is to be used if the incidence of adverse events seems higher than in subjects treated with unopposed oestrogen.[16,17]

10.3.2. Selective Oestrogen Receptor Modulators (SERMs)

Raloxifene (Evista® Eli Lilly and Company, Indianapolis, USA), at a dose of 60 mg/day, is the only current product of this class available for osteoporosis. The definitive study was the Multiple Outcomes of Raloxifene Evaluation (MORE) trial.[18] Raloxifene is licensed for the prophylaxis and treatment of osteoporosis in postmenopausal women. Drugs of this class have both agonist and antagonistic actions on oestrogen receptors. Raloxifene has positive oestrogenic effects on bone. It prevents bone loss but does not stimulate breast or uterine tissues.[19] It has beneficial effects on low-density lipoproteins, raising the possibility of cardiovascular benefits.[20]

The MORE study compared the effects of raloxifene, 60 mg/day and 120 mg/day, with placebo over a period of 4 years in more than 7000 postmenopausal women, with a mean age of 67 years (range, 31–80 years), who had either osteoporosis (a BMD T-score of < -2.5 at the hip or spine) or a morphometric vertebral fracture. Other bone remodelling agents were allowed in the fourth year of treatment. BMD was significantly increased and markers of bone turnover were appropriately suppressed in the actively treated arm of the study. More importantly, there was a reduction in vertebral fracture compared with placebo. At 3 years, 6.6% of subjects treated with the licensed dose of raloxifene (60 mg/day) had sustained at least one new vertebral fracture compared with 10.1% of women receiving placebo. The relative risk (RR) was 0.7; the 95% confidence interval (CI) was 0.5–0.8. The greatest absolute risk reduction was, of course, seen in those with the highest absolute risk, that is those with a prior vertebral fracture (Figure 10.3). No evidence was found that raloxifene reduced the risk of nonvertebral fracture.

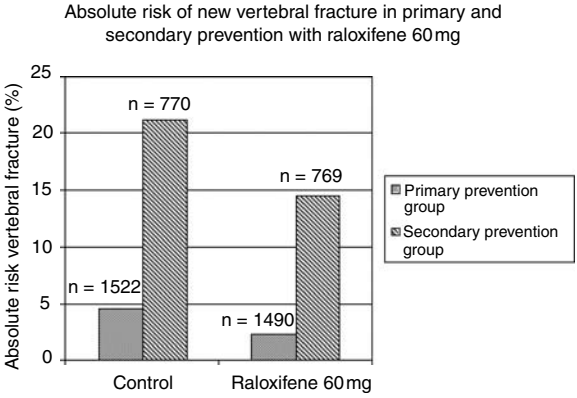


Figure 10.3. The differential effect of raloxifene in primary and secondary prevention of new vertebral fracture. With acknowledgement to Ettinger, B. *et al.* (1999). *JAMA* **282:**637–45.

Women receiving raloxifene had an increased risk of venous thromboembolus compared with those receiving placebo (RR, 3.1; 95% CI, 1.5–6.2). Raloxifene did not cause vaginal bleeding or breast pain but there was an increased risk of hot flushes, leg cramps, and peripheral oedema.

Breast cancer was significantly reduced in the MORE study (RR, 0.3; 95% CI, 0.2–0.6) and further analysis[19] has confirmed a 90% reduction in oestrogen receptor-positive breast cancer only. Data at 4 years found a 72% reduction in invasive breast cancer.[21] The relative efficacy of raloxifene compared with another SERM, tamoxifen, is currently being evaluated in the Study of Tamoxifen and Raloxifene (STAR) trial. Although there was no primary preventative effect in the MORE study at 4 years, there was evidence of a 40% reduction in cardiovascular events in a subset of just over 1000 women with known ischaemic heart disease.[22]

10.3.3. Bisphosphonates

Bisphosphonates are stable analogues of inorganic pyrophosphates, which have made a substantial contribution to the disease area and have come to dominate the market. They act as antiresorptives and all have evidence of effectiveness, that is they reduce markers of bone turnover, increase BMD, and reduce fractures, although not all bisphosphonates have evidence of efficacy both vertebral and nonvertebral fractures. One characteristic of these preparations as a consequence of their mode of action and the cycle of the bone remodelling unit, is that they do not need to be taken daily. Indeed, the once weekly versions are now the most commonly prescribed formulation, but a monthly oral preparation is also available and a quarterly parenteral bisphosphonate for the treatment of osteoporosis is shortly to be released at the time of writing. Parenteral preparations of bisphosphonates, with infrequent dosing schedules, have been available for some time for the management of oncology. Because of the tendency of the oral preparations to cause upper gastrooesophageal symptoms and the almost complete failure of absorption if taken with food, rather complex administration instructions must be complied with. Medication, with the exception of etidronate, which can be swallowed in the middle of a 4-hour fast, should be taken first thing in the morning on an empty stomach with a full glass of tap water and in an upright position that is maintained for at least 30 minutes before food is consumed. Bisphosphonates are contraindicated in subjects with severe renal impairment whose glomerular filtration rate is <35 ml/min.

10.3.3.1. Etidronate

Disodium etidronate (Didronel PMO®, Procter & Gample Pharmaceuticals, Cincinnati, OH, USA) is licensed in the UK for prevention and treatment of osteoporosis in both men and postmenopausal women and treatment of glucocorticoid-induced osteoporosis (GCIOP). The drug is taken as a 14-day pulse of oral etidronate (400 mg), followed by 76 days of calcium supplementation. The original studies were rather limited and could be criticized for a number of reasons.[33,34]

Etidronate never gained a license in the USA as a result. There has never been any convincing evidence from RCTs for its effectiveness in hip fractures, which was only reported in an observational study,[23] and vertebral fracture efficacy was only seen in those with the most severe baseline disease.[35] The drug is now used increasingly rarely and in 2005 only 4.2 % of osteoporosis-related prescriptions were for etidronate preparations.[24]

10.3.3.2. Alendronate

Alendronate (Fosamax®, Merck & Co. Inc., Whitehouse Station, NJ, USA) is licensed for the treatment of osteoporosis in postmenopausal women and men and GCIOP. The drug is available at dosages of 5 mg/day, 10 mg/day, and 70 mg/week. The 10 mg/day and 70 mg/week doses are those for which trial evidence supports the best efficacy. The drug has recently become available generically, but at the time of writing, there is no generic daily dose available. Alendronate is the most commonly prescribed bisphosphonate (Figure 10.3) and has recently become available in combination with 400 IU of colecalciferol (vitamin D3) as Fosavance®, (Merck & Co. Inc., Whitehouse Station, NJ, USA).

Alendronate is the most extensively studied bisphosphonate and has been shown to increase BMD and reduce the risk of vertebral, hip, and other nonvertebral fractures in subjects with prior vertebral fracture and those with low BMD. One arm of the Fracture Intervention Trial (FIT)[25] studied over 2000 women between 55 and 81 years old who had low hip BMD (equivalent to a T-score of -2.1 or more) and a prevalent vertebral fracture over 3 years. The study showed a significantly reduced RR of new morphometric vertebral fracture (0.53, with 95% CI of 0.41–0.68). The hip fracture rate was also reduced in the treatment arm of the study, with a RR of 0.49 (95% CI, 0.23–0.99).

The effect of alendronate in women with low BMD but no prior vertebral fracture was assessed in another arm of the FIT, which was a further 4-year study[26] involving just over 4400 subjects. Although BMD increased, as in the previous trial, a significant fracture risk reduction of 36% (RR, 0.64; 95% CI, 0.5–0.82) was only seen in subjects with a BMD in the osteoporotic range, that is a T-score <-2.5. In the same group, a post-hoc analysis found a 56% reduction in hip fracture (RH, 0.44; 95% CI, 0.18–0.97). The risk reduction of 50% for morphometric vertebral fracture was, again, highest in subjects with osteoporosis (RH, 0.5; 95% CI, 0.31–0.82).

The effect of alendronate in men was studied in a small 2-year RCT[27] of 241 men (mean age, 63 years) who had moderately lowered BMD (T-scores at the hip and lumber spine of ≤ -2.0 and ≤ -1.0, respectively) and at least one prevalent fragility fracture. There were comparable increases in BMD and suppression of bone turnover markers, as seen in earlier studies. Although there was very little difference in overall fracture rates in the treatment and placebo arms of the study, the authors reported a significant reduction in vertebral fractures in the active arm of the study (0.8% in the alendronate group versus 7.1% in the placebor group; $p \leq 0.02$).

Glucocorticoids lead to a reduction in bone quality and BMD and increased fracture risk. Although RCTs are limited in the use of Alendronate in subjects with GCIOP a pooled study of just over 500 men and women on high-dose prednisolone (≥7.5 mg/day for at least 1 year) has been reported.[28] There were significant increases in BMD and suppression of bone turnover markers and a nonsignificant trend towards reduced vertebral fracture (2.3% for subjects receiving Alendronate compared with 3.7% in the placebo group).

There have been many other studies involving post-hoc analysis of the original FIT data. The evidence base supports the use of the daily preparation and the license for the weekly preparation is based on noninferiority trials in which BMD and bone turnover markers were outcome measures.[29] Overall, the drug seems well tolerated, with no differences in adverse reactions when compared with placebo, including upper gastrointestinal symptoms, if the medication is taken correctly. The studies are quite large, with a small drop-out rate, and seem to report a consistent fracture reduction rate that approaches 50%. This impression is reinforced by studies that are pooled in metaanalysis,[30] in which confidence intervals tighten around a RR of 0.52 for vertebral fracture and 0.63 for hip fracture at an alendronate dosage of at least 10 mg/day. The analysis period has been extended to 10 years[31] and this has given reassurance of long-term safety, in addition to an extended offset of effect—the period of time during which the effectiveness of a drug wanes—which could be as much as 5 years. Although the surviving subject numbers are small, there seems also to be a continuing benefit on BMD after 10 years of continuous therapy.

10.3.3.3. Risedronate

Another bisphosphonate, risedronate (Actonel®, Procter & Gamble Pharmaceuticals, Cincinnati, OH, USA) is available in 5 mg/day and 35 mg/week dosages. It works in a very similar way to alendronate and is licensed for the prevention and treatment of osteoporosis in postmenopausal women, to reduce vertebral and hip fractures. It has a license for GCIOP in its daily formulation and, similar to alendronate, is often used "off license" for this condition in its once-weekly form. The drug does not have a license for male osteoporosis, but it seems unlikely that it would be ineffective in this group of subjects. The same precautions for administration of risedronate should be adopted as for alendronate, to ensure minimal gastrointestinal side effects and maximum bioavailability.

The major placebo-controlled trials of effectiveness for fracture reduction are the Vertebral Efficacy with Risedronate Therapy[32] in North Ameria (VERT-NA) and the rest of the world[33] (VERT-MN). The former trial ran for 3 years and included nearly 2500 postmenopausal women who were younger than 85 years and had at least one vertebral fracture. Subjects were randomised to receive oral risedronate (2.5 or 5 mg/day) for 3 years or placebo. The 2.5 mg/day risedronate arm was discontinued after 1 year. The mean T-score at baseline at the hip was −2.6 or −2.7 in the groups assigned to the 2.5 mg/day and 5 mg/day dosages, respectively. The mean T-score at the lumbar spine at baseline was −2.4 in all

subjects regardless of risedronate dosage. At 3 years, the authors reported a 41% reduction in vertebral fracture (RR, 0.59; 95% CI, 0.43–0.82) and 39% reduction in nonvertebral fracture (RR, 0.6; 95% CI, 0.39–0.94). The latter study was similar; it included just over 1200 women and reported a 49% reduction in vertebral fracture (RR, 0.51; 95% CI, 0.36–0.73) and 33% reduction in nonvertebral fracture (RR, 0.67; 95% CI, 0.44–1.04).

A study in which hip fracture was a primary endpoint involved nearly 5500 women who were aged 70–79 years and had osteoporosis and a further 3800 women who were over 80 years old and had clinical risk factors for hip fracture. The entry criteria were a T-score of at least −4.0 or −3.0 and additional nonskeletal risk factors. After 3 years, the RR of hip fracture among all the women treated with risedronate was 0.7 (95% CI, 0.6–0.9). In the 70–79-year-old women with osteoporosis the RR was 0.6 (95% CI, 0.4–0.9). In women selected primarily on the basis of nonskeletal risk factors and who were over 80 years of age there was only a nonsignificant trend towards reduced hip fracture.

Extension studies, of up to 5 years[34,35] and 7 years[36] of treatment, have confirmed the continued benefits on BMD and sustained vertebral fracture risk reduction. A metaanalysis pooled the results of several studies[37] and, although high (as much as 35% in the largest trial), found that the drop-out rate was unlikely to affect the magnitude of the treatment effect. The pooled RR for vertebral fractures was reported as 0.64 (95% CI, 0.54–0.77) for dosages of risedronate in excess of 2.5 mg/day. Similarly, for nonvertebral fracture, the pooled RR in subjects receiving at least 2.5 mg/day of risedronate was 0.73 (95% CI, 0.61–0.87).

The antifracture efficacy and safety in older subjects is an important clinical consideration because age is the greatest predictor of fracture risk and is also associated with a higher risk of gastrooesophageal disease and symptomatology. Subset analysis of pooled data from the three major trials has shown a substantial (81%) reduction in vertebral fracture risk (HR, 50.19; 95% CI, 0.09–0.40) in subjects over 80 years old.[38] Adverse events in subjects receiving risedronate did not differ from the placebo group.

Risedronate has been shown to prevent bone loss in GCIOP[39] and a 70% reduction in vertebral fracture has been observed.[40,41] Similar to alendronate, the evidence base for use of the once-weekly preparation depends on a bridging study.[42]

It is tempting to compare the effects of two similar drugs, such as risedronate and alendronate, but there has not been a direct "head-to-head" study with fracture as an outcome. The Fosamax® Actonel® Comparison Trials (FACTS 1) study[43] demonstrated that, in terms of the improvement in BMD and reduction in bone turnover markers, alendronate seemed superior. However, the relationship between these surrogate measures and fracture outcome is complex and greater clinical efficacy cannot be inferred. Similarly, the results of metaanalysis might imply alendronate has a greater risk reduction, but the populations in the studies are not the same and, in any event, the CIs of the pooled results overlap. There is some evidence that a reduction in vertebral fracture risk is seen as early as 6 months after starting treatment with risedronate although equivalent data are unavailable for alendronate at the same time point. Subjects receiving risedronate had similar adverse reactions to

those receiving placebo, and similar to alendronate, the often expressed opinion that bisphosphonates cause major problems with upper gastrointestinal adverse events is not supported by clinical studies.[44]

10.3.3.4. Ibandronate

The most recently developed bisphosphonate is ibandronate (Bonviva®, Glaxo Smith Kline, Uxbridge, Middlesex, UK). The drug is licensed at an oral dose of 150 mg/month for the treatment of osteoporosis in postmenopausal women, to reduce the risk of vertebral fracture. Ibandronate is a third generation bisphosphonate, which inhibits bone resorption in the same way as the therapies outlined above. Administration advice is similar to that for risedronate and alendronate, but there is a requirement for the subject to spend a period of 1 hour upright before eating or drinking following administration. It is probable that this enhances bioavailability.

The definitive evidence for ibandronate comes from the oral iBandronate Osteoporosis vertebral fracture trial in North America and Europe (BONE) study,[45] which employed a dosage of 2.5 mg/day compared with an equivalent intermittent regimen of 20 mg on alternate days for 12 doses every 3 months. This 3-year study involved nearly 3000 postmenopausal women who were aged 55–80 years and had a BMD T-score of ≤ -2.0 at any one lumbar vertebra and at least one prior vertebral fracture. Over 3 years there was a significant (62%; 95% CI, 40.9–75.1) reduction in morphometric vertebral fracture in the daily administered group, with a 50% (95% CI, 26–60) reduction in the intermittently administered group. Clinical vertebral fracture was also significantly reduced, by 49% and 48% in the daily and intermittent groups, respectively. There was no statistically significant difference in the incidence of nonvertebral fracture.

A simple, once-monthly regimen of ibandronate is currently being investigated in a 2-year, randomized, double-blind, multicentre study (the Monthly Oral iBandronate in LadiEs or MOBILE study) of approximately 1600 postmenopausal women with osteoporosis. After 1-year, results from MOBILE showed that once-monthly regimens of oral ibandronate were as effective as a 2.5 mg once daily oral regimen of ibandronate, with similar tolerability in the treatment of postmenopausal osteoporosis.

European approval has now been given for a 3 mg dose that is administered as a 15 to 30 second intravenous injection every 3 months. Approval was based on a review of data from another noninferiority trial, the Dosing IntraVenous Administration (DIVA) study. Only the year 1 data have been published so far.[46] One problem associated with quarterly injections is an influenza-like response known as an "acute-phase reaction", which lasts 1–2 days in a proportion of subjects, usually after the first dose only.

The significance of the evidence for the effectiveness of once-monthly ibandronate must be interpreted in light of the fact that the recruited subjects were not a high-risk group for hip fracture and only osteopenic at the total-hip region of interest (ROI), with a mean BMD of $T = -1.7$. However, the licensed use of

ibandronate is for the treatment of subjects with osteoporosis who are at risk of vertebral fracture and this is where the evidence lies. The infrequent oral dosage regimen could be a particular advantage to some subjects if concordance is an issue, whether because of the complex dosing regimen or gastrooesophageal side effects.[47]

10.3.4. Parathyroid Hormone (PTH; Teriparatide)

A biologically active 34-amino acid synthetic peptide fragment of PTH (teriparatide; Forsteo®, Eli Lilly Nederland B. V., Houten, Nederland) is licensed for the treatment of postmenopausal osteoporosis, although it also has approval for use in men in the USA. It is administered by daily subcutaneous injection and is supplied in a prefilled 3 ml syringe. The syringe volume contains 750 μg, which is enough for 28 days of the licensed 20 μg/day dosage regimen. The maximum length of treatment is 18 months (see below).

The mode of action differs fundamentally from the previously described preparations because it stimulates osteoblastic activity and cell survival, and, therefore, promotes an anabolic or bone-forming effect, most significantly in skeletal sites with a high proportion of trabecular bone. This has obvious possibilities in regaining lost bone mass and normalizing the microarchitecture.

The definitive study was a randomized, placebo-controlled study (the Fracture Prevention Trial[48]), involving 1637 postmenopausal women with existing vertebral fractures who were treated with 20 μg/day or 40 μg/day doses of teriparatide. The study was designed to last 3 years but was terminated at 18 months because of the occurrence of sarcoma in rats, even though these findings are not thought to be relevant to humans. Teriparatide significantly increased BMD in the lumbar spine and reduced the risk of new vertebral and nonvertebral fractures. The authors reported RRs of vertebral fracture in the 20 μg/day and 40 μg/day ibandronate-treated groups as 0.35 and 0.31, respectively (95% CI, 0.22–0.55 and 0.19–0.50, respectively), compared with placebo. The RR of nonvertebral fragility fracture was 0.47 and 0.46, respectively (95% CI, 0.25–0.88 and 0.25–0.86, respectively). The incidence of new moderate and severe vertebral fractures (defined as a loss of height >26%) was even more effectively reduced, with a RR of 0.1 and 0.22, respectively (95% CI, 0.04–0.27 and 0.11–0.45, respectively). The ibandronate, 40 μg/day dose increased BMD to a greater extent than the ibandronate, 20 μg/day dose, although this was not reflected in the fracture outcomes and the higher dose was more likely to have side effects, which were similar to placebo at the ibandronate, 20 μg/day dose, consisting principally of nausea, dizziness, leg cramps, and headache. In summary, teriparatide is very effective at reducing the risk of further vertebral fracture and nonvertebral fracture at a dose of 20 μg/day. There are no data for its effectiveness in hip-fracture prevention and, indeed, there has been speculation that gains in the vertebrae might be at the expense of cortical sites, such as the hip. Clinical findings with teriparatide treatment have included a decrease in cortical BMD or little change in BMD of the femoral neck or whole body.[49] Biopsy evidence, however, tended to suggest that

changes in cortical bone morphology and geometry occurred that should improve biomechanical competence and resistance to fracture at sites such as the hip, despite what was happening to the BMD.[50]

It might be expected that individual anabolic and antiresorptive agents might be more efficacious in combination. In fact, they are not synergistic in action and concomitant prescription is not recommended. Prior exposure to antiresorptive therapy seems to delay and reduce the anabolic activity of teriparatide, compared with prior treatment with raloxifene, a weaker antiresorptive agent.[51] Although teriparatide does not have a license for GCIOP, there is some evidence that it might be effective in reducing bone loss in these circumstances.[52] In addition, there is evidence from a RCT of >400 men of increases in BMD equivalent to that seen in postmenopausal women.[53]

One of the problems associated with teriparatide is its relatively high cost, at just over GB£270 (£390; US$480) per month or just under GB£5,000 (£7,250; US$9,000) for a full course of treatment. This has led to a restriction on its use within the UK for health-economic reasons.

10.3.5. Strontium Ranelate

Strontium ranelate (Protelos®, Servier, Neuilly-sur-Seine, France) consists of two atoms of strontium and a molecule of ranelic acid, to ensure absorption. Strontium is in the same atomic group as calcium, which it replaces within bone. Protelos® is a daily preparation of granules that are taken as a suspension (2 g in a glass of water) at least 2 hours after food, preferably at bed time. Strontium ranelate is licensed for the treatment of postmenopausal osteoporosis, to reduce the risk of both vertebral and hip fractures. However, the drug is not licensed for prevention of osteoporosis, use in men, or GCIOP.

The drug seems to have a unique mode of action that is not fully understood but seems to uncouple bone turnover and involve both suppression of resorption, by inhibiting the differentiation of preosteoclasts into multicellular osteoclasts, and maintenance of bone formation, through enhanced collagen synthesis and osteoblast replication.[54,55] Although it is always important to differentiate between benefits on BMD and antifracture efficacy of any treatment for osteoporosis, this is even truer for strontium ranelate. Because the atomic weight of strontium (38) is higher than that of calcium (22) and, therefore, attenuates X-ray transmission to a greater extent, the apparent increase in BMD will be partly explained by the percentage of calcium atoms replaced by strontium atoms. Because this is likely to depend on the duration of and compliance with therapy, in addition to the ROI being scanned, it is difficult to define an adjustment factor without a measure of the bone's strontium content.[56]

The evidence base for clinical efficacy depends on two RCTs. The Spinal Osteoporosis Therapeutic Intervention (SOTI) study[57] was a 3-year placebo-controlled RCT involving nearly 1700 postmenopausal women over the age of 50 years who had osteoporosis and at least one vertebral fracture. The drop-out rate was low and the RR for new vertebral fracture was 0.51 and 0.59 at 1 year and

3 years, respectively (95% CI, 0.36–0.75 and 0.48–0.73, respectively). The Treatment of Peripheral Osteoporosis (TROPOS) study[58] was designed to evaluate nonvertebral fracture outcomes also over 3 years. The study involved just over 5000 women with osteoporosis and who were >74 years old or aged between 70 and 74 years but with one additional fracture risk factor (e.g. history of osteoporotic fracture after the menopause, residence in a retirement home, frequent falls, or a maternal history of osteoporotic fracture of the hip, spine, or wrist). The RRs for all vertebral and major nonvertebral (hip, wrist, pelvis and sacrum, ribs and sternum, clavicle, and humerus) fractures were reduced by 16% and 19%, respectively, after 3 years in strontium ranelate-treated subjects [RR, 0.84 (95% CI, 0.702–0.995) 0.81 (95% CI, 0.66–0.98), respectively]. Vertebral fracture risk reduction was similar to that reported in the SOTI study. In the entire study population, there was a 15% reduction in hip fracture, but this was not significant because the study was not powered to detect it. In a subset analysis required by the regulatory authority, which included just under 2000 women over 74 years of age with a T-score of −3.0 on the Hologic European database (which is equivalent to ≤−2.4 on the National Health and Nutrition Examination Survey (NHANES) database used throughout the rest of this chapter), there was a 36% reduction in hip fracture in the treatment arm (RR, 0.64; 95% CI, 0.412–0.997; $p = 0.046$). A recent study supported the effectiveness of strontium ranelate in vertebral fracture reduction independently of existing risk factors, such as baseline BMD, number of vertebral fractures, family history, or smoking status. Efficacy was preserved in older populations of women who were 80 years of age or more.[59]

Strontium ranelate was generally well tolerated, with a slight excess of nausea, diarrhoea, and eczema in the first 3 months. A small, but statistically significant, increase in venous thromboembolism (VTE) subjects on strontium ranelate [odds ratio (OR) = 1.5 at 3 years; 95% CI, 1.1–2.1]. The OR was similar for pulmonary embolus (PE), at 1.7 (95% CI, 1.0–3.1). In absolute terms, this led to six deaths in the treatment arm compared with three in the control group (out of 25 PEs in the strontium ranelate-treated group and 14 PEs in the control group).[60] There is no known plausible biological explanation for this effect but caution should be exercised with its use in subjects at risk of VTE. The effect is much less than that seen with raloxifene, for which the RR is 3.1 (95% CI, 1.5–6.2) for VTE and 4.8 for PE.[18]

10.3.5. Calcium and Vitamin D

There is a positive association between calcium intake and bone mass. Healthy bones need a balanced, calcium-rich diet throughout life. Lifelong inadequate dietary intake is associated with failure to achieve peak bone mass (PBM).[61] Calcium and vitamin D supplementation has been shown to reduce the rate of bone loss in postmenopausal women and in those >65 years of age.[62]

In older women, both adequate levels of dietary calcium and calcium supplements had been thought to reduce fracture risk, with a dose-dependent relationship,[63] and these sort of findings support the recommendations in some

guidelines for higher levels of calcium supplementation in women with osteoporosis than the usual recommended nutrient intake (RNI) of 700 mg/day for individuals >65 years.[64] Three more recent RCTs have failed to demonstrate, on an "intention-to-treat" basis, significant fracture reduction in community-living older people with calcium supplementation alone or in combination with vitamin D.[65–67] Other large-scale observational studies in older women have also supported this view.[68] A metaanalysis from the Cochrane Group recently reported no benefit on fracture outcomes in community-dwelling older people or in those treated with vitamin D alone.[69] Compliance in the RCTs has been noted as a possible explanation for an apparent lack of effect, and an even more recent paper seemed to show a 34% reduction of fracture in approximately 700 compliant women out of just over 1400 subjects, who were >70 years, after 5 years of treatment.[70]

Whatever the controversies regarding the role of calcium, with or without vitamin D supplementation, in community-living older people, calcium (1200 mg/day) and vitamin D (800 IU/day) supplementation can be particularly important in the elderly in the residential and nursing home environment. This group have a much higher risk of hip fracture than community-living older people.[71,72] Research carried out in women living in institutions in France has shown effectiveness[73,74] and cost-effectiveness[75] in preventing hip fracture, with a risk reduction of approximately 30%. It is not clear from these studies whether calcium or vitamin D, or a combination of the two, is the effective agent. It is possible that a primary benefit comes from the correction of vitamin D insufficiency, which is beneficial to falls risk through optimization of neuromuscular strength and coordination rather than through benefits to bone health.

Vitamin D consists of two similar molecules, vitamin D2 (ergocalciferol) and vitamin D3 (colecalciferol). It is absorbed in the gut and synthesised in the skin from a provitamin (7-dehydrocholesterol) under the influence of sunlight; for this reason, it is not, in fact, a vitamin, but is better described as "a steroid hormone". The confusion probably arose from the dramatic effect of cod liver oil supplementation in childhood rickets. Vitamin D is hydroxylated in the liver to 25-hydroxyvitamin D [25(OH)D] and converted into its active form, 1,25-dihydroxyvitamin D [$1,25(OH)_2D$], in the kidney. Vitamin D has not often been studied as an isolated pharmacological intervention for fracture. In a study based in Holland and involving community-living individuals, no effect on fracture was found at a dose of vitamin D, 400 IU/day.[76] Another study found that, even in the residential care environment, there was lack of significant benefit on fracture, although there was a significant reduction in falls if subjects were at least 50% compliant,[77] even if they were not vitamin D-deficient at baseline. However, a metaanalysis of RCTs of vitamin D3 with or without calcium that had fracture as an outcome reported a RR reduction in ambulatory and institutionalized elderly persons of 26% (RR, 0.74; 95% CI, 0.61–0.68) at a dose of vitamin D3 of 700–800 IU/day. This effect disappeared at a dose of vitamin D3 of 400 IU/day.[78]

One problem is that there is controversy regarding the serum levels of 25(OH)D that define insufficiency and deficiency[79] and the daily intake of

vitamin D necessary to maintain these levels. This is complicated because the majority of vitamin D is synthesised in the skin and this process becomes roughly 50% less efficient with ageing.[80] Because of the difficulty of formulating a reasonably acceptable diet that will compensate, replacement therapy often becomes necessary. The RNI for vitamin D is 10 μg/day (400 IU/day). There is some evidence that 800 IU/day is more appropriate for maximum benefit in fall reduction, as described above, and possibly for bone health also.

Despite the role of ultraviolet B (UVB) from the sun's rays, vitamin D insufficiency is common in older people at all latitudes and particularly among women with osteoporosis.[81] Deficiency and insufficiency are more common in those with peripheral fracture,[82,83] especially hip fracture in the UK.[84] Although there is conflicting evidence for benefit in fracture reduction, there is continued evidence for benefit in reducing falls in both individual studies[85] and metaanalyses.[78] Because most peripheral fractures follow a fall,[86] it remains important to optimize vitamin D levels to reduce this risk.

While discussing the role of calcium and vitamin D3 supplementation as an independent intervention for osteoporosis and fracture reduction, it remains important to consider its role alongside specific bone remodelling agents. Some guidelines state that adequate levels of calcium and vitamin D are needed to ensure optimum effects of the treatments for osteoporosis.[91] Accordingly, they recommend that calcium and/or vitamin D supplementation should be provided unless clinicians are confident that subjects receiving osteoporosis treatment have an adequate calcium intake and are vitamin D replete. Other earlier guidance[87–89] recommends that older subjects with osteoporosis receive an adequate intake of calcium and vitamin D. The guidelines all base their recommendations on the evidence base for osteoporosis treatments, primarily bisphosphonates. All RCTs in osteoporosis with fracture endpoints have compared treatment (bisphosphonates, teriparatide, raloxifene, and strontium ranelate) with placebo in subjects who were calcium and vitamin D replete. There is actually no evidence that these therapies are effective on their own because the studies have not been performed with peripheral fracture as an outcome.

10.3.6. Calcitriol

Calcitriol or 1,25$(OH)_2$D (Rocaltrol®, Hoffman-La Roche, Nutley, NJ, USA) is licensed for the treatment of established postmenopausal osteoporosis. It is the active form of vitamin D, which is produced by renal hydroxylation of colecalciferol [25(OH)D]. Because of renal impairment in older people, there is a risk of inadequate levels of the active metabolite. Because vitamin D is essential to the maintenance of adequate bone health and deficiency is associated with hyperparathyroidism and low levels of BMD, calcitriol has been promoted as a treatment for established postmenopausal osteoporosis. Calcitriol was compared with calcium in a single-blind study of just over 600 postmenopausal women with at least one prior vertebral fracture; there was a reduction in the rate of vertebral fracture over 3 years in the active treatment (Calcitriol) arm of the study.[90] During the second

year of treatment, there were 9.3 fractures per 100 subject years in the Calcitriol group compared with 25.0 fractures per 100 subject years in the control group. In the third year, there were 9.9 per 100 subject years in the Calcitriol group compared with 31.5 fractures per 100 subject years in the control group ($p \leqslant 0.001$). On this basis, the medication has a license and is used in some countries but has little of the market in the UK, perhaps because its use requires regular monitoring of serum calcium owing to of the rare occurrence of hypercalcaemia.

10.3.7. Calcitonin

Calcitonin (Miacalcic®, Sandoz International GmbH, Holzkirken, Germany) is available as an intramuscular or subcutaneous injection of 100 IU/day in one or two divided doses. The drug is licensed for acute bone loss associated with immobility for 2 to 4 weeks. Calcitonin seems to have some benefit in reducing pain from acute vertebral fracture, although this use is not supported by robust evidence or its license. The drug has a number of side effects, including flushing, nausea, and diarrhoea and should be used with caution in people with allergies. Hypocalcaemia can be a problem and this must be monitored. A nasal form of calcitonin is available, at a dose of 200 IU/day, to reduce the risk of vertebral fracture in postmenopausal osteoporosis.

The Prevent Recurrence of Osteoporotic Fractures (PROOF) study examined the effect of 100 IU/day, 200 IU/day, and 400 IU/day of calcitonin in just over 1200 postmenopausal women with prior vertebral fracture and a T-score of ≤ -2.0 over 5 years. Only 132 of the 316 subjects assigned to the licensed dose of 200 IU/day completed the study. Of the 287 subjects that completed the 3-year study, there was a 35% reduction of new vertebral fracture (RR, 0.65; 95% CI, 0.47–0.97). Significant fracture reduction was not seen with any other dosage of calcitonin, including the highest dose. Apart from rhinitis, the drug was quite well tolerated. The high drop-out rate and the fact that there no dose-dependent effects were demonstrated means the study can be criticized. Calcitonin is now rarely used in clinical practice.

10.4. Nondrug Treatment Options

This chapter concentrates on the pharmacotherapeutic agents that are promoted to reduce fracture. It must be acknowledged that there is a close relationship between falls and fractures and the absolute risk of fracture following a fall could be between 3% and 5%; with between 20% and 25% of those falls resulting in hip fracture, these studies might underreport the true fall rate.[92,93] The combination of osteoporosis and a recent fall might amplify the fracture risk by a factor as high as 24.8.[94] Although there is now robust evidence that certain interventions can reduce the rate of falling,[95,96] it is very difficult to demonstrate that these interventions can reduce fracture, hospital admission, or nursing home admission because assessment and intervention are far more complex than those for

osteoporosis and the subject numbers needed for a trial powerful enough to show benefit would be unfeasibly large. Nevertheless, on a pragmatic basis, interventions that reduce the rate of falling and the number of individuals who fall might integrate well with strategies designed to improve bone health.

Hip protectors were once considered an effective strategy for fracture reduction, and, indeed, they can be, if worn at the time of the fall. Compliance, however, is a major problem, and on an "intention-to-treat" basis, recent trials involving individual subject randomization have failed to demonstrate a benefit, except possibly within the care home environment.[96]

10.5. Summary

There is now clear evidence from large, well-designed RCTs for effective and worthwhile interventions to reduce the risk of further fractures in subjects who are replete in calcium and vitamin D. In the context of financial constraints in healthcare economics, it is essential to ensure cost-effective prescribing so that only subjects at high risk receive appropriate therapies. These are often the older and less articulate subjects, and opportunistic case finding, as opposed to a systematic approach to care, runs the risk of inequality of access. There is consistent and repeatable evidence that our present approach is failing to identify or appropriately manage the overwhelming majority of even the highest risk subjects, such as those with prior fragility fracture, as described above, or those who are receiving glucocorticoids[97] or who are resident in the extended care setting.[98] Not only do these issues need urgent attention, but also systems are needed to improve the poor concordance and persistence with treatments for osteoporosis.

References

1. Burge, R.T., Worley, D., Johansen, A. *et al*. (2001). The cost of osteoporotic fractures in the UK: projections for 2000–2020. *J Drug Assessment* **4:**71–160.
2. Kanis, J., Brazier, J., Stevenson, M. *et al*. (2002). Treatment of established osteoporosis: a systematic review and cost-utility analysis. http://www.ncchta.org/execsumm/summ629.htm (accessed 4 May 2006). *Health Tech Assessment* **6(29)**.
3. Lawrence, T.M., White, C.T., Wenn, R. *et al*. (2005). The current hospital costs of treating hip fractures. *Injury* **36:**88–91.
4. Department of Health (2005). Hospital Episode Statistics (England). http://www.hesonline.nhs.uk/Ease/servlet/DynamicPageBuild?siteID=1802&categoryID=192&catName=Free%20data (accessed 27 March 2006).
5. Elliot-Gibson, V., Bogoch, E.R., Jamal, S.A. *et al*. (2004). Practice patterns in the diagnosis and treatment of osteoporosis after a fragility fracture: a systematic review. *Osteoporos Int* **15:**767–78.
6. Brankin, E., Caroline, M. and Munro R (2005). Closing the osteoporosis management gap in primary care: a secondary prevention of fracture programme *Curr Med Res Opin* **21:**425–82.

7. No authors listed (1993). Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med* **94:**646–50.
8. Faulkner, K.G. (2005). The tale of the T-score: review and perspective. *Osteoporos Int* **16:**347–52.
9. Masud, T. and Francis, R.M. (2000). The increasing use of peripheral bone densitometry. *BMJ* **321:**396–8.
10. Stone, K., Seeley, D., Lui, L. *et al*. (2003). BMD at multiple sites and risk of fracture of multiple types: long-term results from the study of osteoporotic fractures. *J Bone Miner Res* **18:**1947–54.
11. Schuit, S.C.E., van der Klift, M., Weel, A.E.A.M. *et al*. (2004). Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone* **34:**195–202.
12. Heaney, R.P. (2003). Is the paradigm shifting? *Bone* **33:**457–65.
13. Cauley, J.A., Seeley, D.G., Ensrud, K. *et al*. (1995). Estrogen replacement therapy and fractures in older women. *Ann Intern Med* **122:**9–16.
14. Gallagher, J.C., Rapuri, P.B., Haynatzki, G. *et al*. (2002). Effect of discontinuation of estrogen, calcitriol, and the combination of both on bone density and bone markers. *J Clin Endocrinol Metab* **87:**4914–23.
15. Writing Group for the Women's Health Initiative (2002). I. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* **288:**321–33.
16. Anderson, G.L., Limacher, M., Assaf, A.R. *et al*. (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA* **291:**1701–12.
17. Stefanick, M.L., Anderson, G.L., Margolis, K.L. *et al*. (2006). Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA* **295:**1647–57.
18. Ettinger, B., Black, D.M., Mitlak, B.H. *et al*. (1999). Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. *JAMA* **282:**637–45.
19. Cummings, S.R., Eckert, S., Krueger, K.A. *et al*. (1999). The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the MORE randomized trial. *JAMA* **281:**2189–97.
20. Riggs, B.L. and Hartmann, L.C. (2003). Selective estrogen-receptor modulators—mechanisms of action and application to clinical practice. *N Engl J Med* **348:**618–29.
21. Cauley, J.A., Norton, L., Lippman, M.E. *et al*. (2001). Continued breast cancer risk reduction in postmenopausal women treated with raloxifene: 4-year results from the MORE trial. *Breast Cancer Res Treat* **65:**125–34.
22. Barrett-Connor, E., Grady, D., Sashegyi, A. *et al*. (2002). Raloxifene and cardiovascular events in osteoporotic postmenopausal women: four-year results from the MORE (Multiple Outcomes of Raloxifene Evaluation) randomized trial. *JAMA* **287:**847–57.
23. van Staa, T.P., Abenhaim, L. and Cooper, C. (1998). Use of cyclical etidronate and prevention of non-vertebral fractures. *Rheumatology* **37:**87–94.
24. Department of Health (2005). Prescription Cost Analysis. http://www.ic.nhs.uk/pubs/prescostanalysis2005 [accessed 20 April 2006].
25. Black, D.M., Cummings, S.R., Karpf, D.B. *et al*. (1996). Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet* **348:**1535–41.

26. Cummings, S.R., Black, D.M., Thompson, D.E. *et al*. (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* **280:**2077–82.
27. Orwoll, E., Ettinger, M., Weiss, S. *et al*. (2000). Alendronate for the treatment of osteoporosis in men. *N Engl J Med* **343:**604–10.
28. Saag, K.G., Emkey, R., Schnitzer, T.J. *et al*. (1998). Alendronate for the prevention and treatment of glucocorticoid-induced osteoporosis. *N Engl J Med* **339:**292–9.
29. Rizzoli, R., Greenspan, S.L., Bone, G., 3rd *et al*.; The Alendronate Once Weekly Study Group (2002). Two-year results of once-weekly administration of alendronate 70 mg for the treatment of postmenopausal osteoporosis. *J Bone Miner Res* **17:**1988–96.
30. Cranney, A., Wells, G., Willan, A. *et al*. (2002). II. Meta-analysis of alendronate for the treatment of postmenopausal women. *Endocr Rev* **23:**508–16.
31. Bone, H.G., Hosking, D., Devogelaer, J.-P. *et al*. (2004). Ten years' experience with alendronate for osteoporosis in postmenopausal women. *N Engl J Med* **350:**1189–99.
32. Harris, S.T., Watts, N.B., Genant, H.K. *et al*. (1999). Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. *JAMA* **282:**1344–52.
33. Reginster, J.Y., Minne, H.W., Sorensen, O.H. *et al*. (2000). Randomized trial of the effects of risedronate on vertebral fractures in women with established postmenopausal osteoporosis. *Osteoporos Int* **11:**83–91.
34. Sorensen, O.H., Crawford, G.M., Mulder, H. *et al*. (2003). Long-term efficacy of risedronate: a 5-year placebo-controlled clinical experience. *Bone* **32:**120–6.
35. Ste-Marie, L.G., Sod, E., Johnson, T. *et al*. (2004). Five years of treatment with risedronate and its effects on bone safety in women with postmenopausal osteoporosis. *Calcif Tiss Int* **75(6):**469–76.
36. Mellstrom, D.D., Sorensen, O.H., Goemaere, S. *et al*. (2004). Seven years of treatment with risedronate in women with postmenopausal osteoporosis. *Calcif Tiss Int* **75(6):**462–8.
37. Cranney, A., Tugwell, P., Adachi, J. *et al*. (2002). III. Meta-analysis of risedronate for the treatment of postmenopausal osteoporosis. *Endocr Rev* **23:**517–23.
38. Boonen, S., McClung, M.R., Eastell, R. *et al*. (2004). Safety and efficacy of risedronate in reducing fracture risk in osteoporotic women aged 80 and older: implications for the use of antiresorptive agents in the old and oldest old. *J Am Geriatr Soc* **52:**1832–9.
39. Cohen, S., Levy, R.M., Keller, M. *et al*. (1999). Risedronate therapy prevents corticosteroid-induced bone loss: a twelve-month, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum* **42:**2309–18.
40. Reid, D.M., Hughes, R.A., Laan, R.F.J.M. *et al*. (2000). Efficacy and safety of daily risedronate in the treatment of corticosteroid-induced osteoporosis in men and women: a randomized trial. *J Bone Miner Res* **15:**1006–13.
41. Wallach, S., Cohen, S., Reid, D.M. *et al*. (2000). Effects of risedronate treatment on bone density and vertebral fracture in patients on corticosteroid therapy. *Calcif Tiss Int* **67:**277–85.
42. Harris, S.T., Watts, N.B., Li, Z. *et al*. (2004). Two-year efficacy and tolerability of risedronate once a week for the treatment of women with postmenopausal osteoporosis. *Curr Med Res Opin* **20:**757–64.
43. Reid, D.M., Hosking, D., Kendler, D. *et al*. (2006). Alendronic acid produces greater effects than risedronic acid on bone density and turnover in postmenopausal women with osteoporosis: results of FACTS1-International. *Clin Drug Invest* **26:**63–74.
44. Taggart, H., Bolognese, M.A., Lindsay, R.L. *et al*. (2002). Upper gastrointestinal safety of risedronate: a pooled analysis. *Mayo Clinic Proc* **77:**262–70.

45. Chesnut, C., 3rd, Skag, A., Christiansen, C. *et al*. (2004). Effects of oral ibandronate administered daily or intermittently on fracture risk in postmenopausal osteoporosis. *J Bone Miner Res* **19:**1241–9.
46. Adami, S., Felsenberg, D., Christiansen, C. *et al*. (2004). Efficacy and safety of ibandronate given by intravenous injection once every 3 months. *Bone* **34:**881–9.
47. Epstein, S., Delmas, P.D., Emkey, R. *et al*. (2006). Oral ibandronate in the management of postmenopausal osteoporosis: Review of upper gastrointestinal safety. *Maturitas* **54:**1–10.
48. Neer, R.M., Arnaud, C.D., Zanchetta, J.R. *et al*. (2001). Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* **344:**1434–41.
49. Rubin, M.R., Cosman, F., Cosman, F. *et al*. (2002). The anabolic effects of parathyroid hormone. *Osteoporos Int* **13:**267–77.
50. Jiang, Y., Zhao, J.J., Mitlak, B.H. *et al*. (2003). Recombinant human parathyroid hormone (1–34) [Teriparatide] improves both cortical and cancellous bone structure. *J Bone Miner Res* **18:**1932–41.
51. Ettinger, B., San Martin, J., Crans, G. *et al*. (2004). Differential effects of Teriparatide on BMD after treatment with raloxifene or alendronate. *J Bone Miner Res* **19:**745–51.
52. Lane, N.E., Sanchez, S., Modin, G.W. *et al*. (1998). Parathyroid hormone treatment can reverse corticosteroid-induced osteoporosis. Results of a randomized controlled clinical trial. *J Clin Invest* **102:**1627–33.
53. Orwoll, E.S., Scheele, W.H., Paul, S. *et al*. (2003). The effect of teriparatide [Human Parathyroid Hormone (1–34)] therapy on bone density in men with osteoporosis. *J Bone Miner Res* **18:**9–17.
54. Marie, P.J., Ammann, P., Boivin, G. *et al*. (2001). Mechanisms of action and therapeutic potential of strontium in bone. *Calcif Tiss Int* **69:**121–9.
55. Marie, P.J. (2005). Strontium ranelate: a novel mode of action optimizing bone formation and resorption. *Osteoporos Int* **16:**S7–S10.
56. Nielsen, S.P., Slosman, D., Sorensen, O.H. *et al*. (1999). Influence of strontium on bone mineral density and bone mineral content measurements by dual X-ray absorptiometry. *J Clin Densitom* **2:**371–9.
57. Meunier, P.J., Roux, C., Seeman, E. *et al*. (2004). The effects of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. *N Engl J Med* **350:**459–68.
58. Reginster, J.Y., Seeman, E., De Vernejoul, M.C. *et al*. (2005). Strontium ranelate reduces the risk of nonvertebral fractures in postmenopausal women with osteoporosis: treatment of peripheral osteoporosis (TROPOS) study. *J Clin Endocrinol Metab* **90:**2816–22.
59. Roux, C., Reginster, J.-Y., Fechtenbaum, J. *et al*. (2006). Vertebral fracture risk reduction with strontium ranelate in women with postmenopausal osteoporosis is independent of baseline risk factors. *J Bone Miner Res* **21:**536–42.
60. Committee for Proprietary Medicinal Products (CPMAP). European Public Assessment Report. Protelos: scientific discussion [online] http://www.emea.eu.int/humandocs/PDFs/EPAR/protelos/121604en6.pdf (accessed 18 April 2006).
61. Heaney, R.P. (2000). Calcium, dairy products and osteoporosis. *J Am Coll Nutr* **19:**83S–99.
62. Dawson-Hughes, B., Harris, S.S., Krall, E.A. *et al*. (1997). Effect of calcium and vitamin D supplementation on bone density in men and women 65 years of age or older. *N Engl J Med* **337:**670–6.
63. Cumming, R.G. and Nevitt, M.C. (1997). Calcium for prevention of osteoporotic fractures in postmenopausal women. *J Bone Miner Res* **12:**1321–9.

64. Department of Health (1998). Nutrition and bone health: with particular reference to calcium and vitamin D. Report on Health and Social Subjects, 49. London: Department of Health.
65. Porthouse, J., Cockayne, S., King, C. *et al*. (2005). Randomised controlled trial of calcium and supplementation with cholecalciferol (vitamin D3) for prevention of fractures in primary care. *BMJ* **330:**1003–6.
66. The Record Trial Group (2005). Oral vitamin D3 and calcium for secondary prevention of low-trauma fractures in elderly people (Randomised Evaluation of Calcium Or vitamin D, RECORD): a randomised placebo-controlled trial. *Lancet* **365:**1621–8.
67. Jackson, R.D., LaCroix, A.Z., Gass, M. *et al*. (2006). Calcium plus vitamin D supplementation and the risk of fractures. *N Engl J Med* **354:**669–83.
68. Michaelsson, K., Melhus, H., Bellocco, R. *et al*. (2003). Dietary calcium and vitamin D intake in relation to osteoporotic fracture risk. *Bone* **32:**694–703.
69. Avenell, A., Gillespie, W., Gillespie, L. *et al*. (2005). Vitamin D and vitamin D analogues for preventing fractures associated with involutional and post-menopausal osteoporosis. *The Cochrane Database of Systematic Reviews* Issue 3. CD000227.
70. Prince, R.L., Devine, A., Dhaliwal, S.S. *et al*. (2006). Effects of calcium supplementation on clinical fracture and bone structure: results of a 5-year, double-blind, placebo-controlled trial in elderly women. *Arch Intern Med* **166:**869–75.
71. Brennan, J., Johansen, A., Butler, J. *et al*. (2003). Place of residence and risk of fracture in older people: a population-based study of over 65-year-olds in Cardiff. *Osteoporos Int* **14:**515–19.
72. Cali, C.M. and Kiel, D.P. (1995). An epidemiologic study of fall-related fractures among institutionalized older people. *J Am Geriatr Soc* **43:**1336–40.
73. Chapuy, M.C., Arlot, M.E., Duboeuf, F. *et al*. (1992). Vitamin D3 and calcium to prevent hip fractures in the elderly women. *N Engl J Med* **372:**1637–42.
74. Chapuy, M.C., Arlot, M.E., Delmans, P.D. *et al*. (1994). Effect of calcium and cholecalciferol treatment for three years on hip fractures in elderly women. *BMJ* **308:**1081–2.
75. Lilliu, H., Pamphile, R., Chapuy, M.-C. *et al*. (2003). Calcium-vitamin D3 supplementation is cost-effective in hip fractures prevention. *Maturitas* **44:**299–305.
76. Lips, P., Graafmans, W.C., Ooms, M.E. *et al*. (1996). Vitamin D supplementation and fracture incidence in elderly persons: a randomized, placebo-controlled clinical trial. *Ann Intern Med* **124:**400–6.
77. Flicker, L., MacInnis, R.J., Stein, M.S. *et al*. (2005). Should older people in residential care receive vitamin D to prevent falls? Results of a randomized trial. *J Am Geriatr Soc* **53:**1881–8.
78. Bischoff-Ferrari, H.A., Willett, W.C., Wong, J.B. *et al*. (2005). Fracture prevention with vitamin D supplementation: a meta-analysis of randomized controlled trials. *JAMA* **293:**2257–64.
79. Dawson-Hughes, B., Heaney, R.P., Holick, M.F. *et al*. (2005). Estimates of optimal vitamin D status. *Osteoporos Int* **16:**713–16.
80. MacLaughlin, J. and Holick, M.F. (1985). Aging decreases the capacity of human skin to produce vitamin D3. *J Clin Invest* **76:**1536–8.
81. Holick, M.F., Siris, E.S., Binkley, N. *et al*. (2005). Prevalence of vitamin D inadequacy among postmenopausal North American women receiving osteoporosis therapy. *J Clin Endocrinol Metab* **90:**3215–24.
82. Simonelli, C., Weiss, T.W., Morancey, J. *et al*. (2005). Prevalence of vitamin D inadequacy in a minimal trauma fracture population. *Curr Med Res Opin* **21:**1069–74.
83. Seton, M., Jackson, V., Lasser, K.E. *et al*. (2005). Low 25-hydroxyvitamin D and osteopenia are prevalent in persons ≥ 55 yr with fracture at any site: a prospective,

observational study of persons fracturing in the community. *J Clin Densitom* **8:**454–60.
84. Dixon, T., Mitchell, P., Beringer, T. *et al*. (2006). An overview of the prevalence of 25-hydroxy-vitamin D inadequacy amongst elderly patients with or without fragility fracture in the United Kingdom. *Curr Med Res Opin* **22:**405–15.
85. Bischoff-Ferrari, H.A., Orav, E.J. and Dawson-Hughes, B. (2006). Effect of cholecalciferol plus calcium on falling in ambulatory older men and women: a 3-year randomized controlled trial. *Arch Intern Med* **166:**424–30.
86. Youm, T., Koval, K.J. and Kummer, F.J., Zuckerman, J.D. (1999). Do all hip fractures result from a fall? *Am J Orthop* **28:**190–4.
87. Royal College of Physicians and the Bone and Tooth Society of Great Britain (2000). *Osteoporosis. Clinical guidelines for prevention and treatment. Update on pharmacological interventions and an algorithm for management*. London: Royal College of Physicians.
88. Scottish Intercollegiate Guidelines Network (2003). Management of osteoporosis: National clinical guideline 71. http://www.sign.ac.uk/guidelines/fulltext/71/index.html (accessed 7 April 2006).
89. Department of Health (2001). *National Service Framework for Older People*. London: Department of Health.
90. Tilyard, M.W., Spears, G.F., Thomson, J. *et al*. (1992). Treatment of postmenopausal osteoporosis with calcitriol or calcium. *N Engl J Med* **326:**357–62.
91. National Institute for Health and Clinical Excellence (2005). *Bisphosphonates (alendronate, etidronate, risedronate), selective oestrogen receptor modulators (raloxifene) and parathyroid hormone (teriparatide) for the secondary prevention of osteoporotic fragility fractures in postmenopausal women. Technology appraisal guidance 87*. London: NICE. Available at http://www.nice.org.uk/page=aspx?o=TA087guidance accessed 23/11/06
92. Tinetti, M.E., Speechley, M. and Ginter, S.F. (1988). Risk factors for falls among elderly persons living in the community. *N Engl J Med* **319:**1701–7.
93. Pluijm, S., Smit, J., Tromp, E. *et al*. (2006). A risk profile for identifying community-dwelling elderly with a high risk of recurrent falling: results of a 3-year prospective study. *Osteoporos Int* 17:417–25.
94. Geusens, P., Autier, P., Boonen, S. *et al*. (2002). The relationship among history of falls, osteoporosis, and fractures in postmenopausal women. *Arch Phys Med Rehabil* **83:**903–6.
95. Chang, J.T., Morton, S.C., Rubenstein, L.Z. *et al*. (2004). Interventions for the prevention of falls in older adults: systematic review and meta-analysis of randomised clinical trials. *BMJ* **328:**680.
96. Gillespie, L.D., Gillespie, W.J., Robertson, M.C. *et al*. (2003). Interventions for preventing falls in elderly people. *The Cochrane Database of Systematic Reviews* Issue 4. CD000340.
97. Feldstein, A., Elmer, P., Nichols, G. *et al*. (2005). Practice patterns in patients at risk for glucocorticoid-induced osteoporosis. *Osteoporos Int* **16:**2168–74.
98. Aspray, T.J., Stevenson, P., Abdy, S.E. *et al*. (2006). Low bone mineral density measurements in care home residents—a treatable cause of fractures. *Age Ageing* **35:**37–41.

11

The Ideal Drug for Treatment of Osteoporosis

David J. Hosking

11.1. Introduction

The ideal drug for treatment of osteoporosis is one that replaces lost bone and restores its disorganized microarchitecture so that fracture risk is reduced to that of the normal population. The drug should be free of side effects and sufficiently inexpensive for widespread use. Its beneficial effects on bone should persist for a significant time once therapy is withdrawn, opening the way to intermittent courses of treatment. It should be effective when given by a variety of routes, to improve subject acceptance and maintain compliance. Such a drug is not available, but several of the currently available compounds fulfil some (but not all) of the requirements and point the way to the future optimization of therapy. There have also been several important recent advances in our understanding of the cell biology of bone, which could also lead to the development of better treatment.

Although fracture is the main clinical consequence of osteoporosis and the ideal drug must reduce this risk, there are a number of different issues involved. Because osteoporosis is caused by a loss of bone, leading to architectural deterioration and ultimately fracture,[1] an effective drug must offset this bone loss. However, despite achieving this goal, some drugs do not reduce fracture incidence,[2] which is the ultimate test of drug response. Finally, the relevance of the drug to the management of osteoporosis must be evaluated in terms of safety, tolerability, and cost-effectiveness. The choice of drug will also depend on the use to which it will be put. Thus, the characteristics of a drug used for prevention of osteoporosis will be different from those of a drug needed for treatment of established bone loss. It might, therefore, be unrealistic to expect a single compound to fulfil both these requirements.

There are a number of general characteristics required for an effective drug for the management of osteoporosis, which are summarized below:

1. The drug must increase bone mass
2. There must be a dose–response relationship

3. The drug must increase bone mineral density (BMD) at all skeletal sites prone to fracture
4. The drug must reduce fracture incidence in clinical trials
5. The drug must be safe and tolerable
6. The drug must work in all types of osteoporosis.

11.2. Increase in Bone Mass

Loss of bone is the initiating factor in the development of all types of osteoporosis and all effective treatments must reverse this process. However, the amount by which this must occur depends on clinical circumstances. Prevention of osteoporosis requires that a normal bone mass is preserved in the presence of a stimulus that would otherwise cause bone to be lost—for example the onset of the menopause or the introduction of high-dose glucocorticoid therapy. In a strict sense, bone is not "gained" but its loss is "prevented". In practice, many women starting postmenopausal hormone-replacement therapy (HRT) for prevention of osteoporosis will achieve a real gain in bone of 3–5% during the first few years because of the perimenopausal increase in bone turnover and effects of antiresorptive therapy. However, if the amount of bone loss prevented is incorporated (approximately 1–2% per annum), 10 years of HRT could result in an overall gain in bone mass of 13–14%.[3]

By contrast, the treatment of established bone loss requires an agent that will restore all or part of the deficit, although it is recognised that architectural destruction cannot always be repaired. In this context, the increase in bone mass is equal to the sum of the bone gained plus the amount of bone loss prevented (Figure 11.1).

These different requirements will influence the choice of therapeutic agent and this is, generally, wider for prevention than treatment. Although the latter, generally, requires the most powerful agents, in terms of an effect on the bone remodelling cycle, prevention can be achieved with either the maximum dose of a weak drug or a low dose of a powerful drug.

11.2.1. Mechanism of Bone Gain (Table 11.1)

Bone loss in osteoporosis is caused by an imbalance between bone resorption and formation, which are the two linked components of the bone remodelling cycle by which skeletal microdamage (stress fracture) is repaired. In most situations, particularly in the postmenopausal woman, there is an excess of resorption compared with formation. By contrast, glucocorticoids and excessive alcohol cause osteoporosis largely thr ough inhibition of bone formation. All current therapies for osteoporosis function by modifying bone remodelling. Ideally, the choice would be for a drug that stimulated bone formation and, although several agents are under clinical investigation, only teriparatide (parathyroid hormone [PTH])[48] is

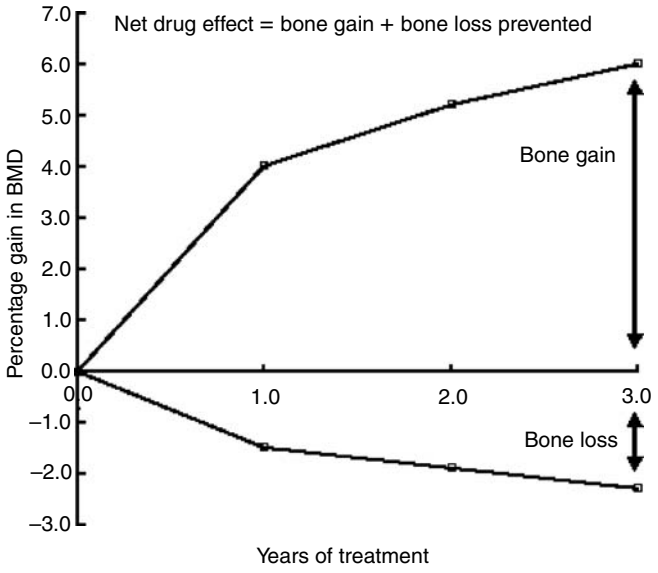


Figure 11.1. Diagram to show the concept of bone loss prevented and bone gain.

Table 11.1. Mechanisms of bone gain with antiresorptive drugs.

Temporary uncoupling of bone turnover
Infilling of remodelling space
Preservation of microarchitecture/formation surfaces
More complete secondary mineralization
Decreased resorption depth/increase mean wall thickness

Table 11.2. Drugs for the treatment of osteoporosis.

Inhibitors of bone resorption	Oestrogens
	Selective oestrogen receptor modulators (SERMs)
	Bisphosphonates
	Vitamin D/hydroxylated metabolites
	Antibody to RANK ligand*
	Cathesin K antagonist*
Stimulators of bone formation	Parathyroid hormone
	Strontium ranelate

* Drugs in current clinical trials.

currently commercially available (Table 11.2). As a consequence, the majority of drugs currently used for osteoporosis treatment are those that inhibit bone resorption. Although a more recent development with drugs such as strontium ranelate has been the introduction of "dual-action" agents, which seem to inhibit bone resorption and stimulate bone formation.[54]

Bone remodelling follows an orderly sequence (Figure 11.2), whereby resorption is followed by formation. These two processes are linked through the local production of cytokines, which depends on the process of bone resorption. As bone resorption becomes progressively inhibited by drug therapy, the flow of cytokines also diminishes, in addition to decreased bone formation. An inevitable consequence of the use of antiresorptive drugs is that, because overall bone turnover decreases, it limits the potential for bone gain to a relatively short period of 3 to 5 years.[4,5] By contrast, drugs that stimulate bone formation directly are not limited in this way and seem to have the potential to cause much larger gains in BMD.[6]

All antiresorptive drugs produce their major effect in the first year or so of treatment (most clinical trials have evaluated responses at 6 and/or 12 months). This, again, is related to the characteristics of the bone remodelling cycle. Because osteoclastic resorption declines under the influence of drug therapy, there will, as stated above, be a corresponding decline in bone formation. However, whereas osteoclasts resorb bone over a period of a few weeks, osteoblasts refill the resorption lacunae over a period of about 3 months. As a consequence, there will be a temporal dissociation between a rapid inhibition of bone resorption and a much slower inhibition of bone formation. This leads to a temporary (and progressively declining) gain in bone owing to an excess of bone formation compared with resorption.

Another important mechanism by which bone is gained early results from the finite interval between the end of bone resorption and the beginning of bone

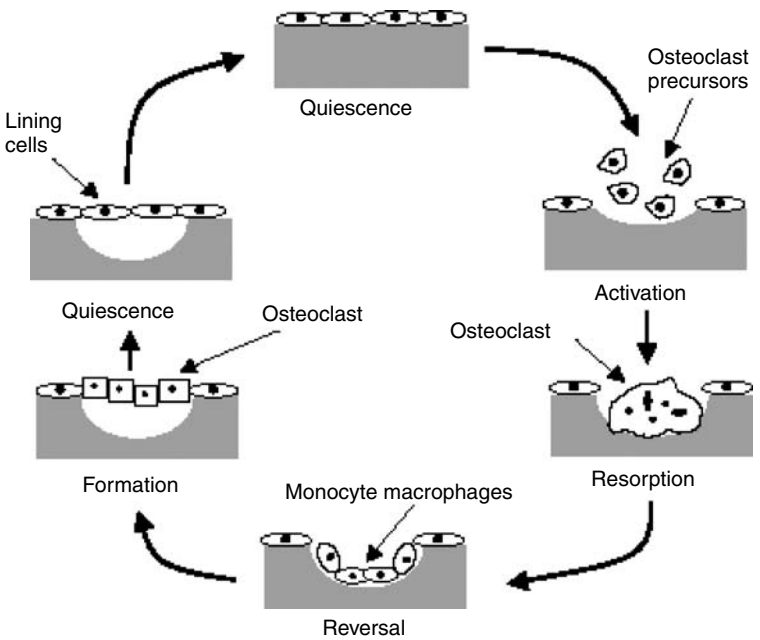


Figure 11.2. The bone remodelling cycle.

formation (reversal). At this point, the resorption lacunae have been excavated by the osteoclast (which has either moved away or undergone apoptosis), but until the cavity is filled in by osteoblastic bone formation, it represents an actual (but ultimately remedial) deficit of bone termed "the remodelling space".[7] As osteoclastic bone resorption is progressively inhibited by drug therapy, the size of the remodelling space decreases. This is reflected in a real, but time-limited, gain in bone mass, until resorption reaches a new steady state. The amount of gain in bone mass owing to this process will depend on the antiresorptive activity of a particular drug and this is summarized in Table 11.3. Because remodelling is much more active in trabecular, compared with cortical bone, the gain at the former site will be greater.

Several studies have shown continuing gain of bone mass after the first year of therapy and this probably relates to progressive mineralization of bone while it matures.[8] When bone first mineralizes, only approximately 80% of the final amount is deposited within the first few months (primary mineralization). As bone ages, it becomes more densely mineralized (without further change in its trabecular dimensions). One of the consequences of the suppression of bone resorption is that bone tends to become older, because there is less chance of it being remodelled, and its mineralization will, therefore, become more complete.

Finally, there is the possibility that, with inhibition of osteoclastic resorption, there will also be a decrease in the depth of the resorption lacunae, which, when refilled by osteoblastic formation, results in an increase in mean wall (trabecular) thickness, but such changes are difficult to demonstrate histomorphometrically.

Another possible consequence of osteoclast inhibition is that there will be less chance of trabecular perforation. This has a disproportionate effect on mechanical integrity relative to bone mass but its reduction by antiresorptive therapy preserves surfaces on which bone can be reformed. As such, it makes a minor contribution to the gain of bone mass.

The gain in BMD with antiresorptive therapy is owing to the temporary excess in bone formation compared with resorption, because of a reduction in osteoclast function. Anabolic agents, such as PTH or the N-terminal fragment PTH(1–34), build bone by stimulating osteoblastic bone formation without the need for prior resorption, leading to deposition of new bone on the periosteal and endosteal

Table 11.3. Relative potency of drugs for osteoporosis treatment.

		Percentage gain in BMD at 3 years		
Drug	Optimal dose	Lumbar spine	Femoral neck	Reference
Etidronate	400 mg	4.7	1.0	43
Alendronate	70 mg/week	6.8	4.8	5
Risedronate	35 mg/week	5.4	1.6	21
Ibandronate	150 mg/month	4.9	2.3	69
Zoledronate	4 mg IV			71
HRT	0.625 mg	5.1	3.2	44
Raloxifene	60 mg	2.6	2.1	24

surfaces of bone.[45–47] The effect of this action is to increase the area of the cortex so that the same applied load (from a fall, for example) imposes less stress on the bone. The evidence from an animal model of osteoporosis is that treatment with PTH(1–34) also increases the thickness and connectivity of existing trabeculae.[45] Note that PTH at doses given in current formulations is bone-anabolic.[50] This should not be confused with the consistent, high PTH levels found in primary hyperparathyroidism, which lead to osteoporosis. It seems that the small stimulation caused by a bolus injection of PTH once daily is sufficient to stimulate the anabolic effect on osteoblasts without enhancing excessive osteoclastic action.

11.2.2. Measurement of Response

Although bone gain can be measured by dual-energy absorptiometry (DXA), the technique only has sensitivity for measuring a real change in BMD of approximately ± 3.5% at the lumbar spine. This means that it is only practical to make measurements after approximately 1 year of treatment. Biochemical markers of bone turnover are more immediately responsive and provide an early insight into the changes in turnover, which should be subsequently translated into bone gain. However, the sensitivity of current bone markers in reliably identifying changes within an individual are relatively poor, although results are much better in group studies. Bone resorption falls by 40% to 60% within 6 weeks of treatment with the most powerful antiresorptive agents, and a subsequent reduction in formation markers reaches a nadir after approximately 12 weeks.[9,10] This underlies the transient uncoupling of resorption and formation, which contributes to bone gain and also confirms response to, and compliance with, treatment. Biochemical markers are of less value in monitoring the early response to weaker inhibitors of bone resorption, which rarely reduce bone turnover by 40%. In these circumstances, measurement of BMD by DXA is the only available option. Treatment with strontium ranelate presents a unique problem because strontium has a higher atomic number than calcium so bone crystals containing strontium seem more radiodense than hydroxyapatite and thus increases in BMD with this agent overestimate the true gain in BMD.[54] For the same reasons, bone turnover markers have been used to monitor the changes that occur when treatment is withdrawn. This is of considerable practical and theoretical importance in trying to identify crucial pathophysiological mechanisms. Antiresorptive agents can influence the osteoclast either through cell-surface receptors (calcitonin, oestrogen, and vitamin D) or through uptake during the process of bone resorption (bisphosphonates). It was hoped that the long half-life of bisphosphonates in bone, owing to their adsorption to hydroxyapatite crystals,[11] might prolong their inhibitory osteoclastic effect once treatment was withdrawn. Current data using BMD and/or biochemical markers of bone turnover show that this is a variable effect depending on the drug dose and duration of treatment.[10,12,55,56] If administered at a low dose for a short duration, the inhibitory effect of bisphosphonates seems to wear off only slightly more slowly than that of hormones such as oestrogen.[12,57,58] There is little evidence of the maintenance of BMD following the withdrawal of

anabolic agents. In animals, the changes produced are lost once treatment is stopped.[45,51–53] At the end of treatment, the new bone does not seem to be completely mineralized. Current experience suggests that mineralization will continue once treatment is withdrawn provided that an increase in bone resorption can be prevented by subsequent antiresorptive therapy.[59] In women taking PTH(1–34) in combination with established HRT, the gain in BMD was maintained 1 year after discontinuation of PTH(1–34).[49] Whether, in the future, sequential therapy with an anabolic agent, such as PTH, is followed by treatment with a bisphosphonate or a selective oestrogen receptor modulator (SERM) to complete the mineralization of new bone will move closer to the ideal is not clear at present. The initial evidence, from a study of PTH(1–34) in combination with established HRT[49] and a further study of 12 months of PTH followed by 12 months of alendronate treatment,[50] is encouraging.

11.2.3. Clinical Significance of Bone Gain

There is not a close relationship between the amount of bone gained and reduction of new vertebral fracture, but there is a more consistent, positive relationship with respect to nonvertebral fracture.[60,61] For example, both raloxifene (a SERM)[14] and alendronate (a bisphosphonate)[15] reduce the incidence of new vertebral fracture by 50% despite gains in lumbar spine BMD at 3 years of 3.2% and 8.6%, respectively. However, as above, the main contribution to reduction of this type of fracture risk is infilling of the remodelling space and prevention of trabecular perforation. Prevention of nonvertebral fracture largely depends on the inhibition of endosteal resorption, which is the major process by which cortical thickness is reduced with ageing. Endosteal bone turnover is extremely rapid[62] and requires a potent antiresorptive agent for its control. This can only be achieved by potent nitrogen-containing bisphosphonates, such as alendronate and risedronate.[63]

Another issue is that of the relationship between bone gain and bone quality. This is illustrated by a recent study of the role of nasal calcitonin Prevent Recurrence of Osteoporotic Fractures (PROOF) study[16] and an older study of high-dose sodium fluoride.[2] Over a 3-year period, a dose of 200 IU of nasal salmon calcitonin increased lumbar spine BMD by 1.4% but reduced the incidence of new vertebral fracture by 36%. Bone gain was an unlikely explanation for this response, which was attributed to an improvement in bone quality, although this cannot be assessed by current techniques. The converse situation arose with sodium fluoride, which produced large increases in lumbar spine BMD (10% per annum) but did not reduce the incidence of vertebral fracture. This was thought to be because of the incorporation of fluoride into the hydroxyapatite crystals of bone, which, because fluorapatite is known to be brittle, might explain the lack of effect on fracture. These studies emphasise that an increase in BMD cannot be taken as a surrogate market for fracture reduction, nor does it necessarily follow that a failure to gain bone will not reduce fracture risk.

There was initial concern that the bone laid down during bisphosphonate therapy might be structurally abnormal because of the development of defective mineralization with high doses of the first generation bisphosphonate etidronate.[17] Subsequent studies with lower doses of etidronate (400 mg/day for 2 weeks every 3 months) showed that this was not a problem.[18] Alendronate treatment in baboons showed that the bone of the lumbar vertebrae formed during treatment was of normal structure and mechanical integrity.[18] Bone biopsies from subjects with Paget's disease who were treated with alendronate also showed the deposition of normal lamella bone during therapy.[19] There is, therefore, current evidence that the bone formed during treatment with currently available therapies is structurally normal. However, none of these therapies will restore the mechanical integrity of perforated trabeculae and this poses a limitation on the extent to which a gain in bone *per se* will reduce fracture risk to normal. Studies with PTH(1–34) have shown that bone deposition is increased in both animal models[47] and a biopsy study in humans[46] and that the trabecular bone microarchitecture is improved. These studies provide an understanding of the structural basis of the changes to bone architecture following treatment with PTH(1–34), providing supporting evidence that an anabolic-induced increase in BMD offsets the fracture propensity owing to architectural destruction.[48] Development of therapies based on cytokines and growth factors that modulate bone remodelling at the cellular level might offer the prospect of renewal of the microarchitecture, another requirement for the ideal drug.

11.3. Dose Response

Most fracture trials in osteoporosis have involved a prior phase II dose-ranging study to identify the optimum dose of the new drug. This usually involves three dose levels, with a minimum and increments at two and four times the minimum dose. In most studies, the maximum dose beyond which there is no additional benefit is considered as the optimum dose (e.g. alendronate), whereas in other studies in which the difference between the maximum effective dose and the next dose level down is relatively small, the lower dose might be chosen for clinical use (e.g. raloxifene).

The dose response can also be determined by the relationship of the dose to adverse events, for example dose–response studies with etidronate showed that there was a risk of defective mineralization with the most effective doses (10–20 mg/kg body weight/day), which limited treatment to a less effective dose of 5 mg/kg body weight/day (400 mg/day).[17] It also was the rationale for the use of cyclical etidronate for treatment of osteoporosis whereby the drug is given for 2 weeks every 13 weeks to avoid this side effect.[20] A similar rationale was the basis for the choice of 20 mcg as the optimum dose of teriparatide, for which there were small benefits compared with the larger 40 mcg dose but a greater incidence of adverse events.[48] The subsequently introduced (second and third generation) bisphosphonates have greater potency for inhibiting bone resorption at doses that

do not inhibit mineralization and this issue has not had a limiting effect on choice of optimum dose with the newer compounds. These issues are well illustrated by the phase III studies of alendronate at doses of 5 mg/day, 10 mg/day, and 20 mg/day[15] which demonstrated that the two higher doses produced an identical response, indicating that a dose of alendronate, 10 mg/day was optimal. By contrast, risedronate was only studied at two doses (2.5 mg/day and 5 mg/day); however, although a dose–response relationship was clearly identified, the maximum effective dose is unknown.[21] Typical dose responses for alendronate and raloxifene are given in Table 11.4. Two anabolic agents have also demonstrated a dose–response relationship in clinical trials (Table 11.4): in one study in which subjects received either 20 μg or 40 μg of PTH(1–34) compared with placebo[48] and another study in which subjects received either 50 μg, 70 μg or 100 μg of PTH.[50]

If the dose range has been well chosen, the phase III study should enable selection of the optimum dose according to clinical need. In general, smaller doses of antiresorptive drugs are needed for prevention of osteoporosis (maintenance of BMD) than treatment of the established disease, for which a gain in BMD is required. The dose–response study can also provide an indication of the probable maintenance dose once the maximum gain in BMD has been achieved. Finally, the dose–response study will indicate whether there are importance differences in the time course of bone gain. In general, most phase III studies of the major classes of antiresorptive drug show the same pattern of response (Figure 11.3a). By contrast, in a chronic condition such as osteoporosis, a suboptimal early response from a relatively low dose of a new drug that is compensated for by a more prolonged later phase of gain might indicate that a relatively low dose might be optimal (Figure 11.3b). The dose response is usually assessed by changes in bone turnover markers or BMD because of their greater sensitivity and, therefore, the need for smaller clinical trials.

Table 11.4. Dose response with currently available drugs in terms of gain in lumbar spine BMD.

Drug	Treatment duration (years)	Dose (mg)	Percentage gain in BMD	Reference
Alendronate	3	5	5.2	15
		10	8.2	
		20	7.8	
Raloxifene	2	30	1.3	14
		60	1.6	
		150	2.2	
PTH(1–34)	2	20 μg	9.0	48
		40 μg	13.0	
PTH	1	50 μg	4.3	50
		75 μg	6.9	
		100 μg	9.2	

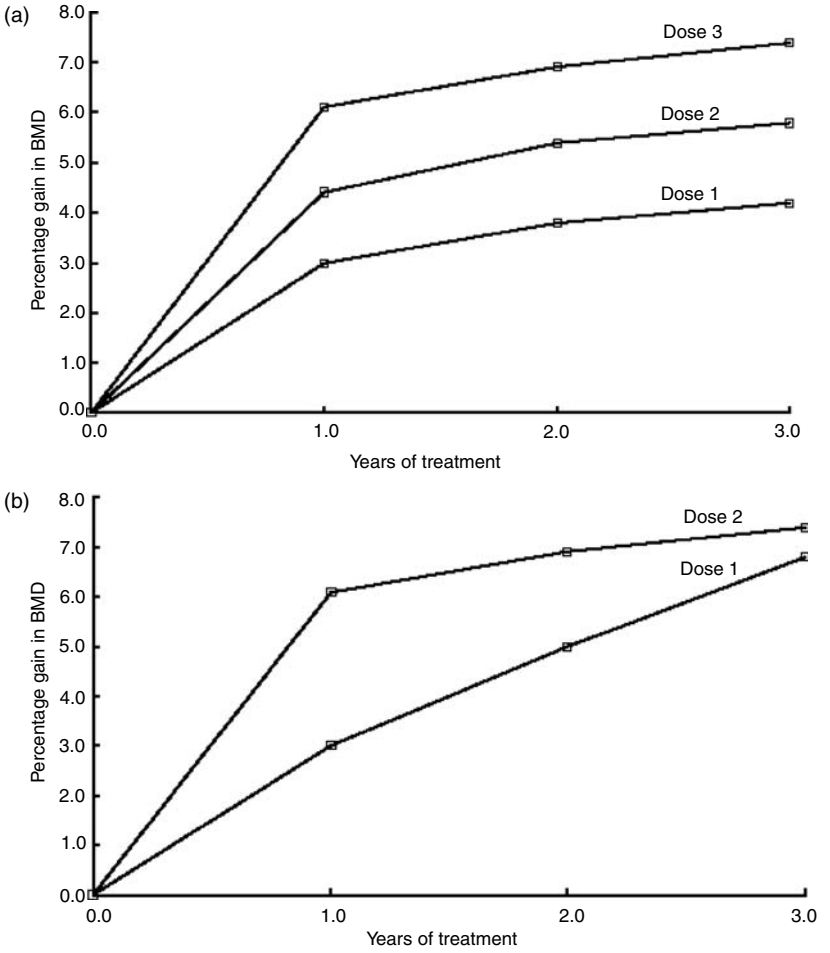


Figure 11.3. Examples of the change in BMD with different doses **(a)** The pattern of dose response if the shape of the dose response does not vary with dose. **(b)** The pattern of dose response if, at lower doses, there is prolonged gain in BMD.

Nasal salmon calcitonin and PTH(1–34) have been used in dose–response studies using fracture endpoints[16,48] and it illustrates some of the problems that could follow. In the calcitonin study, there was no significant difference between the gain in BMD at the lumbar spine in the chosen dose range (100 IU, 200 IU, and 400 IU). However, when fracture reduction at the lumbar spine was evaluated, there was a significant reduction of new fracture, by 36% in subjects treated with the 200 IU dose, whereas there was no benefit from the lower or higher doses. The failure of the 400 IU dose to show a response at least equal to the 200 IU dose response presents a problem in understanding the mechanisms of fracture reduction with this agent.

A similar effect was found in the PTH(1–34) study, in which women with prior vertebral fracture were randomized to receive 20 μg or 40 μg PTH(1–34) or

placebo. The incidence of new vertebral fracture showed a 75% reduction relative to placebo, but no significant dose response, despite a dose response with regard to BMD.[48] There was also no difference in the reduction of nonvertebral fracture between the 20 μg and 40 μg groups (38% at both doses).

The dose–response relationship for the ideal drug should show a clear maximal dose effect (which implies little difference between the two highest doses). The lowest dose should also show a significant benefit relative to placebo, to give scope for choice, depending on whether the clinical goal is prevention or treatment of osteoporosis. The dose-ranging study should be of sufficient duration for the drug to reach a steady-state concentration, to indicate the duration of treatment needed to achieve the maximum gain in BMD. Although it would be ideal to assess dose response using fracture endpoints, rather than using changes in bone turnover markers or BMD, this is unrealistic because of the huge size (and cost) of the clinical trials that would be necessary.

11.4. Distribution of Bone Gain Within the Skeleton

Measurement of BMD at a particular region in the skeleton is the best predictor of fracture at that site[22] and it follows, therefore, that the ideal osteoporotic drug should increase BMD at all sites prone to fracture. To understand the implications of this requirement, it is necessary to distinguish between the remodelling rates of trabecular and cortical bone. The skeleton contains 80% of its content as dense cortical bone that remodels at approximately 20% of the rate of trabecular bone. The consequence is that changes in BMD with treatment, or disease evolution, are generally much more obvious at trabecular sites, such as the spine, compared with cortical sites, such as the hip. However, the importance of the hip in terms of the mortality and morbidity of osteoporotic fracture[23] makes it crucial that new drugs are effective at the hip (and other cortical sites, such as the appendicular skeleton). A number of specific issues have emerged from recent clinical trials of osteoporotic treatments. These centre around poor, or absent, bone gain at the femoral neck, lack of change in BMD at the distal radius, and redistribution of bone from cortical to trabecular sites.

Owing to the low rate of remodelling of cortical bone, it is important that clinical trials extend to a period of at least 3 years, so that the full effect of a new drug can be assessed as the bone approaches a new steady state. In these circumstances, the most powerful antiresorptive drugs increase femoral neck BMD by approximately 6% over 3 years.[15,21] The crucial increase in BMD, above which the incidence of hip fracture will decrease, is unknown. A recent study of raloxifene achieved a BMD gain at the femoral neck of 2.1–2.4% but this was not associated with a reduction in hip or appendicular fracture[24] and, perhaps, gives some indication of the magnitude of change required to reduce the incidence of fracture in cortical bone.

Distal radial (Colles) fracture is also a common consequence of osteoporosis and there might be important qualitative and quantitative differences between the classes of antiresorptive agent at this site. This is illustrated by the Early Postmenopausal

Intervention Cohort (EPIC) study, which compared the abilities of HRT and low-dose alendronate in preventing early postmenopausal bone loss.[12] Despite similar effects of HRT and alendronate (5 mg/day) in preventing bone loss at the lumbar spine and hip, alendronate was only able to attenuate, but not prevent, bone loss at the distal radius, whereas HRT was effective. This could be a dose-dependent effect, that is alendronate (10 mg/day) prevented bone loss at all skeletal sites in older subjects in the phase III study.[15] However, this data cannot necessarily be extrapolated to the early postmenopausal population and the failure of alendronate could reflect an important qualitative difference between HRT and bisphosphonates.

Finally, there is the issue of redistribution of bone mineral from one site to another during osteoporosis treatment. This was seen in early studies of sodium fluoride[2] and might underlie the occasional development of intracortical fracture with low-dose sodium monofluorophosphate.[25] In the early sodium fluoride studies, some of the large gains in BMD at the lumbar spine seem to be achieved by redistribution of bone mineral from femoral neck.

This problem seems to have been avoided by the use of lower doses of fluoride (15 mg/day), which, however, only increased BMD by 2.33% at the femoral neck compared with 12.6% at the lumbar spine.[25] However, 10% of the subjects in this study developed painful intracortical fractures at the distal end of the weight-bearing bones and this could reflect mobilization of cortical bone, although BMD was not measured at these sites. There were similar concerns about the use of synthetic human PTH [hPTH(1–34)] for treatment of osteoporosis because primary hyperparathyroidism seems to be associated with a preferential loss of cortical bone.[26] However, a recent clinical trial of PTH(1–34) showed a significant gain in BMD at the lumbar spine (13.0%) and femoral neck (2.7%) and an 8.0% gain in total body bone mineral content (80% cortical bone),[6] although other cortical sites were not measured.

The ideal drug for osteoporosis treatment must, therefore, be capable of increasing BMD at all the skeletal sites prone to fracture. The practical implication is that BMD must be measured at both the lumbar spine and the hip to evaluate the relative responses of trabecular and cortical bone. Ideally, measurements should also be made at the distal radius because this cortical site can respond differently to some therapeutic agents. Finally, measurement of total body BMD (predominantly reflecting changes in cortical bone) is a useful adjunct to confirm a real increase in BMD, rather than the redistribution of bone. Because fracture risk doubles for every 1 standard deviation (SD) decrease in BMD, the final gain in bone mass at each fracture site should be approximately 10%, to equate with a 50% reduction of fracture risk.

11.5. Fracture Incidence

This is the most crucial and stringent test of a drug's relevance to treatment of osteoporosis. It has a profound influence on the design of clinical trials, which must focus on high-risk subjects to contain the numbers of subjects required to show that the new treatment reduces the risk of new fracture. In general,

this means that such trials focus on the elderly (>65 years of age) who have already lost substantial amounts of bone and have experienced at least one prior fracture.[16,20,21,24,25,27]

All trials evaluate the antifracture efficacy of the new drug against placebo, but because it is unethical to use a true placebo (an agent that has no effect on bone) now that treatment is available to reduce fracture incidence, the placebo (often calcium with or without vitamin D) will have intrinsic, but usually weak, antifracture efficacy.[28,29] No recent trial has compared a new drug with the currently best available therapy because the difference is likely to be small and this would increase the study sample size and, therefore, its cost. Because the new drug is not tested in isolation, but in combination with placebo (test drug plus placebo versus placebo alone), logic dictates that the drug should be used in a similar combination in clinical practice. However, a reasonable compromise is to make sure that subjects are calcium and vitamin D replete (assuming that these agents comprised the placebo) before introduction of the new drug.

The next issue to be considered is the difference between the relative risk (RR) of a new fracture and the absolute reduction in the number of new fractures. Whereas the former is the usual basis for demonstrating a significant difference between test drug and placebo, it is the latter that is important in clinical practice. This has led to the introduction of the concept of the "number needed to treat" (NNT), to indicate the number of subjects who must be treated with a new drug to prevent one fracture.

This can be illustrated by a recent study of raloxifene in women with and without a previous vertebral fracture.[24] Although both groups experienced a RR reduction of a new fracture of just under 50%, the fracture rate in subjects in the placebo group without a prior fracture was 4.5% compared with 21.2% of subjects in the placebo group who had a prevalent vertebral fracture. It is, therefore, clear that more fractures would be prevented by raloxifene in the prevalent fracture group compared with those without a prior fracture. The NNT was, therefore, 16 and 46, respectively (for treatment with raloxifene, 60 mg/day), which will have a profound influence on how the drug is used in clinical practice. The ideal drug, therefore, needs the NNT to be as small as possible; where available, the NNT for current therapies are shown in Table 11.5.

Table 11.5. Number of subjects needed to treat (NNT) to prevent one vertebral or non-vertebral fracture with currently available drugs after 3 years of treatment.

Drug	T-score	Number treated	NNT, vertebral fracture	NNT, nonvertebral fracture
Etidronate	<−2.5	156	32	NS
Alendronate	<−2.0	1022	14	22
Risedronate	<−2.0	696	20	31
Raloxifene	<−2.5	769	16	NS
Nasal calcitonin	<−2.0	270	13	NS
Calcitriol		213	7	19
PTH(1–34)		1093	5	27

NS = not significant.

One further aspect of this analysis is that it is important to distinguish between trials in which there are reductions of new fractures in the treatment group with a constant fracture rate in the placebo group[21,24,27] and those in which the fracture rate is constant in the treatment group but increases in the placebo group.[30] The latter type of trial is more difficult to evaluate because the general expectation, during the relatively short period of most clinical trials (2 to 3 years), is that fracture rates in the placebo group will be relatively constant. This raises doubts about subject selection and randomization.

The crucial issue, however, when evaluating a new drug is whether the incidence of both axial and appendicular fractures is reduced. This is the major difference between currently available therapies. Only alendronate,[27] risedronate,[21] and teriparatide[48] seem to reduce both types of fracture, whereas all other drugs only reduce the incidence of axial fractures.[16,20,24,30]

The significance of this difference can usually be minimized in practice by evaluating clinical and densitometric evidence of probable the future fracture type and choosing therapy appropriately. However, the ideal drug must reduce fracture incidence at all skeletal sites. The only exception to this rule is if drugs are being evaluated for primary prevention of osteoporosis, in which it is assumed that prevention of bone loss at all sites will be translated into subsequent fracture protection. Because the actual fracture incidence in this type of population is likely to be low, clinical trials with a fracture endpoint would be prohibitively large and costly.

The advantages of stimulators of bone formation have already been considered and early indications are that they have a similar propensity to reduce fractures compared with antiresorptives (Table 11.5). In theory, there would be competing pathogenetic factors at work. Stimulators of osteoblastic activity would increase bone at sites not previously undergoing resorption, in addition to sites that have just been resorbed. Both of these mechanisms should strengthen weakened trabeculae and contribute to a reduction in fracture. However, if bone resorption is increased, as it commonly is in postmenopausal osteoporosis, this protective effect will be offset by an increase in the remodelling space.[7] The ideal treatment would, therefore, be a compound that both stimulates bone formation and inhibits bone resorption; strontium ranelate is the first of this type of agent to be introduced into clinical practice.

11.6. Compliance with Treatment

Bisphosphonates are strongly adsorbed to bone surfaces and this makes intermittent administration possible. The added convenience of the once-weekly dosing regimens of bisphosphonates[64,65] is reflected in their current dominance in the osteoporosis market. Poor compliance with daily alendronate and risedronate therapy, which was approximately 30% after 12 months, improved to 44% with once-weekly treatment.[66] The observation that ibandronate retained its effect on bone turnover after intermittent administration, with an interval between doses of 9 weeks,[67,68] opened the way for the use of once-monthly ibandronate regimens,[65] with an expectation that compliance would be further enhanced. Bisphosphonates

are poorly absorbed from the upper small intestine and the need to take these drugs while fasting is inconvenient. Cathepsin K antagonists are a new class of osteoclast inhibitor that are currently starting clinical trials; these agents do not have to be given while fasting and have the additional advantage of a once-weekly administration regimen, which might improve long-term subject acceptability.

The only certain way of ensuring good compliance is to administer the drug by injection and several approaches are currently being explored. An extension of the oral ibandronate studies investigated the use of 3-monthly intravenous injections of the drug,[70] whereas evidence has been presented that an annual intravenous injection of 4 mg of zoledronate controls bone turnover during the subsequent year.[71] Vertebral fracture efficacy will be accepted for intravenous ibandronate if it shows the same change in BMD and bone turnover as the 2.5 mg oral dose, which has been shown to protect against vertebral, but not nonvertebral, fracture (except in a small high-risk group).[68] Fracture protection with intravenous zoledronate has yet to be demonstrated. A limiting factor in the general uptake of intravenous bisphosphonates is the "acute-phase reaction", an influenza-like reaction that occurs in approximately 10% of subjects during the first 72 hours following injection. This is caused by stimulation of γδ T cells by accumulated substrates of the enzyme farnesyl diphosphate synthase, which is inhibited by nitrogen-containing bisphosphonates.[72] Strangely, an acute-phase reaction is much less prevalent after subsequent injections and the initial reaction can be modified by symptomatic treatment. It remains to be seen how this effect might influence the use of intravenous bisphosphonates in primary care, which is the setting in which most of these drugs are administered.

The most exciting development in the field of osteoporosis has been the introduction of antibodies to the receptor for activation of NF-KB (RANK) ligand, one of the key factors in activation and maturation of osteoclasts. Although suppression of bone turnover increases with the dose of antibody, its most important effect is to prolong the residency of the antibody in the circulation.[73] Clinical trials are currently underway with AMG 162, a RANK ligand antibody, which is given every 6 months by subcutaneous injection and thus has several practical advantages compared with the intravenous administration of bisphosphonates. If these trials show protection against fractures, this will offer the potential for a significant advance in long-term compliance.

11.7. Safety and Tolerability

This is a major issue because, even if a drug is very effective, it will not be used if it is associated with severe side effects. An adverse event in a clinical trial is defined by the US Food and Drug Administration (FDA) as "any untoward event occurring during the course of the trial, but these are usually subdivided into those that are probably, possibly, and possibly not, drug-related." The only objective way of deciding whether an adverse event is drug-related is as part of a double-blind, placebo-controlled clinical trial. The importance of the placebo group in arriving

at this decision is shown in Table 11.6, which summarizes the adverse events of three recent alendronate studies. Adverse events are common in the placebo group but would have been less likely to be so attributed had the clinician known that the subject was taking a placebo.

Although it is self-evident that the ideal drug should have no significant side effects, it is important to explore what this means in practice. The main aim of phase III and fracture-endpoint studies is either to identify the optimum dose or to confirm that the drug prevents fractures. The sample size is an important consideration because this influences the length of time needed for recruitment and also the cost of the study. Most pharmaceutical companies are naturally reluctant to take "all comers" into a clinical trial and prefer to exclude subjects who might be particularly prone to side effects of the new drug and could, therefore, drop out of the study. This tends to reduce the ability of the trial to identify either the nature or the frequency of important side effects that, as a consequence, could surface for the first time during postmarketing surveillance. Alendronate is a good example, because the phase III and Fracture Intervention Trial (FIT) studies[15,27] of the drug showed no excess of upper gastrointestinal side effects, perhaps because upper gastrointestinal symptoms were an exclusion criterion but these became more obvious when the drug was introduced into routine clinical practice.[31]

The converse situation can sometimes arise and this is illustrated by a recent raloxifene study where "beneficial" side effects were seen. There are two types of oestrogen receptor, α and β,[32] and drugs such as tamoxifen could be modified to act as antagonists at the oestrogen type α receptor (no oestrogen-mediated effect on the endometrium or breast) and agonists at the oestrogen type β receptor (in bone and the vasculature). As a consequence, the first of the SERMs (Raloxifene) could protect against osteoporosis through its oestrogen-mediated agonist effect, in addition to having the beneficial (oestrogen type α receptor antagonist) effect of reducing the risk of breast cancer without endometrial hyperplasia.[24]

By contrast, long-term use of HRT is associated with the risk of breast cancer[33] and there is a need to avoid endometrial hyperplasia with progestogens, which might add to the side-effect profile of these compounds. Large trials have shifted the balance between the beneficial and unwanted effects of HRT,[74,75] with a shift from this type of therapy to bisphosphonates if this balance is more acceptable.

Table 11.6. Overall safety of alendronate.

Percentage of subjects with:	FOSIT Placebo	Alendronate 10 mg	Phase III Placebo	Alendronate 10 mg	FIT Placebo	Alendronate 10 mg
Any adverse event (AE)	69.7%	67.9%	90.2%	88.3%	99.4%	99.4%
Drug related AE	18.0%	19.1%	25.4%	27.0%	9.3%	9.3%
Serious AE	6.3%	6.5%	17.4%	13.8%	32.4%	31.9%
Withdrawn	**6.4%**	**5.6%**	**6.0%**	**4.1%**	**10.1%**	**9.1%**

Furthermore, the main side effect of hydroxylated vitamin D metabolites (e.g. calcitriol, which is used in osteoporosis to suppress PTH secretion and reduce bone turnover) is hypercalcaemia.[34] Analogues of vitamin D have been developed that reduce bone turnover but do not stimulate calcium absorption, and thereby limit the risk of hypercalcaemia.[35]

In a study of the effect of PTH(1–34) on BMD and fracture, there were a greater number of withdrawals because of adverse events in the PTH(1–34)-treated group compared with the placebo-treated group.[48] Nausea and headache were the commonest reported adverse events. Osteosarcoma has been found in rats given lifelong daily injections of PTH(1–34) and increased cortical porosity has been reported. Neither of these adverse events have been reported in humans.[46,48] Most of the currently available literature on PTH concerns its N-terminal fragments, predominantly PTH(1–34). The intact PTH entity is currently in clinical development and seems to have a comparable efficacy and safety profile to PTH (1–34).[76]

For some drugs, analogues are not available but side effects can often be avoided or minimized by other means. Thus, the poor absorption of bisphosphonates can be improved by taking the drugs on an empty stomach. Similarly, the oesophageal irritation that can be associated with alendronate and risedronate can be reduced by taking the drugs with adequate water while upright. The defective bone mineralization that can occur with etidronate can be avoided by appropriate dose reduction, whereas hypercalcaemia associated with calcitriol can be reduced by taking the drug without food before bed time. Exacerbation of menopausal symptoms with raloxifene can be circumvented by avoiding the drug within the few years following the menopause. Adverse effects of oral HRT on hepatic clotting factors or the side effects of parenteral calcitonin can be reduced by alternative routes of administration (e.g. transdermal HRT or nasal calcitonin).

11.8. Efficacy in Different Types of Osteoporosis

The final requirement for the ideal treatment for osteoporosis is that it should be effective in all forms of the disease. It cannot be assumed that because osteoporosis is caused by an excess of bone resorption compared with formation that any inhibitor of resorption will be effective. This is particularly relevant to the quantitative aspect of drug response, for which it cannot be assumed that a particular drug will be equipotent in postmenopausal osteoporosis (accelerated bone resorption) and glucocorticoid-induced osteoporosis (inhibition of bone formation). Differences between the forms of the disease will also emerge with respect to fracture reduction by different structural mechanisms. In the example cited above, there are structural differences in the pattern of development of osteoporosis that could have a profound effect on a drug's ability to reduce fracture. In postmenopausal osteoporosis, the trabeculae become perforated and lose connectivity, whereas in glucocorticoid-induced osteoporosis the trabeculae become progressively thinned, with relative

preservation of connectivity.[36] For these reasons, each drug must be tested clinically in each form of osteoporosis to quantitate the gain in BMD and confirm antifracture efficacy.

The cost of osteoporosis trials, and those with fracture endpoints in particular, has meant that very few drugs have been tested across the spectrum of osteoporotic subtypes. Both etidronate and alendronate have been shown to reduce fracture in postmenopausal women[20,27] and to prevent bone loss in glucocorticoid-induced osteoporosis.[37,38] For both these drugs, the net bone gain (bone gained plus prevented bone loss) was similar in both types of osteoporosis studied.

These two agents have also been shown to be effective both in early prevention[12,39] and in treatment.[20,27] Studies evaluating bisphosphonates in male osteoporosis show that both etidronate and alendronate are capable of increasing lumbar spine BMD,[40,77] although the studies were too small to assess fracture risk.

11.9. Summary—The Future

Currently, we do not have the ideal treatment for osteoporosis, but the last decade has seen the introduction of drugs that increase BMD and reduce fracture. The main limitation of current treatment is the reliance on antiresorptive agents that have the inbuilt propensity to restrict bone gain by reducing overall bone turnover with time. The development of anabolic agents, such as PTH, which stimulate osteoblastic bone formation without the need for prior resorption, offer the potential for progressive bone gain without the limitations inherent with antiresorptive agents and might also reverse the deterioration in the bone microarchitecture that makes such an important contribution to fracture. Moreover, the growing understanding of the mechanisms by which bone formation is regulated at a cellular level[41] opens the possibility for the development of a whole new class of osteoporotic drugs. However, although it has been shown that systemic administration of PTH is effective in stimulating bone formation *in vivo*,[50–53] the major challenge with the newer agents will be to deliver them directly to the bone-forming surface. A more immediate prospect is the possibility of exploring the relationship between formation stimulators and resorption inhibitors, to evaluate their role in long-term reversal of osteoporosis.

The optimistic view is that the pace of drug development will accelerate with increasing understanding of the cellular regulation of bone remodelling. This, in turn, should improve our ability to reduce fracture more effectively and, hopefully, with less risk of adverse events.

References

1. (1993). Consensus Development Conference: diagnosis, prophylaxis and treatment of osteoporosis. *Am J Med* **94:**646–50.
2. Riggs, B.L., Hodgson, S.F., O'Fallon, W.M. *et al*. (1990). Effect of fluoride treatment on fracture rate in postmenopausal women with osteoporosis. *N Engl J Med* **22:**802–9.

3. Eiken, P., Kolthoff, N. and Pors Nielsen, S. (1996). Effects of 10 years' hormone replacement therapy on bone mineral content in postmenopausal women. *Bone* **19:**1915–35.
4. Storm, T., Killerup, G., Thamsborg, G. *et al*. (1996). Five years of clinical experience with intermittent cyclical etidronate for postmenopausal osteoporosis. *J Rheumatol* **23:**1560–4.
5. Devogelaer, J.P., Broll, H., Correa-Potter, R. *et al*. (1996). Oral alendronate induced progressive increases in bone mass of the spine, hip and total body over 3 years in postmenopausal women with osteoporosis. *Bone* **18:**141–50.
6. Lindsay, R., Nieves, J., Formica, C. *et al*. (1997). Randomised controlled study of effect of parathyroid hormone on vertebral bone mass and fracture incidence among postmenopausal women on oestrogen with osteoporosis. *Lancet* **350:**550–5.
7. Mosekilde, L., Eriksen, E.F. and Charles, P. (1991). Hypercalcaemia of malignancy: pathophysiology, diagnosis and treatment. *Crit Rev Oncol Haematol* **11:**1–27.
8. Meunier, P.J. and Bolvin, G. (1997). Bone mineral density reflects bone mass but also the degree of mineralisation of bone: therapeutic implications. *Bone* **21:**373–7.
9. Chestnut, C.H., McClung, M.R., Ensrud, K.E. *et al*. (1995). Alendronate treatment of the postmenopausal osteoporotic woman: effect of multiple dosages on bone mass and bone remodelling. *Am J Med* **99:**44–152.
10. Garnero, P., Shih, W.J., Gineyts, E. *et al*. (1994). Comparison of new biochemical markers of bone turnover in late postmenopausal osteoporotic women in response to alendronate treatment. *J Clin Endocrinol Metab* **79:**1693–700.
11. Fleisch, H. (1993). New bisphosphonates in osteoporosis. *Osteoporos Int Suppl* **2:**S15–22.
12. Ravn, P., Bidstrup, M., Wasnich, R.D. *et al*. (1999). Alendronate and estrogen-progestin prevent bone loss in healthy postmenopausal women: 4 year results from the Early Postmenopausal Intervention Cohort (EPIC) study, a randomised clinical trial. *Ann Intern Med* **11:**935–42.
13. Christiansen, C., Christensen, M.S. and Transbol, I. (1981). Bone mass after withdrawal of oestrogen replacement. *Lancet* **1:**1053–4.
14. Delmas, P.D., Bjarnason, N.H., Mitlak, B.H. *et al*. (1997). Effects of raloxifene on bone mineral density, serum cholesterol concentrations, and uterine endometrium in postmenopausal women. *N Engl J Med* **337:**1641–7.
15. Liberman, U.A., Weiss, S.R., Broll, J. *et al*. (1995). Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med* **333:**1437–43.
16. Chestnut, C.H., 3rd, Silverman, S., Andriano, K., *et al*. (2000). A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: the prevent recurrence of osteoporotic fractures study. PROOF study group. *Am J Med* **109:**267–76.
17. Khairi, M.R.A., Altman, R.D., DeRosa, G.P. *et al*. (1977). Sodium etidronate in the treatment of Paget's disease of bone. *Ann Intern Med* **87:**656–63.
18. Balena, R., Toolan, B.C., Shea, M. *et al*. (1993). The effect of 2-year treatment with the aminobisphosphonate alendronate on bone metabolism, bone histomorphometry and bone strength in ovariectomised non human primates. *J Clin Invest* **92:**2577–86.
19. Reid, I.R., Nicholson, G.C., Weinstein, R.S. *et al*. (1996). Biochemical and radiologic improvement in Paget's disease of bone treated with alendronate: A randomized, placebo-controlled trial. *Am J Med* **171:**341–8.

20. Storm, T., Thamsborg, G., Steiniche, T. *et al*. (1990). Effect of intermittent cyclical etidronate therapy on bone mass and fracture rate in women with postmenopausal osteoporosis. *N Engl J Med* **322:**1265–71.
21. Harris, S.T., Watts, N.B., Genant, H.K. *et al*. (1999). Effects of risedronate treatment on vertebral and non-vertebral fractures in women with postmenopausal osteoporosis. *JAMA* **282:**1344–52.
22. Marshall, D., Johnell, O. and Wedel, H. (1996). Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. *BMJ* **312:** 1254–9.
23. Barrett-Connor, E. (1995). The economic and human cost of osteoporotic fracture. *Am J Med* **98(Suppl 2A):**3S–8S.
24. Ettinger, B., Black, D.M., Mitlak, B. *et al*. (1999). Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: Results from a 3-year randomized clinical trial. *JAMA* **282:**637–45.
25. Ringe, J.D., Kippshoven, C., Coster, A. *et al*. (1999). Therapy of established postmenopausal osteoporosis with monofluorophosphate plus calcium: dose-related effects on bone density and fracture rate. *Osteoporos Int* **9:**171–8.
26. Adami, S., Braga, V., Squaranti, R. *et al*. (1998). Bone measurements in asymptomatic primary hyperparathyroidism. *Bone* **22:**565–70.
27. Black, D.M., Cummings, S.R., Karpf, D.B. *et al*. (1996). Randomised trial of effects of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet* **348:**1535–41.
28. Reid, I.R., Ames, R.W., Evans, M.C. *et al*. (1995). Long-term effects of calcium supplementation on bone loss and fractures in post-menopausal women: a randomized controlled trial. *Am J Med* **98:**331–5.
29. Dawson-Hughes, B., Harris, S.S., Krall, E.A. *et al*. (1997). Effect of calcium and Vitamin D supplementation on bone density in men and women 65 years of age or older. *N Engl J Med* **337:**670–6.
30. Tilyard, M.W., Spears, G.F.S., Thomson, J. (1992). Treatment of postmenopausal osteoporosis with calcitriol or calcium. *N Engl J Med* **326:**357–62.
31. De Groen, P.C., Lubbe, D.F., Hirsch, L.J. *et al*. (1996). Esophagitis associated with the use of alendronate. *N Engl J Med* **335:**1016–21.
32. Kumbar, V., Green, S., Stack, G. *et al*. (1987). Functional domains of the human estrogen receptor. *Cell* **51:**941–51.
33. Colditz, G.A., Hankinson, S.E., Hunter, D.J. *et al*. (1995). The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *N Engl J Med* **332:**1589–93.
34. Sambrook, P., Birmingham, J., Kelly, P. *et al*. (1993). Prevention of corticosteroid osteoporosis. A comparison of calcium, calcitriol, and calcitonin. *N Engl J Med* **328:**1747–52.
35. Tan, A.U., Levine, B.S., Mazess, R.B. *et al*. (1997). Effective suppression of parathyroid hormone by 1 alpha-hydroxy-vitamin D_2 in haemodialysis patients with moderate to severe secondary hyperparathyroidism. *Kidney Int* **51:**317–23.
36. Aaron, J.E., Francis, R.M., Peacock, M. *et al*. (1989). Contrasting micro-anatomy of idiopathic and corticosteroid induced osteoporosis. *Clin Orthop* **243:**294–305.
37. Adachi, J.D., Bensen, W.G., Brown, J. *et al*. (1997). Intermittent etidronate therapy to prevent corticosteroid induced osteoporosis. *N Engl J Med* **337:**382–7.
38. Saag, K., Emkey, R., Schnitzer, T.J. *et al*. (1998). Alendronate for the prevention and treatment of glucocorticoid-induced osteoporosis. *N Engl J Med* **339:**292–9.

39. Herd, R.J.M., Balena, R., Blake, G.M. *et al*. (1997). The prevention of early postmenopausal bone loss by cyclical etidronate therapy: a 2-year, double blind, placebo controlled study. *Am J Med* **103:**92–9.
40. Anderson, F.H., Francis, R.M., Bishop, D.J. *et al*. (1997). Effect of intermittent cyclical disodium etidronate on bone mineral density in men with vertebral fractures. *Age Ageing* **156:**359–65.
41. Jilka, R.L. (1998). Cytokines, bone remodeling and estrogen deficiency: a 1998 update. *Bone* **23:**75–81.
42. Jilka, R.L., Weinstein, R.S., Bellido, T. *et al*. (1999). Increased bone formation by prevention of osteoblast apoptosis with parathyroid hormone. *J Clin Invest* **104:**439–46.
43. Harris, S.T., Watts, N.B., Jackson, R.D. *et al*. (1993). Four year study of intermittent cyclical Etidronate treatment of post-menopausal osteoporosis; three years of blinded therapy followed by one year of open therapy. *Am J Med* **95:**557–67.
44. Hosking, D.J., Chilvers, C.E.D., Christiansen, C. *et al*. (1998). Prevention of bone loss with Alendronate in postmenopausal women under 60 years of age. *N Engl J Med* **338:**485–92.
45. Seeman, E. and Delmas, P.D. (2001). Reconstructing the skeleton with intermittent parathyroid hormone. *Trends Endocrinol Metab* **12:**281–3.
46. Dempster, D.W., Cosman, F., Kurland, E.S. *et al*. (2001). Effects of daily treatment with parathyroid hormone on bone microarchitecture and turnover in patients with osteoporosis a paired biopsy study. *J Bone Miner Res* **16:**1846–53.
47. Zhou, H., Shen, V., Dempster, D.W. *et al*. (2001). Continuous parathyroid hormone and estrogen administration increases vertebral cancellous bone volume and cortical width in the estrogen-deficient rat. *J Bone Miner Res* **16:**1300–7.
48. Neer, R.M., Arnaud, C.D., Zanchetta, J.R. *et al*. (2001). Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* **344:**1434–41.
49. Cosman, F., Nieves, J., Woelfert, L. *et al*. (2001). Parathyroid hormone added to established hormone therapy: effects on vertebral fracture and maintenance of bone mass after parathyroid hormone withdrawal. *J Bone Miner Res* **16:**925–31.
50. Rittmaster, R.S., Bolognese, M., Ettinger, M.P. *et al*. (2000). Enhancement of bone mass in osteoporotic women with parathyroid hormone followed by alendronate. *J Clin Endocrinol Metab* **85:**2129–34.
51. Samnegård, E., Akhter, M.P. and Recker, R.R. (2001). Maintenance of vertebral body bone mass and strength created by human parathyroid hormone treatment in ovariectomized rats. *Bone* **28:**414–22.
52. Samnegård, E., Iwaniec, U.T., Cullen, D.M. *et al*. (2001). Maintenance of cortical bone in human parathyroid hormone(1–84)-treated ovariectomized rats. *Bone* **28:**251–60.
53. Iwaniec, U.T., Samnegård, E., Cullen, D.M. *et al*. (2001). Maintenance of cancellous bone in ovariectomized, human parathyroid hormone [hPTH(1–84)]-treated rats by estrogen, risedronate, or reduced hPTH, *Bone* **29:**352–60.
54. Meunier, P.J., Roux, C., Seeman, E. *et al*. (2004). The effect of strontium ranelate on the risk of vertebral fracture in women with postmenopausal osteoporosis. *N Engl J Med* **350:**459–68.
55. Tonino, R.P., Meunier, P.J., Emkey, R. *et al*. (2000). Skeletal benefits of alendronate; 7-year treatment of postmenopausal osteoporotic women. *J Clin Endocrinol Metab* **85:**3109–15.
56. Bone, H.G., Hosking, D., Devogelaer, J-P. *et al*. (2004). Ten years experience with alendronate for osteoporosis in postmenopausal women. *N Engl J Med* **350:**1189–99.

57. McClung, M.R., Wasnich, R.D., Hosking, D.J. *et al*. (2004). Prevention of postmenopausal bone loss: six year results from the Early Postmenopausal Intervention Cohort study. *J Clin Endocrinol Metab* **89:**4879–85.
58. Wasnich, R.D., Bagger, Y.Z., Hosking, D.J. *et al*. (2005). Changes in bone density and turnover after alendronate or estrogen withdrawal. *Menopause* **11:**622–30.
59. Black, D.M., Bilezikian, J.P., Ensrud, K.E. *et al*. (2005). One year of alendronate after one year of parathyroid hormone (1–84) for osteoporosis. *N Engl J Med* **353:**555–65.
60. Wasnich, R.D. and Miller, P.D. (2000). Antifracture efficacy of antiresorptive agents are related to changes in bone density. *J Clin Endocrinol Metab* **85:**231–6.
61. Hochberg, M.C., Greenspan, S., Wasnich, R.D. *et al*. (2002). Changes in bone density and turnover explain the reductions in incidence of non-vertebral fractures that occur during treatment with antiresorptive agents. *J Clin Endocrinol Metab* **87:**1586–92.
62. Parfitt, A.M. (2002). Misconceptions (2): turnover is always higher in cancellous than cortical bone. *Bone* **30:**807–9.
63. Cranney, A., Guyatt, G., Griffin, L. *et al*. (2002). Summary of meta-analysis of therapies for postmenopausal osteoporosis. (2002). *Endocr Rev* **23:**570–8.
64. Schnitzer, T., Bone, H.G., Crepaldi, G. *et al*. (2000). Therapeutic equivalence of alendronate 70 mg once-weekly and alendronate 10 mg daily in the treatment of osteoporosis. *Aging Clin Exp Res* **12:**1–12.
65. Brown, J.P., Kendler, D.L., McClung, M.R. *et al*. (2002). The efficacy and tolerability of risedronate once a week for the treatment of postmenopausal osteoporosis. *Calcif Tissue Int* **71:**103–11.
66. Cramer, J.A., Amonkar, M.M., Hebborn, A. *et al*. (2005). Compliance and persistence with bisphosphonate dosing regimens among women with postmenopausal osteoporosis. *Curr Med Res Opin* **21:**1453–60.
67. Delmas, P.D., Recker, R.R., Chestnut, C.H., 3rd, *et al*. (2004). Daily and intermittent oral ibandronate normalize bone turnover and provide significant reduction in vertebral fracture risk: results from the BONE study. *Osteoporos Int* **15:**792–8.
68. Chestnut, C.H., 3rd, Skag, A., Christiansen, C. *et al*. (2004). Effects of oral ibandronate administered daily or intermittently on fracture risk in postmenopausal osteoporosis. *J Bone Miner Res* **19:**1241–9.
69. Miller, P.D., McClung, M.R., Macovei, L. *et al*. (2005). Monthly oral ibandronate therapy in postmenopausal osteoporosis: 1-year results from the MOBILE Study. *J Bone Miner Res* **20:**1315–22.
70. Adami, S., Felsenberg, D., Christiansen, C. *et al*. (2004). Efficacy and safety of ibandronate given by intravenous injection once every 3 months. *Bone* **34:**881–9.
71. Reid, I.R., Brown, J.P., Burckhardt, P. *et al*. (2002). Intravenous zoledronic acid in postmenopausal women with low bone mineral density. *N Engl J Med* **346:**653–61.
72. Thompson, K. and Rogers, M.J. (2004). Statins prevent bisphosphonate-induced γ,δ-T- cell proliferation and activation *in vivo*. *J Bone Miner Res* **19:**278–88.
73. Bekker, P.J., Holloway, D.L., Rasmussen, A.S. *et al*. (2004). A single-dose placebo-controlled study of AMG 162, a fully human monoclonal antibody to RANKL, in postmenopausal women. *J Bone Miner Res* **19:**1059–66.
74. Writing Group for the Women's Health Initiative Investigators (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* **288:** 321–33.
75. Beral, V.; for the Million Women Study Collaborators (2003). Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **362:**419–27.

76. Black, D.M., Greenspan, S.L., Ensrud, K.E. *et al*. (2003). The effects of parathyroid hormone and alendronate alone or in combination in postmenopausal osteoporosis. *N Engl J Med* **349:**1207–15.
77. Orwoll, E., Ettingham, M., Vice, S. *et al*. (2000). Alendronate for the treatment of osteoporosis in men. *N Eng J Med* **343:**604–10.

12

Instrument Measurements in Osteoporosis Clinical Trials: Evaluating the Endpoints

Colin G. Miller

12.1. Introduction

Medical instruments can be used in one of four primary ways: for screening, diagnosis, prognosis, and monitoring the natural history of the disease or therapeutic intervention. Good quantitative endpoints in clinical trials are usually obtained from instruments measuring a physiological parameter that is relevant to the anticipated effect of the molecular entity under investigation, and nowhere is this more apparent than in the field of osteoporosis. The surrogate endpoint of choice, bone mineral density (BMD), is in fact a recognised diagnostic endpoint in its own right in that the World Health Organization (WHO) criterion for defining osteoporosis in an individual is a BMD that is >2.5 standard deviations (SD) below peak bone mass. However, in the arena of clinical trials the choice of endpoint is not as simple as this (see Section 2.3).

It is becoming standard practice in some therapeutic arenas to contract out the measurement or imaging part of the trial to an experienced organization. In the field of osteoporosis, it is now a recognised standard that a dual-energy X-ray absorptiometry (DXA) Quality Assurance (QA) centre is used and it is a requirement for FDA registration of an NME using a BMD outcome. There are several measurements available to the trialist in establishing a clinical programme to develop a new drug for osteoporosis. These include the use of DXA, Quantitative Computed Tomography (QCT), various ultrasonometry techniques, and vertebral morphometry by several methodologies. However, by whatever means the measurements are obtained, eight basic criteria have to be weighed and balanced before any instrument measurements are taken in a clinical trial. The instruments must include the following:

1. Able to discriminate between normal and disease states
2. Acceptably precise and accurate
3. Reliable

4. Relevant
5. Acceptable to regulatory agencies
6. Of acceptable cost to the trialist
7. Acceptable to the subject
8. Safe for the subject and operator.

For many instruments that are used in trials, these parameters are not evaluated. Some measurements have become so commonplace and acceptable to the regulatory agencies, that they are immediately accepted. An example would be blood pressure measurement by sphygmomanometer cuff in hypertensive subjects to evaluate a beta-blocker or calcium-channel blocker.

However, a closer review of these criteria can ensure that trials are conducted more efficiently, even with known instruments. A case in point arose a few years ago with the random zero sphygmomanometer that caused error in several clinical trials.[1] Care on the part of the trialists and contracting out to specialists in the field could have prevented the problem. However, the precedent has been set for proper quality control (QC) and quality assurance (QA) for instruments in clinical trials, which in some instances is now the expected approach by the regulatory agencies.[2]

Discussion of these eight criteria will help the trialist to understand the factors affecting clinical measurements. It should also reduce the potential error in instruments in clinical trials.

12.2. Ability to Discriminate Between Normal and Disease States

The greater the numerical spread, the "better" the instrument is at differentiating between comparators or between active comparator and placebo. Because the young normal range could be anatomical-site-dependent and method-dependent, this requires the definition of "normal" and "abnormal". For the fully quantitative methodologies under discussion here (DXA, ultrasonometry, and QCT), the population should be defined by a population of young healthy individuals aged between 20 and 40 years (or, arguably, 20 and 35 years), without any history of bone disease or medication usage likely to affect bone. The normal population should also be drawn from a geographically dispersed population, to avoid local regional differences. Owing to phenotypic variation, a separate population must be assessed for each of the major ethnic groups (for example Caucasian, African, and Asian). There is some argument that this must be considered within subpopulations, for example north European and south European, but on the whole, for clinical trial use, this is pushing the argument too far and is not relevant, because we are mainly interested in monitoring not diagnosing.

The definition of osteoporosis as a bone mass >2.5 SD below the young normal mean has come about because of the inverse relationship between DXA and the age-related increase in fracture risk. Currently, the role of ultrasonometry in this paradigm is uncertain. There are prospective data available on three ultrasonometers (Lunar (GE Healthcare, Madison WI, USA), Sahara (Hologic Inc, Bedford MA,

USA), and Cuba (Cooper Surgical, Trumbull, CT, USA))[3–5] for their ability to diagnose risk of fracture. However, there is less information available on other systems. Low bone mass would not be a clinical issue *per se*, if it were not an indicator for increased risk of fracture. In the field of osteoporosis, fracture must be the ultimate endpoint by which all the instruments are evaluated.

This leads on to another debate: what is a vertebral fracture? The ability to diagnose a hip fracture is, generally, very easy with a good radiograph of the hip. However, what constitutes a vertebral fracture and which definition should be used in a particular trial requires some evaluation and determination *a priori*. There are several different methodologies available, both fully quantitative and semiquantitative. The methodology for placing markers for the quantitative assessment of vertebrae is shown in Figure 7.16. The major issue is defining when a vertebra has deformed sufficiently to be classed as a fracture. For example, does a reduction in anterior height of 15% define a fracture, or should it be 20%? This can be the difference of 1 mm or less. If there has been a deformity, how do we know for certain, without seeing a radiograph of the vertebra before the suspected injury? The latter, of course, is nearly always impossible to obtain, so a comparison must be made either with a defined population or with other vertebrae within the subject's spine. Which vertebrae should then be compared and are they deformed? This is obviously important on two fronts:

1. The evaluation of the normal ranges for the new instruments. Both the major DXA manufacturers that provide morphometric X-ray analysis (MXA) use a reference range derived using radiographic methods of morphometry. The Hologic software High Definition Instant Vertebral Assessment (IVA-HD) uses the McCloskey[6] and Minne[7] methods, using the fourth lumber vertebra (L4) as the reference vertebra owing to poor visualization of the thoracic vertebrae. The Lunar Dual-Energy Vertebral Assessment (DVA) uses a method similar to the Minne method, using the second to fourth lumber vertebrae (L2 to L4) as reference vertebrae. Owing to the magnification inherent in the radiographic images, the vertebral heights from MXA are 20–30% lower than from the radiographic technique.[8] MXA reference ranges have been derived, and show a greater sensitivity for detecting vertebral fracture than using a radiographic reference range.

2. Defining the pure endpoint for clinical trials. This latter issue has been debated several times openly and within closed doors, to provide consensus for a major clinical programme.[9–11] The general consensus has been to enrol subjects with a vertebral deformity of $>20\%$ compared with the vertebrae above and below the one in question or compared with the fourth cervical vertebra (C4), as proposed by the Minne index. All future incident fractures are normally considered if the deformity deteriorates by 20% or more, or occurs at a new vertebra. However, this is not the only methodology that has been employed. For the semiquantitative methodologies, the reader would be better advised to consult the original references.[12,13]

12.3. Precision and Accuracy

Precision is the measure of the reproducibility of the measurement. This is usually assessed as the percentage coefficient of variation (CV%; see Section 4.8). The lower the CV%, the better the precision and the easier it is to detect small changes in BMD. This has to be factored into the power calculations, which will determine the study size. If, however, the precision is very much less than the SD of the population mean, precision is not such an issue. For example, the SD of BMD in a group of subjects will be approximately 0.1 g/cm^2, with a mean of 1.0 g/cm^2, that is approximately 10%. The long-term precision of most DXA equipment at the lumbar spine is between 1% and 3%, depending on the study population. In the study design, it is clear that the population SD dominates. A study carried out using equipment with a precision of 3% rather than 1% will require at least 10% more subjects in each study group.

Precision is not to be confused with accuracy, which is how close the measurement is to the actual quantity being measured. An example of the difference between precision and accuracy for target shooting is shown in Figure 12.1. For a clinical trial in which a measurement is the primary inclusion or exclusion parameter, the baseline measurement calls for high accuracy. At enrolment, a comparison is made of the individual and a normal reference population, to assess the degree of disease. Precision then becomes more important for all future measurements to ensure they compare to baseline.

Accuracy within the densitometry field is difficult to elucidate fully. For DXA, all the manufacturers use a different standard for calibration. Therefore, there is no absolute standard. Ultrasonometry poses a further set of challenges because it is still unknown what a unit of change represents with respect to the

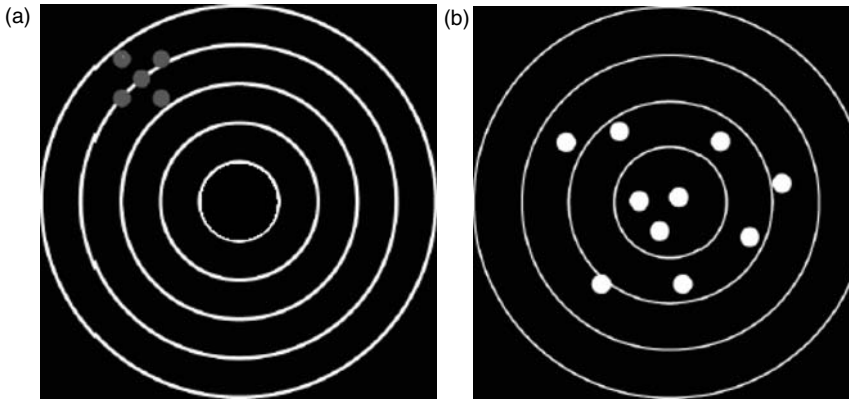


Figure 12.1. The difference between precision (a) and accuracy. (b) Precision is a measure of the consistency result, that is all the bullets hit the target in the same place. Accuracy is the ability to get the right result, that is hit the bull's-eye.

bone density and architecture. For speed or velocity of sound measurements, there is some relationship to density and elasticity. With attenuation measurements such as BUA, there is not only a major component of density in the measurement, but also some part with respect to the architecture of bone. Until these are fully elucidated, the accuracy of ultrasonometry devices will remain elusive. QCT is, arguably, the only methodology available that allows a true determination of accuracy.

If a measurement is not used for selection of enrolment, precision is the overriding parameter to consider. There is an inherent assumption when selecting a particular instrument, or imaging system, that the manufacturer has ensured that it measures accurately. However, calibration checks must be performed and should be considered before starting the trial to ensure these assumptions are correct. With the sphygmomanometer example earlier, the problem would not have arisen if a calibration check had been performed before the start of the study.

Precision and discrimination are two parameters that have to be evaluated together. The poorer the discrimination, the smaller the precision error that is required to distinguish between cohorts or populations. Taking it to the extreme, an instrument with a precision of, for example, 10% would be of little value if the difference between normal and disease states was only 10%, or even 15%. A parameter combining dynamic range and precision was developed in the field of the ultrasonic assessment of bone, the standardized coefficient of variation (SCV%).[14] This is the CV% multiplied by the mean and divided by the dynamic range of the measurement. The smaller the SCV%, the better the instrument. Because the measure of precision is affected by the scale being used, it is impossible to compare instruments using different scales. The SCV% allows for this comparison. Other statistical methodologies have also been proposed[15,16] to overcome these problems.

Table 4.1 shows the comparison of DXA and ultrasonometry in terms of SCV%. DXA of the spine is still the optimum measurement if evaluated in these terms.

12.4. Reliable

Clinical trials in osteoporosis can last many years. There are now trials for which follow-on studies mean subjects have been followed for up to 8 years on the same instrument. If an instrument continually breaks down, it is questionable whether it should be used in a clinical trial. The chance of losing some important data points could become too great. Before buying or leasing a particular instrument, researchers should visit other centres to learn of the problems with equipment, suppliers, and support organizations. The ergonomics and ease of use from subject and staff viewpoints are also important, in addition to the ease with which database information can be extracted for analysis purposes.

A change in instrument calibration can also be considered as affecting the reliability of equipment. Is there a way the calibration and linearity can be evaluated

on an ongoing basis? The technologist operating the instrument should measure a standard or calibration phantom daily. The BMD of the phantom can then be checked chronologically for changes and drifts in calibration. This then must be more fully evaluated by the DXA QA centre, as discussed in Chapter 8.

12.5. Relevant

Is the measurement going to provide useful information about the product under examination? It would not be the first time a measurement would be taken because an investigator or medical advisor has been given incorrect information on the output of the instrument or how the measurement will evaluate the product under examination. In addition, the use of a particular instrument could be appropriate, but the study is statistically underpowered to be able to use the data. This has obvious ethical implications. However, in evaluating new molecular entities (NMEs), changes in BMD, as assessed by DXA, have become the standard. Knowledge of the limitations and assumptions must be known, particularly if evaluating true anabolic bone NMEs, for two reasons. First, BMD measurements with DXA require a good soft-tissue baseline measurement, which can adversely affect the BMD results. Second, DXA measurements of BMD are based on the software identifying the difference between soft tissue and bone. This is performed either by a "simple" thresholding technique (where all pixels in the image over a particular density are assumed to be bone) or by setting the bone edge at the point for which the second-order differential of the attenuation profile across the bone is zero. With an anabolic compound, there is true bone deposition, which could steepen the profile in the area where the thresholding occurs. In these instances, it can cause an apparent increase in the area of the bone. This is an artefact of the DXA software but has the negative secondary effect of reducing the increase in BMD (because BMD = Bone Mineral Content/area).

The choice of measurement site will affect how relevant the measurement is. Does the site itself give adequate discrimination between normal and disease states? In practice, trabecular sites are better than cortical sites, and axial better sites than peripheral sites, for discrimination of antiresorptive compounds. When considering monitoring, both the precision and the expected treatment response must be considered in light of outcome. If the outcome is change in BMD, DXA of the lumbar spine remains the method of choice. It is, however, unreliable in the elderly in whom degenerative changes in the spine could lead to misinterpretation of the change in BMD over a long period.[17] If the outcome is fracture risk, the issue is somewhat more complex. Although a reduction in BMD of 1 SD relates to a doubling of fracture risk, following treatment in subjects with a preexisting fracture, an increase in BMD of a few percentage points can reduce the risk of subsequent vertebral fracture by 50%.[18] Thus, an increase in BMD owing to treatment cannot be directly related to the reduction of fracture risk as it is in age-related bone loss.

12.6. Accepted by Regulatory Agencies

If the study is for registration purposes and the data being collected are essential, confirmation of the acceptability of the measurement is appropriate before the start of the study, not at filing. Many measurements, however, can produce data that are useful supporting documentation, so that their use is appropriate in a well-designed clinical programme. Most study sponsors are seeking to meet the rigorous requirements of the US Food and Drug Administration (FDA) and will use methodologies that are licensed for diagnosis and monitoring by the FDA. DXA and QCT are fully accepted methodologies for most agencies, although QCT is not the methodology of choice because of the high radiation dose to the subject. The FDA will not accept evidence based on QCT alone, although it might be acceptable in a phase II trial and an indication to move to the next stage of development, providing additional DXA data are available in the phase III programme.

Currently, although ultrasonometry is licensed in the USA and the Lunar Achilles has a license for monitoring therapy, the FDA does not accept data of this sort for the evaluation of NMEs. There have been no submissions to date elsewhere, based on ultrasonometry data.

12.7. Acceptable Cost

This is a difficult item to define because it depends on the drug, its stage of development, and the market in which it is to be used. An additional factor is the probable cost of reimbursement when the drug is on the market. Clinical trials are one method by which clinics obtain funding to purchase new equipment, sometimes directly for the particular trial. Using DXA equipment as an example, a few years ago, when the instrumentation was new, only a few sites had the equipment, and pharmaceutical companies wishing to perform clinical trials often had to purchase the instruments for the selected clinical sites. Currently, there are many such instruments (at the time of writing there are about 14,000 DXA instruments around the world, of which an estimated 6,000 can be found in the USA alone). There is no reason for the pharmaceutical company to purchase such equipment, because, with careful selection, investigators can be chosen who have access to such instruments. QCT is much more expensive, although a hospital with an underused Computed Tomography (CT) instrument can buy the necessary software and hardware to enable QCT to be performed. Cost alone should not be the driving force behind this decision, however. Centres chosen to participate in a trial should have a good track record in successfully administering a trial, good QC data, demonstrating that the equipment is reliable, and adequate scientific and technical support. These add overheads to a service that are well worth the investment to provide high-quality data.

Table 12.1. The acceptability of different modalities for monitoring bone density in clinical trials.

	DXA Lumbar spine	Femur	Forearm	QCT Lumbar spine	QUS Calcaneus
Discrimination	√√√	√√√	√√√	√√√	√√
Precise and accurate	√√√	√√	√√	√√√	√√
Reliable	√√√	√√√	√√√	√√√	√
Relevant	√√√	√√√	√	√√√	√√
Acceptable to FDA	√√√	√√√	√	√√√	√
Cost	√√	√√	√√	√	√√√
Acceptable to subject	√√√	√√	√√	√	√√√
Safe	√√	√√	√√	√	√√√

12.8. Acceptable to the Subject

There is only so much inconvenience and measurement that a subject will tolerate. This will vary considerably between subjects, but the way subjects are treated at the investigator site will also have a significant influence on the acceptability of the procedure. During a phase I or II trial, in which a battery of tests is being performed, the investigational team are usually highly involved with the trial and spend a great deal of time with each subject. In these situations, subjects are more likely to tolerate discomfort, particularly when they believe they are being altruistic for humankind. However, this is not the case in the vast majority of phase III or IV studies, or in the routine clinical setting in which the NME is the anticipated treatment of the future. Therefore, it is essential that the measurements are acceptable to the subject, who is normally required to undergo repeat evaluations at each visit. Poor tolerability to the measurement will lead to increased subject drop-out rate and leave the results of the trial questionable. Good investigational staff at the site can make or break a trial in terms of acceptability for the subject. Nowhere is this more apparent than when instrumentation is being used that might appear frightening or overwhelming to the subject. The acceptability of each methodology is summarized in Table 12.1.

12.9. Safe for the Subject and Operator

With any task we perform in life, there are increased risks for injury and harm. Having a measurement taken increases an individual's risk of harm, and could expose the operator to increased risk. Having our height or weight measured is one end of the spectrum, because the risk is no more than continuing our everyday activities. Towards the other end of the spectrum, we could include a complex CT image being performed during which the subject receives a highly significant dose of radiation (see Chapter 3). Alternatively, an invasive procedure such as angiography has increased risks for the subject. Most operators are well

trained and do not expose themselves to undue risk, but it should be considered. In the angiographic example, from the imaging perspective, operators and attending physicians have to wear lead aprons and do receive some additional radiation exposure compared with the normal background dose. The safety issue is one that should be part of the Institutional Review Board or Independent Ethics Committee (IRB/IEC) deliberations before granting the conduct of the trial.

DXA technologists can safely be in the same room as the instrument without hazard. If the operator is 2 m from the scanning arm, there is no problem with radiation safety. Of some concern are the new, peripheral DXA machines. Because these are relatively portable, they are perceived as inherently safe. The scattered radiation dose is quite high, however, and operators should remain at least 2 m away from the machines while an exposure is made. Ultrasound in contrast produces no ionizing radiation and is, therefore, inherently very safe.

The ergonomics of the equipment also must be considered from the operator's viewpoint. Is it difficult to gain access to the subject for positioning purposes? Is there a C-arm that must be rotated by hand? For small technologists, this can be a problem. Are the ergonomics of the workstation acceptable?

12.10. New Technologies

In the last few years there has been a major increase in the development of medical imaging. The aims are generally either to provide more comprehensive information for early go/no go decisions with NME's or to improve the information about the compound's effect in the Phase III studies. Some of the "up and coming" techniques that are available include the following:

1. Hip Structural Analysis by Tom Beck.[19] At the time of writing this is being adopted by Hologic for use in their densitometers.
2. Feature extraction from plane radiographs of the femur and spine by Imaging Therapeutics (Foster City CA, USA).[20,21]
3. A new comprehensive series of parameters for the further evaluation of the spine and hip with QCT developed by Mindways Software Inc (Austin, TX, USA).[22]
4. Finite Element Analysis from QCT data of the spine by Tony Keaveny.[23,24]
5. Active Shape Modeling of both the vertebral body and proximal femur by, respectively, Optasia Medical Ltd (Cheadle, UK) and David Reid's group in Aberdeen.[25]
6. Trabecular structure by MRI developed by Majumdar et al in San Francisco[26] or commercially available by MicroMRI Inc (Philadelphia, PA, USA).[27,28]

While this list is by no means comprehensive, it does provide a good flavour of what is up and coming. However, each technique will require further evaluation and elucidation of its role within the clinical trial arena and how it stacks up against the metrics presented in this chapter.

12.11. Summary

In conclusion, the use of any instrument in a clinical trial is a multifactorial process, which are summarized in Table 12.1. In this table, it can be seen that DXA of the spine or femur remain the methods of choice for assessment of an NME. As technology becomes more complex, however, and imaging modalities provide us with more information, it is possible to become blinded to technology if a simpler device could be used. A case in point is, again, using a DXA instrument; some devices are set up to measure body composition. This is becoming a key parameter in antiobesity trials. From the measurement of the body composition, a subject's weight can be indirectly evaluated. It was with some dismay that the author was confronted with a situation in which this measurement was being used as the primary assessment of weight rather than the good old-fashioned weighing scales. A pair of well-calibrated scales would have given far more precise and accurate data, even if it had been as an additional measurement.

Having a central review of the data is becoming the standard in many areas in which the trialist does not have the knowledge base to judge, or even understand, the numbers being generated and whether they are correct. Using a centralized review of the data also improves precision significantly, ultimately at a cost benefit. As mentioned in the introduction, some form of QA review is now required by the regulatory agencies for DXA instruments in clinical trials.

The use of a well-qualified team to review the data from an instrument is not unlike having blood and sera samples being sent to a central laboratory for analysis. This latter system is now standard practice in the pharmaceutical industry. For developing imaging systems, it is becoming necessary to contract these measurements out to qualified contract organizations. This is now being performed routinely for vertebral morphometry and DXA measurements for BMD.

Another major advantage this offers to the trialist is the centralization of all the image data, enabling good archival and retrieval mechanisms. The data should all be entered into a database and the relevant endpoints downloaded to the sponsoring company in a timely fashion. The net effect is to reduce the time from the last subject measurement to a clean lockable database. The knowledge that the data are of optimal quality and have undergone a "two-person" review similar to all the other trial data is, in its own right, grounds for ensuring that good central QC is performed. The use of all instruments in clinical trials should be carefully evaluated. For routine and well-established techniques, this will be no more than a momentary mental check. However, for more complex equipment, particularly involving imaging of some kind, a full evaluation is a valuable investment of time compared with the ultimate cost of the trial. The use of a good QA contractor should be seriously considered, and the use of academic centres to conduct this kind of work, particularly if they have no previous track record, should also be carefully considered.

For the development of an antiosteoporosis NME, these criteria have been more fully evaluated than for some other therapeutic environments. This puts the trialist at an advantage, which is further supported by well-established and long-running QA laboratories that can provide full support and technical know-how.

References

1. Conroy, R.M., O'Brien, E., O'Malley, K. *et al*. (1993). Measurement error in the Hawksley Random Zero Sphygmomanometer; what damage has been done and what can we learn? *BMJ* **306:**13 19–22.
2. CPMP Notes for Guidance on Postmenopausal Osteoporosis in Women (2001). http://www.emea.eu.int/pdfs/human/ewp/055295en.pdf accessed 13/11/06
3. Bauer, D.C., Gluer, C.C., Cauley, J.A. *et al*. (1997). Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women. A prospective study. Study of Osteoporotic Fractures Research Group. *Arch Intern Med* **157:**629–34.
4. Hans, D., Dargent-Molina, P., Schott, A.M. *et al*. (1996). Ultrasonographic heel measurements to predict hip fracture in elderly women: the EPIDOS prospective study. *Lancet* **348:**511–14.
5. Pluijm, S.M.F., Graafmans, W.C., Bouter, L.M. *et al*. (1999). Ultrasound measurements for the prediction of osteoporotic fractures in elderly people. *Osteoporos Int* **9:**550–6.
6. McCloskey, E.V., Spector, T.D., Eyres, K.S. *et al*. (1993). The assessment of vertebral deformity: A method for use in population studies and clinical trials. *Osteoporos Int* **3:**138–47.
7. Minne, H.W., Leidig, G., Wuster, C. *et al*. (1998). A newly developed spine deformity index (SDI) to quantitate vertebral crush fractures in patients with osteoporosis. *Bone Miner* **3:**335–49.
8. Cawte, S.A. (1999). Morphometric X-ray absorptiometry using the Hologic QDR 2000, MPhil Thesis. Nottingham: University of Nottingham.
9. Black, D.M., Cummings, S.R., Karpf, D.B. *et al*. (1996). Randomized trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. *Lancet* **348:**1535–41.
10. National Osteoporosis Foundation Working Group on Vertebral Fractures (1995). Report Assessing Vertebral Fractures. *J Bone Miner Res* **10:**518–22.
11. Genant, H.K., Jergas, M. and van Kuijk, C. (1995). *Vertebral Fractures in Osteoporosis*. San Fransisco: Radiology Research Education Foundation.
12. Genant, H.K., Wu, C.Y., van Kuijk, C. *et al*. (1993). Vertebral fracture assessment using a semiquantitative technique. *J Bone Miner Res* **8:**1137–48.
13. Jergas, M., Genant, H.K. and Nevitt, M.C. (1995). Standardized visual assessment of osteoporotic vertebral deformities. In *Vertebral Fractures in Osteoporosis* (Genant, H.K., Jergas, M. and van Kuijk, C., eds), pp. 149–61. San Francisco: Radiology Research Education Foundation.
14. Miller, C.G., Herd, R.J.M., Ramalingham, I. *et al*. (1993). Ultrasonic velocity measurements through the calcaneus; which velocity should be measured? *Osteoporos Int* **3:**31–5.
15. Morris, M., Peretz, A., Tjeka, R. *et al*. (1995). Quantitative ultrasound bone measurements: Normal values and comparison with bone mineral density by dual X-ray absorptiometry. *Calcif Tissue Int* **57:**6–10.
16. Langton, C.M. (1997). ZSD: A universal parameter for ultrasound precision. *Physiol Meas* **18:**67–72.
17. Liu, G., Peacock, M., Eilam, O. *et al*. (1997). Effect of osteoarthritis in the lumbar spine and hip on bone mineral density and diagnosis of osteoporosis in elderly men and women. *Osteoporos Int* **7:**564–9.

18. Cummings, S.R., Black, D.M., Thompson, D.E. *et al*. (1998). Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. *JAMA* **280:**2077–82.
19. Beck, T.J., Ruff, C.B., Warden, K.E. *et al*. (1990). Predicting femoral neck strength from bone mineral data. A structural approach. *Invest Radiol* **25:**6–18.
20. Steines, D. *et al*. (2005). Predicting hip fracture in a sub-group of subjects from the study of osteoporotic fractures using automated structural measurements of proximal femur in pelvic radiographs. Presented as a plenary poster at the 27th annual meeting of the ASBMR, Nashville, Tennessee. Available at http://www.imatx.com/ImaTx_ASBMR_2005_Hip_Fracture_Prediction.PDF (Accessed 11/07/07).
21. Arnaud, C. *et al*. (2005). Anabolic changes in vertebral trabecular microstructure in osteoporotic postmenopausal women treated for 2 years with human Parathyroid hormone fragment (hPTH 1-34) as determined by a new, automated noninvasive imaging technology that uses ordinary lateral spine radiographs. Presented at the 27th annual meeting of the ASBMR, Nashville, Tennessee. Available at http://www.imatx.com/ImaTx-Spine-Imaging.PDF (Accessed 11/07/07).
22. Brown, J.K., Qiao, Q., Weigert, J.M. *et al.* (2005). Improved precision of hip structure analysis using optimized projection images from segmented 3D CT scans of the hip. *J Bone Miner Res* **20(Suppl 1):** S337.
23. Niebur, G.L., Feldstein, M.J., Yuen, J.C. *et al*. (2000). High-resolution finite element models with tissue strength asymmetry accurately predict failure of trabecular bone. *Journal of Biomechanics* **33:**1575–83.
24. Liebschner, M.A.K., Kopperdahl, D.L., Rosenberg, W.S. *et al*. (2003). Finite element modeling of the human thoracolumbar spine. *Spine* **28:**559–65.
25. Gregory, J.S., Testi, D., Stewart, A. *et al*. (2004). A method for assessment of the shape of the proximal femur and its relationship to osteoporotic hip fracture. *Osteoporos Int* **15:**5–11.
26. Carballido-Gamio, J., Majumdar, S. (2006). Clinical utility of microarchitecture measurements of trabecular bone. *Curr Osteoporos Rep* **4:**64–70. (Review).
27. Gomberg, B.R., Saha, P.K., Wehrli, F.W. (2005). Method for cortical bone structural analysis from magnetic resonance images. *Acad Radiol* **12:**1320–32.
28. Gomberg, B.R., Saha, P.K., Song, H.K. *et al*. (2000). Topological analysis of trabecular bone MR images. *IEEE Trans Med Imaging* **19:**166–74.

13

Biochemical Markers of Bone Turnover

Penny Blackwell, Ian M. Godber and Nigel Lawson

13.1. Introduction

The routine clinical assessment of bone mineral density (BMD) is best undertaken by the use of dual-energy X-ray absorptiometry (DXA) and ultrasound scans. These techniques will establish BMD at a particular time. Using serial DXA measurements it is possible to measure a change in BMD over a set period of time. It is presumed that these measured changes in BMD are caused by alterations in bone turnover, but they are not direct measurements of bone turnover. Furthermore, DXA scans and ultrasound can only indicate that loss of BMD has occurred, a single measurement cannot indicate that bone loss *is* occurring, and might lead to a lowered BMD in the future. These radiological and ultrasound techniques have other limitations, not least inherent imprecision of the methods, which means that there have to be considerable changes in bone turnover before changes in BMD can be noticed. For example, the 1–2% imprecision of DXA measurement limits scanning to 6-monthly intervals, so that observed changes are certain to be owing to bone loss and not imprecision. The skilled nature of these techniques combined with the need for special equipment, and in the case of DXA exposure to ionizing radiation, has fueled the search for useful and reliable markers of bone turnover.

Biochemical markers of bone turnover are based on the measurement either of peptides, enzymes, and other small molecules synthesised by osteoclasts and osteoblasts or of osteoclast-generated degradation products of bone matrix. These factors are usually measured in the urine or blood. Biochemical markers provide information about whole-body bone turnover and are minimally invasive. Obviously, these markers can be more frequently assessed than BMD, and the two assessments complement each other, because they can assess two different parameters—bone *turnover* and bone *density*. Ideally, biochemical markers of bone turnover should correlate to *changes* in BMD and have a use in predicting

Table 13.1. Bone markers.

Formation markers (usually only measured in serum)	Resorption markers (usually measured in urine, but some also measured in serum)
Alkaline phosphatase (ALP) Total (TALP) Bone-specific (BALP)	Hydroxyproline (OHP) Galactosyl hydroxylysine (GHyL)
Osteocalcin (OC)	Collagen crosslinks Actual crosslinks (free and total) deoxypyridinoline (DPyd) pyridinoline (Pyd)
Propeptides of type I collagen N-terminal (PINP) C-terminal (PICP)	Peptide bound crosslinks: N-terminal (NTx) C-terminal telopeptide (CTx, ICTP) Tartrate-resistant acid phosphatase (TRACP)

which subject groups are at risk of developing low BMD and hence increased risk of fracture.

Bone markers fall into two main categories: markers of resorption and formation (Table 13.1).

13.2. Markers of Resorption

13.2.1. Hydroxyproline (OHP)

OHP constitutes 13% of the amino acids in collagen, which are released into the circulation on degradation of the bone matrix.[1] However, OHP is present in the N-terminal extension peptide cleaved from procollagen, hence its presence in urine might not be wholly indicative of resorption, but of formation too. Circulatory OHP is not reincorporated into new bone but reabsorbed by the renal tubules. The liver degrades 80% and 10% enters the urine as small peptides or larger ones derived from the N-terminal propeptide of type I collagen (PINP).

OHP lacks specificity for bone collagen. Dietary collagen in urine is indistinguishable from bone collagen, so all urine samples must be taken after an overnight fast and abstinence from foods containing high concentrations of collagen,[1] such as products containing gelatin. OHP is also found in the C1q component of complement.[2] Other confounding factors contributing to the variability of OHP levels include connective tissue disorders, other OHP-containing proteins,[3] and a diurnal rhythm, with peak excretion just after midnight. High levels of OHP are seen in infants, falling by 5 years of age, and declining again at puberty to adult levels.

The usefulness of OHP is blunted by contribution from the diet, lack of specificity for type I collagen, and lengthy chemical assays compared with other bone markers. OHP assays are not considered sensitive enough for individual diagnostic and therapeutic monitoring, and assay performance is inconsistent, compared with some of the newer markers. However, OHP can provide useful information if bone turnover is markedly elevated, such as in Paget's disease.

Because 90% of OHP is peptide-bound, samples require hydrolysis before assaying, to remove the measurable OHP from peptides. The first analyses of OHP were colorimetric reactions, based on the oxidation of OHP to pyrrole-2-carboxylic acid. This compound is heated to form a pyrrole, which is extracted with an organic solvent, such as toluene, and reacted with *p*-dimethylaminobenzaldehyde to produce a chromophore, which can be measured spectrophotometrically. However, many chromophores interfere with this method and a modified colorimetric assay was developed that uses a resin to remove interfering substances and hydrolyse peptides.

Since these early colorimetric methods, high-performance liquid chromatography (HPLC) techniques were developed, which improved the reproducibility and precision of OHP measurement.[4] OHP can be derivatized with compounds that enable detection, such as phenylisothiocyanate or NBD-chloride, the products of which can be detected by ultraviolet or fluorescence spectrophotometry, respectively.

Reference ranges for urinary OHP vary between techniques, but employing the relatively accurate and precise HPLC techniques, early morning fasting values in normal adults range from 12 μmol/mmol to 25 μmol/mmol of creatinine. Although OHP can demonstrate gross changes in bone turnover, for example in subjects with Paget's disease who can develop urinary OHP levels up to fivefold higher than the reference range, subtle changes in bone turnover cannot be detected using OHP.

13.2.2. Galactosyl Hydroxylysine (GHyL)

GHyL is present in bone collagen in the pyridinium crosslink and released on osteoclastic resorption. It is less abundant than OHP and not absorbed from the diet. GHyL is more specific for bone collagen, but is also present in complement. Similar to OHP, assays are time-consuming, involving quantitation by HPLC. Hydroxylysine has not been extensively studied or validated against reference techniques, such as calcium kinetics or histomorphometry, and is of questionable use as a resorption marker.[5] However, it has been suggested that it could have a potential use in the prediction of bone metastases in subjects with breast cancer.[6]

13.2.3. Collagen Crosslinks

Osteoclastic degradation of bone yields peptide fragments, some of which contain crosslinks.[1] Fragments range in size from the free crosslink to peptide segments of the telopeptides containing the crosslink. Small fragments are cleared by the kidney and can be detected in urine. Studies *in vitro* have shown that osteoclastic resorption yields only peptide fragments. Two-thirds of these small peptides in urine contain sequences specific for type I collagen, conferring added specificity to measurement of the peptide-linked crosslink as a marker of bone resorption.[7] It has been suggested that because the majority of crosslinks are found at the N-terminal end of collagen, measuring peptides from this end, rather than the C-terminus, should provide a more sensitive index of bone resorption.

Several parameters can be measured (Figure 13.1). The crosslink molecules themselves, pyridinoline (Pyd) and deoxypyridinoline (DPyd), can be measured,

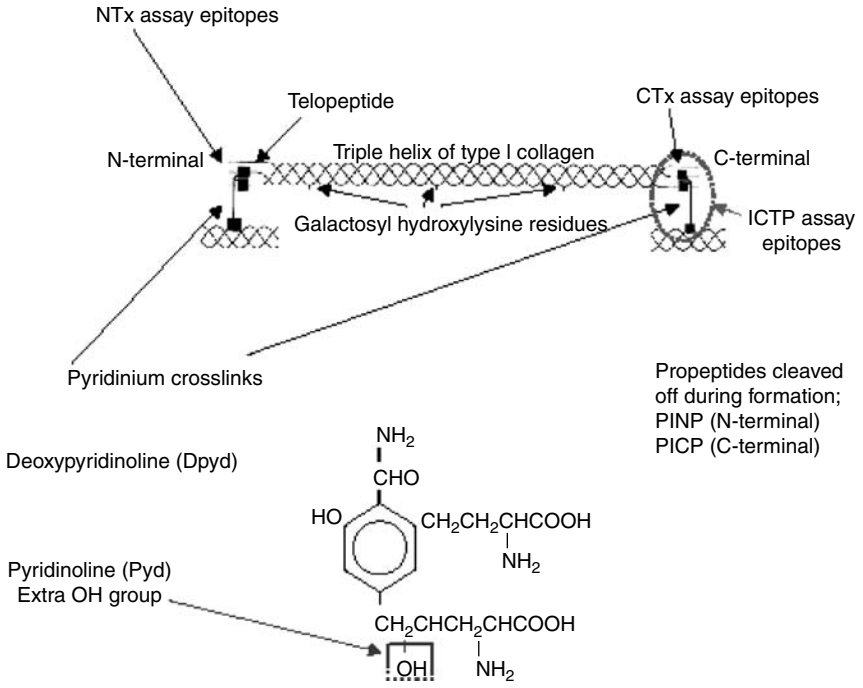


Figure 13.1. The origin of several markers of bone turnover.

either in their free forms or as total crosslinks after hydrolysis. Alternatively, the peptide-linked crosslinks, N-terminal telopeptide (NTx), C-terminal telopeptide (CTx or crosslaps), or a variant of this assay, type I collagen C-terminal peptide (ICTP), can be measured.

Free crosslink measurement detects only the crosslink moiety itself, whereas peptide-bound crosslinks detect the crosslinks remaining bound to small peptides. Peptide-linked crosslinks are thought to be the physiological breakdown product of bone resorption. It is also possible to simultaneously measure total, in addition to free, crosslinks[8] in the serum and urine by hydrolysing samples to strip Pyds from the peptides on which they are released from collagen. Hydrolysis involves boiling samples in 6 mol/L of hydrochloric acid and neutralizing them with a high salt concentration, which could be a potential source of crosslink destruction.[1]

13.2.4. Pyd Crosslinks

Pyd crosslinks stabilize and strengthen collagen fibrils,[9] which are formed by the action of lysyl oxidase on lysine and hydroxylysine residues in the telopeptide domains of the collagen molecule (Figure 13.1). Aldehydes result, which condense with hydroxylysyl or lysyl residues on adjacent collagen molecules to form

mature crosslinks: Pyd (three hydroxylysine molecules) and DPyd (two hydroxylysine molecules plus one lysine molecule). They are released on osteoclastic bone resorption and excreted by the kidney.[10]

Crosslinks are found in all types of collagen and are present in the diet in animal products, but are not absorbed in the gut and so do not interfere with measurement.[11] In tissues other than bone and dentine, Pyd predominates over DPyd. In type I collagen, Pyd predominates, but less so; therefore, DPyd is considered the most bone-specific crosslink. DPyd concentrations in soft tissue collagen can be as high as those in bone, but bone represents the major reservoir of collagen in the body and turnover is faster than in other connective tissues. Also, the ratio of Pyd: DPyd in bone and urine is similar in both media, evidence for the majority of crosslinks in urine having an origin in bone.[7]

Free, or total, crosslinks can be measured by HPLC, radioimmunoassay (RIA; Quest Diagnostics, San Clemente, CA, USA), polyclonal antibody-based Enzyme-linked immunosorbent assay (ELISA; Quidel Corporation, San Diego, CA, USA), and automated immunoassay (ACS:180, Novartis Vaccines & Diagnostics Inc, Emeryville, CA, USA; Immuno 1, Bayer Healthcare Diagnostics, Tarrytown, NY, USA; Immulite, DPC Biermann, Bad Nauheim, Germany).

The first HPLC method used to measure crosslinks was developed in 1984 by Eyre *et al*.[12] and was further modified by Kollerup *et al*.[13] Isocratic, ion-paired, reverse-phase HPLC was employed, using the natural fluorescence of the crosslinks for detection of Pyd and DPyd in urine. This method remained in general use for some time, using different solvents and detection systems.

Although HPLC is a very accurate and precise technique, it does require a high degree of technical skill, thus encouraging the development of commercial immunoassay methods. One of the first immunoassays to measure free Pyds used a polyclonal antibody-based ELISA, and more recently, a monoclonal antibody assay was developed that detected the more bone-specific DPyd.[14] The antibody exhibits $<1\%$ crossreactivity with Pyd and the assay time is considerably shorter: 3 hours in total. The DPyd antibody (Quidel Corporation) has now been patented and the last 2 years have seen the introduction of DPyd as a test option on various automated immunoassay analysers, such as the ACS:180 and DPC Immulite. This has revolutionized the measurement of DPyd, removing the manual aspect of the analysis to make it more precise and speeding up the process of analysis.

Total crosslinks have been measured in serum, with the best results obtained for Pyd.[15] To measure the total crosslinks, samples must be hydrolysed to strip the Pyds from the small peptides on which they are released from collagen into the urine. The hydrolysis protocol that is endorsed by the Quidel Corporation involves boiling the samples in 6 mol/L of hydrochloric acid and then neutralizing the acidity with a high concentration of salt. Hydrolysis, and thus, measurement of total crosslinks can only be performed using the Quidel Corporation method or HPLC. It is not possible to use hydrolysed samples in the automated immunoassay systems because the dilution step has an impact on the sensitivity of the assay, in addition to the high salt concentration. As discussed earlier, hydrolysis could also be a potential source of destruction of the crosslinks.

There has been a lack of common reference standards for the measurement of Pyds, which is reflected by the fact that many of the previously published reference ranges for DPyd vary considerably between reports.[16] Almost all of the standards are now biological in origin, having been isolated from children's urine or adult bone.

Healthy children excrete 10 to 15 times more DPyd than young adults; adult DPyd concentrations increase until the ninth decade.[17] Total DPyd exhibits considerable circadian variation, with peak values at 6 a.m. and trough levels at 5.30 p.m.[7]

Urine collection should be standardized. Some workers use a fasting, early-morning urine sample to allow for circadian rhythm, correcting the results for creatinine,[11] and some use 24-hour collections. However, DPyd measurements in these two types of sample do not correlate,[10] and the early-morning samples are recommended, with the result expressed per mmol of creatinine. Samples are stable at −20°C for 18 months[13,18] and are resistant to up to 10 freeze–thaw cycles;[11,19] these data have been extrapolated to suggest that DPyd measurements will be stable for up to 25 years at −80°C.

Reference ranges for DPyd have been established for adults >25 years using the commercial enzyme immunoassay for "Pyrilinks-D" (Quidel Corporation). These ranges are quoted as 3.4–7.4 nmol/mmol of creatinine for females and 2.3–5.4 nmol/mmol of creatinine for males. The age of the subject should also be taken into account, with the excretion of DPyd rising throughout childhood, peaking at 12–14 years of age.[20] Recent studies have also proposed that diabetic adolescents excrete higher concentrations of DPyd than normal adolescents.[20]

Measurement of free, or total, crosslinks remains controversial. Several clinical studies have shown that changes in the level of free DPyd do not reflect change in total DPyd excretion in subjects on bisphosphonates or hormone-replacement therapy (HRT).[21,22] Robins[11] suggests that bisphosphonates could affect a pathway in the degradative metabolism of bone collagen, altering the proportions of small peptide-linked and free crosslinks. Renal handling of crosslinks might also be altered in renal impairment and crosslink fragments could be produced, some of which might remain immunoreactive.

13.2.5. Peptide-Linked Crosslinks

Assays for peptide-linked crosslinks include NTx, crosslaps or CTx, and ICTP. The first two analytes are usually measured in urine, whereas the ICTP assay is performed using serum. NTx is a type I collagen-specific peptide sequence in the α_2(I) molecule,[23] containing the crosslink.[1]

CTx is a type I collagen-specific sequence at the C-terminal end of the molecule,[24,25] containing some of the helical part of the α1 and α2 collagen chains, which are strongly conserved in collagen types II and III, which could crossreact. Although traditionally measured in urine, kits have recently come onto the market for the measurement of CTx in serum. It has been suggested that serum CTx assays are more specific to bone resorption than other measurements and results

show a good correlation to changes in spinal BMD in response to antiresorptive therapy with pamidronate.[26]

The antibody raised to ICTP has not been fully characterized,[27] but recognises a 12 kDa antigen in the serum. Correlation to urinary resorption markers is poor, but serum measurement confers advantages compared with ICTP for use in bone disease with renal impairment, although crosslink excretion is not significantly altered by renal function.

NTx is often measured by a monoclonal antibody-based ELISA method, in which the antibody is raised to a type I collagen-specific peptide sequence in the α_2(I) molecule.[24] The sequence (gln-tyr-asp-gly-lys-gly-val-gly) is a product of osteoclast degradation, in which the lysine molecule is involved in the crosslink itself.[28] Standards for the analysis of NTx are based on collagenase-digested human bone collagen. Because the antibody is raised to the specific amino acid sequence, the assay does not recognise the free crosslink or telopeptides alone.

CTx is usually measured by a polyclonal antibody-based method, the antibody being raised to a type I collagen-specific sequence at the C-terminal end of the molecule (glu-lys-ala-his-asp-gly-gly-arg).[25] This was further developed, and then evaluated, by Bonde *et al*.[25] However, this antigen also consists of some of the helical part of the α1 and α2 collagen chains, both of which are strongly conserved in different types of collagen. Thus, there is a possibility of crossreactivity with collagen types II and III. The newer serum CTx assays use two monoclonal antibodies, each recognising characteristic linear octapeptides.

ICTP is measured by a monoclonal immunoassay.[25]

NTx levels are highest in infancy and childhood,[24] peaking at puberty, and falling to a plateau in adulthood, with a rise after the menopause.[29] NTx levels exhibits a diurnal rhythm, similar in peaks and troughs to DPyd levels. In comparative studies, NTx shows a greater increase at the menopause than total, or free, Pyds and a greater suppression on treatment with bisphosphonates. Studies investigating CTx have not been so numerous, but show a similar increase at the menopause and decrease with bisphosphonate therapy, suggesting that NTx is more bone-specific.[22] Similar to the other crosslinks, ICTP exhibits a circadian rhythm, with higher levels at 4.00 a.m. than midafternoon. Serum ICTP increases 20% after the menopause and decreases by only 10% after 1 year's HRT, possibly making it insensitive to slower rates of bone turnover.[30] However, ICTP has been shown to be useful in assessing resorption in subjects with bone secondaries.

Reference ranges for NTx are usually derived from early-morning urine samples and are expressed as nmol of bone collagen equivalents (BCE) per mmol of creatinine. Quoted values for adults are in the range of 4–92 nmol BCE/mmol of creatinine. NTx has been measured in serum, with mean values of 16.0 nmol/L quoted. Mean CTx concentrations in the urine range from 227 μg/mmol of creatinine in premenopausal women to 429 μg/mmol of creatinine in postmenopausal women.[31] In children, mean CTx values can be as high as 1849 μg/mmol of creatinine.[32] The upper reference value for ICTP is 4.6 μg/L of plasma.

13.2.6. Tartrate-Resistant Acid Phosphatase (TRACP)

There are five isoenzymes of acid phosphatase (ACP), which are found in the bone, prostate, platelets, red blood cells, and spleen, respectively. The bone-specific enzyme is produced by osteoclasts and excreted into the sealed off resorbing compartment. Bone-specific ACP leaks into the circulation during bone resorption and after the resorbing compartment is released.[33] TRACP is measured in serum.

Isoenzymes can be separated and analysed by kinetic methods, according to their susceptibility to inhibition by tartrate. This is not very specific, but separates ACP activity owing to bone from other ACPs, such as erythrocyte and prostatic ACPs. Charge differences enable isoenzyme separation by electrophoresis, and recently, immunoassays have been developed[34] using monoclonal antibodies to measure bone-specific ACP more specifically. Only small studies, to date, have used TRACP to assess bone turnover, so measurement and validation of this parameter should be considered incomplete.[35] Reference values seem to be highly dependent on the assay employed, with quoted adult mean values varying from 13 g/L[35] to 197 g/L.[34] A recent study has suggested that the TRACP circulates in serum as part of a calcium-containing complex, and that the TRACP must be removed from this complex before analysis.[35]

13.3. Markers of Formation

13.3.1. Total Alkaline Phosphatase (TALP) and Bone-Specific ALP

Elevation of TALP in skeletal disorders has been recognised for >60 years. ALP activity is widespread, encoded for by genes at four different loci: the tissue non-specific (tns) gene (bone, liver, kidney, and early placental forms) on chromosome 1 and intestinal, mature placental and germ-cell ALP on chromosome 2.[36] The tns gene produces proteins with 50% homology[37] and tissue-specific differences between these isoforms are caused by carbohydrate side-chain variations and the degree to which they are sialated.

Many theories for the role of ALP in bone formation exist, based on its ability to increase local concentrations and transport of phosphate and to destroy inhibitors of mineral crystal growth. A role in bone mineralization is supported in hypophosphatasia, an autosomal recessively inherited deficiency of tns ALP that causes defective bone and teeth mineralization.[38] In human osteoblast-like cells, ALP activity is proportional to phosphate concentration, and release of ALP from its phospholipid anchor is inversely proportional to calcium concentration, suggesting a role in initiating mineralization.[39]

ALP in healthy adults is derived equally from both liver and bone. The enzyme is anchored to phosphatidyl inositol moieties on the extracellular cell surface, but can be converted to a soluble, circulating form by phospholipases. These fractions of membrane-attached ALP increase in hepatobiliary disease, because of the action of detergent-like bile acids.[40] Altered glycosylation patterns can also occur,

some of these forms being detected when bone-specific ALP is measured.[41] Bone-specific ALP can also arise from excess or "used" enzyme, and so the relationship between bone-specific ALP and the bone-formation rate might not be simple or constant within an individual.

Interest in separating the two major isoforms has exploited differences between liver-specific and bone-specific ALPs. Measurement of bone-specific ALP improves sensitivity and specificity for bone formation, but separation from the liver isoform is technically difficult owing to structural similarities. The problems are further compounded by heterogeneity of the isoforms in different disease states, which could react unexpectedly during analysis.[41] The criteria that have been exploited to separate the isoforms include differences in heat stability, urea sensitivity, electrophoretic mobility, carbohydrate moieties, and immunochemical characteristics.[42]

13.3.1.1. Heat Stability

At 56°C the placental ALP isoform is completely stable, the liver enzyme shows intermediate stability [half-life ($t_{1/2}$), 7.6 ± 1.5 min], and the bone-specific ALP is very labile ($t_{1/2}$, 1.9 ± 0.4 min). ALP is almost completely destroyed by heating to this temperature for 10 minutes. TALP activity is measured in heated and unheated serum. If the heated ALP activity is <20% of the unheated activity, the sample consists primarily of bone-specific ALP. Thus, the method will give only a qualitative, or at best, a semiquantitative estimate for bone-specific ALP.[41]

13.3.1.2. Electrophoresis

This is still considered the "gold standard" method for separating bone-specific and liver-specific ALP enzymes. When serum containing ALP isoenzymes is separated electrophoretically, the liver-specific ALP, which carries the highest negative charge, moves most rapidly towards the anode, followed (in order) by the placental, bone, intestinal, and, rarely seen, kidney isoforms. The bone-specific ALP forms a somewhat diffuse zone, overlapping to some extent with the liver-specific ALP band. Improvements in electrophoretic separation can be made by treating the sample with wheatgerm lectin (WGL)[43] or neuraminidase;[44] the latter preferentially strips sialic acid residues from bone-specific ALP, retarding its mobility in relation to liver-specific ALP by decreasing the negative charge on the molecule. However, neuraminidase will also strip sialic acid residues from liver-specific ALP with time, until both isoforms reach the same isoelectric point (pI) and thus identical electrophoretic mobility. The timing of neuraminidase incubation is thus crucial. Treatment of the electrophoresis gel with WGL can also enhance separation, because WGL binds preferentially to bone-specific ALP, again retarding its mobility. This was first achieved by using cellulose acetate gels[43] and then using agarose gels.[45] Although the WGL gels have better separation, the precision of this method seems variable.[43,45,46]

With many electrophoretic techniques, there is some overlap between the liver and bone isoforms, even in subjects with Paget's disease. Quantitation is thus

imprecise and the method is time-consuming, allowing only 15 samples to be assayed at once.

13.3.1.3. WGL Precipitation

WGL precipitation was first used in 1984 by Rosalki and Ying Foo[43] to separate bone-specific and liver-specific ALPs, and the method has been further optimized.[47,48] WGL binds to *N*-acetylglucosamine residues on glycoproteins, binding preferentially, but not exclusively, to the bone isoform. This method indirectly measures bone-specific ALP, that is TALP activity in the serum is determined, the bone-specific ALP is precipitated out of solution with WGL, and the remaining TALP activity in the supernatant is determined.

WGL displays considerable between-batch heterogeneity, altering the affinity of the lectin for bone-specific ALP. For this reason and because at high concentrations of WGL liver-specific ALP can also be bound, WGL batches require standardization before they can be used. Standardization and assay performance is operator-dependent, as determined by the measurements of precision obtained by different workers, and also depends on what material has been used to standardize the lectin. For example, cord blood, human bone, animal blood, and serum from subjects with Paget's disease have all been used. The type of sample that is used to standardize a batch of WGL does have implications for the accuracy of the method. This was noted by Farley *et al*.[49] who found that if serum from a Paget's subject with a high TALP was used to standardize a batch of lectin, skeletal and hepatic ALPs were very poorly resolved. This might be because it is not known how much the liver isoform is contributing to TALP in these subjects. Using cord blood as the standard, if liver-specific ALP is absent, the results from the WGL method and heat inactivation methods agree.[50]

In 1993, the original Rosalki and Ying Foo method was manufactured and sold in kit form as Iso-ALP® (Roche Diagnostics GmbH, Mannheim, Germany); the kit was evaluated in three countries by five different laboratories.[51] The kit contains already standardized lectin in solution and also incorporates a quality control (QC) sample.

13.3.1.4. Immunoassay

Monoclonal antibodies that distinguish bone-specific ALP from liver-specific ALP were initially developed by Lawson *et al*.[52] and Hill and Wolfert,[53] leading to an immunoassay which showed greater affinity for the bone isoform. Two immunoassays were marketed: an ELISA from the Quidel Corporation (Alkphase B®) and a two-site immunoradiometric assay (IRMA) called Hybritech Ostase® (Beckman Coulter, Fullerton CA, USA). These immunoassays cannot be directly compared with the other methods for quantification of bone-specific ALP, because antibody methods will measure mass, rather than activity of the enzyme.

Garnero and Delmas[54] and Pangrahi *et al*.[55] reported that the IRMA assay had a 16% crossreactivity with liver-specific ALP. There is a good correlation between

this assay and gel electrophoresis, and also to a WGL precipitation method if standardized with cord blood. England *et al*.[56] demonstrated similar crossreactivity with the antibody in the Ostase® kit. Van Hoof *et al*.[57] further validated the IRMA, comparing it with agarose gel electrophoresis in 293 subjects. Their work shows the IRMA is suitable as a screening method, but when high values for bone-specific ALP are found with the Ostase® method, electrophoresis should be used to rule out the possibility of crossreactivity with liver-specific ALP. The ELISA method, using monoclonal antibodies and marketed by the Quidel Corporation, showed a 3–10% crossreactivity with the liver isoform for this method with intra-batch and interbatch Coefficients of Variation (CV) $< 10\%$. There was also a high correlation between this method and WGL precipitation($r = 0.99$).

Because the immunoassay methods are not totally specific for bone-specific ALP, it is advisable to perform liver function tests on samples, in which a raised γ-glutamyl tranferase level might cast doubt on the validity of the bone-specific ALP result.

Physiologically high levels of bone-specific ALP are seen in infancy, falling through childhood, increasing prepubertally, and falling again to adult levels.[58] In healthy adults, liver and bone isoforms constitute ~50% each of the TALP. An increase of two to three times normal might be observed in the third trimester of pregnancy, with an increase in bone-specific ALP, in addition to the placental isoform.[59] The level of bone-specific ALP increases with age and the menopause in healthy populations.[47,60,61]

13.3.2. Osteocalcin

Osteocalcin (bone Gla protein) is the most abundant, noncollagenous protein within the bone matrix, synthesised by mature osteoblasts. It is chemotactic and might aid osteoclast recruitment and activation at resorption sites. Osteocalcin is a 49-amino acid protein, containing three γ-carboxylated glutamic acid residues that bind calcium ions to stabilize the α-helical structure of the molecule and bind osteocalcin to hydroxyapatite.[62] Carboxylation is vitamin K-dependent and research suggests that vitamin K enhances mineralization by modulating osteoblast activity.[63] Anticoagulants that antagonize vitamin K decrease osteocalcin carboxylation, leading to a reduced ability to bind to hydroxyapatite, and increased circulating "undercarboxylated" osteocalcin.[55] Levels of undercarboxylated osteocalcin increase with age,[64] with less incorporated into bone and more entering the circulation. Low vitamin K levels have been found in osteoporotic subjects with femoral neck fractures.[65] Osteocalcin is also modified by 1, 25 dihydroxyvitamin D at the level of the gene.[66]

Osteocalcin is synthesised by osteoblasts as preproosteocalcin, with two sequences being removed by peptidase cleavage before secretion of the mature protein.[67] It is mostly incorporated into bone, but a small amount enters the circulation to be cleared by the kidneys and liver; osteocalcin has an average half-life of 10 minutes.[68] Postsynthesis, osteocalcin is metabolised so that one-third is intact, one-third is a large N-terminal fragment, and one-third is of "midmolecule" length, or smaller.[69] Theoretically, osteocalcin measurement should only

reflect bone formation. However, it is possible that fragments of osteocalcin are released on resorption. In addition, catabolism of the intact molecule can lead to immunoreactive fragments in the circulation. It is unclear whether serum osteocalcin levels correspond to matrix synthesis or mineralization.[70–72]

Osteocalcin heterogeneity has led to variation in the results obtained for serum concentrations, which has, in turn, led to a lack of clarity regarding its function.[69] Each of the commercial assays, including those manufactured by Quest Diagnostics, Quidel Corporation, Diagnostic Systems Laboratories (Webster, Texas, USA), CISbio International (Marcoule, France), DiaSorin Inc (Stillwater, MN, USA) and IDS (Fountain Hills, AZ, USA) give different answers. This heterogeneity in assays is reflected in healthy adults, but can become more marked in subjects with renal failure and Paget's disease.[69] Further problems associated with osteocalcin measurement relate to the instability of intact osteocalcin in serum samples and the poor comparability of different methods. Serum values can differ more than twofold with respect to the same osteocalcin standard, even when immunoassays are said to be specific for the intact molecule.[73] A complete review of the different assays for osteocalcin has been recently published.[74]

In healthy adults, it is believed that the major circulating fragments are the intact molecule and a large N-terminal fragment (amino acid residues 1 to 43). Levels of other immunoreactive fragments (mainly from the C-terminal) are increased in renal failure.[75] Antibodies picking up the intact and N-terminal fragments are considered to be the most accurate in assessment of bone formation.

Osteocalcin levels are raised in children, peaking at puberty, and falling to adult levels; changes correlate to growth velocity.[76] Levels of osteocalcin rise in men >60 years and after the menopause in women. In pregnancy, osteocalcin decreases throughout the first and second trimesters, returning to normal just before delivery.[77] Osteocalcin levels exhibit a diurnal rhythm, with an early-morning peak (~2.00 a.m.), which is 10% to 30% higher than nadir values at noon.[78] Peak cortisol levels precede a low osteocalcin concentration by 4 hours,[79] which is a glucocorticoid that functions to suppress osteocalcin gene expression.[80]

To counteract some of the fragment problems and clarify the osteocalcin situation, several two-site IRMAs have been produced that only recognise intact osteocalcin and the large N-terminal fragment. One of the earlier two-site IRMAs was developed by Garnero *et al*.[69] using two monoclonal antibodies and standardizing the assay with human osteocalcin. This method detects intact osteocalcin and the large midregion peptide fragment (1–44) found to make up 50% of the total osteocalcin concentration in normal subjects and Paget's disease subjects, and up to 75% of the total osteocalcin concentration in chronic renal failure subjects.

This assay is the basis of the Quest Diagnostics osteocalcin kit, which uses two polyclonal antibodies to amino acids 20 to 36 and 1 to 19. Considering the instability of osteocalcin, Garnero *et al*.[69] found that 50% to 70% of immunoreactivity was lost in serum samples after 24 hours at room temperature. The instability of osteocalcin in serum is primarily due to a labile six-amino acid C-terminal sequence, cleaved off from the molecule *in vivo* and *in vitro*. Using an antibody that binds to the 20–36 region of the osteocalcin peptide ensures that the C-terminal

cleaved sequence is not detected, improving the sensitivity of the assay. Several other double antibody assays have been developed, in an attempt to improve sensitivity and specificity; the majority are two-site enzyme immunoassays. All of these assays use a variation of monoclonal and polyclonal antibodies raised against various osteocalcin sequences, but in the main, are all standardized with human osteocalcin.[69,81–85]

A chemiluminescence immunoassay was developed for osteocalcin by Kao *et al*.[86] and a chemiluminescent immunometric assay was developed by the Quest Diagnostics.

Various studies have considered the comparability of osteocalcin assays and standardization. Power *et al*.[87] showed that, although various assays showed no differences between different methods when assaying osteocalcin in bone, markedly different answers were obtained using serum samples. Delmas *et al*.[88] demonstrated marked differences in osteocalcin concentrations between eight different laboratories. Masters *et al*.[73] showed similar variations when comparing eight commercial kits in a range of healthy subjects and subjects with metabolic bone disease. In primary hyperparathyroidism, in which bone formation and resorption remain coupled, there was reasonable agreement between the kits, whereas in Paget's disease, in which bone formation and resorption are uncoupled, osteocalcin was not sensitive enough for detection of disease or monitoring response to treatment. In osteoporosis, a wide range of osteocalcin values was obtained in the same subject with different kits, even if expressed as multiples of the control means. In one subject, osteocalcin was 95% of the control value with one assay, but 250% with another. Unfortunately, this discrepancy was not resolved by the use of the respective manufacturers' reference ranges. One assay showed 8 out of 10 osteoporotic subjects to have an above normal osteocalcin level, whereas another assay classified these subjects as nine normal values and one low value. Finally, a review by Diego *et al*.[89] comparing three RIAs and three IRMAs concluded that the six assays did not recognise the same fragments of osteocalcin. In a paper by Colford *et al*.[90] five osteocalcin assays were compared with respect to tracer specificity, fragment interference, and calibration.

Serum or plasma can be used for osteocalcin measurement, but haemolysed and lipaemic samples should be avoided, because erythrocyte proteases can degrade osteocalcin[91] and osteocalcin binds to lipids, rendering it nonimmunoreactive. Samples are also very sensitive to freeze–thawing, which should be avoided owing to the labile nature of the protein.

Reference ranges are supplied by most osteocalcin kit manufacturers. These, however, vary, and commonly are not calculated for the populations in which osteocalcin is likely to be measured, such as postmenopausal women or children.[74] For example, the NovoCalcin® (Quidel Corporation) competitive immunoassay for osteocalcin quotes reference ranges of 3.7–10.0 ng/mL for females and 3.4–9.1 ng/mL for males. Because of variations in the reference ranges, it is recommended that laboratories develop their own, using a particular assay and sampling from the clinical population in which the assay will be used.

13.3.3. Procollagen Propeptides

Collagen is synthesised as procollagen, which contains extension peptides that are endoproteolytically cleaved from amino (PINP) and carboxy (PICP) terminal ends, in a stoichiometric relationship with collagen biosynthesis.[92] Both peptides can be measured in serum. Processing of type I collagen from soft tissues contributes to PICP and PINP pools, but the effect should be small, because the quantity of bone being formed is far greater than soft tissues.[92]

PICP is a globular protein rich in oligosaccharide side chains,[93] which is cleared by the hepatic mannose receptor ($t_{\frac{1}{2}}$, 6–8 min).[94] Subjects have been identified who have inherited defective clearance leading to elevated levels,[95] but without apparent pathological consequence.

PINP is a trimeric, elongated protein, which is held together by a domain of collagenous triple helix. The antigenicity of PINP arises from intact PINP and a smaller (Col I) domain of the pro αI (I) chain. This is the most immunogenic, amino-terminal part of the chain, thought to be a degradation product of PINP. Different assays detect these epitopes to different extents. PINP is cleared by scavenger receptors on the endothelial cells of the liver,[96,97] and the Col I domain is cleared by the kidney.[93]

After cleavage from collagen, the propeptides enter the circulation or bone matrix. PINP and PICP circulate in microgramme concentrations and many immunoassays for these propeptides of type I collagen have been developed.[27,94,98–100] Homology between PICP and type III collagen (PIIICP) caused crossreactivity in the early assays,[101] but PIIICP circulates at very low concentrations, contributing little to measured results.

PICP assays have been produced by Taubman *et al*.,[102] Simon *et al*.,[101] Melkko *et al*.[103] (marketed by Orion Diagnostica (Espoo, Finland) and the Quidel Corporation), and Pedersen and Bonde.[98] In many of these assays, collagen from cultured skin or lung fibroblasts is digested by bacterial collagenase to yield the propeptides to which polyclonal antibodies are then raised. However, it is now recognised that cleavage by bacterial collagenase produces PICP with an amino-terminal end distinct from that produced *in vivo*.[103] To rectify this, Pedersen and Bonde's assay uses free procollagen peptide from human foetal fibroblasts as the standard and tracer. In this assay, PICP and PIIICP are also separated. Melkko *et al*.'s RIA uses polyclonal rabbit antibodies to human skin fibroblast PICP. Only native PICP proteins are used as the reference antigen in all of these assays and because endogenous antigen in serum is homogeneous, all assays should give similar results.

PINP assays have been developed by Davis and Madri,[104] Rasmusson *et al*.,[105] Ebeling *et al*.,[106] Linkhart *et al*.,[99] and Melkko *et al*.[92] Some of these assays have used synthetic amino acids as the immunogen and for the preparation of the standards and tracer. One assay used amino acids 7–24 of the α chain of PINP and Linkhart *et al*.'s assay using peptide 23–34 because this is the most antigenic site containing the Col I domain. Melkko *et al*.[92] isolated intact PINP from pleural fluid of cancer subjects and treated this with collagenase to yield the amino-terminal Col I domain of PINP, which remained trimeric.

Recent reports have suggested that both PICP and PINP could have some use in detecting and monitoring bone metastases in breast and prostate cancers. Higher values of PINP compared with PICP are found in subjects with active metastases, suggesting that the PINP: PICP ratio could be a useful marker of bone secondary progression. Other groups have shown that PINP alone can predict changes in BMD in postmenopausal breast cancer subjects, and PINP seems to be more responsive to changes in bone turnover than PICP. In fracture and osteoporosis intervention studies, PINP has been shown to be very sensitive in detecting changes in bone turnover. However, it is still unclear regarding whether PINP will be more useful or provide any extra information than bone-specific ALP measurement. Other studies show that changes in PINP and PICP are sometimes discordant and suggest the ratio of these two propeptides could be a useful clinical measurement.

As for many other markers of bone turnover, reference ranges vary with the assay used. In adults' serum, PINP concentrations are in the range of 20–90 μg/L,[92,107–109] whereas the reference range for PICP has been quoted in SI units as 0.36–1.44 nmol/L in healthy women >30 years.[98] The concentrations of PICP have been found to vary to a greater extent, especially with age. Manufacturers such as Orion Diagnostica, who produce a PICP RIA, quote age-related reference values, with a mean of 50–170 μg/L for males and 38–202 μg/L for females, which were observed to decline with age. Specific paediatric and adolescent reference ranges have also been determined.[110–112]

13.4. Summary

Studies of biochemical bone markers have assessed the clinical validity and usefulness of individual or groups of markers in various disease states and in response to therapy. These have been a mix of comparative resorption *or* formation *or* resorption and formation markers and are mainly cross-sectional. Earlier studies compared markers with direct measures of bone status or activity (calcium kinetics, histomorphometry, or bone density), whereas newer studies compare markers against each other or changes in BMD over time. Several reviews have summarized these findings.[13,113–119] Different markers can be useful in different disease states, depending on the mechanism of altered bone turnover.

In osteoporosis, DPyd and NTx are the most significantly elevated markers in subjects with hip fracture and are discriminatory for detecting increased turnover.[117,120,121] Bisphosphonate treatment reduces the levels of resorption markers within 1 to 3 months and formation markers within 3 to 9 months.[122,123] Only DPyd and NTx decline to premenopausal levels on bisphosphonate treatment, with the percentage decrease at 3 months for PICP, ALP, osteocalcin, Pyrilinks-D (DPD), and NTx correlating to increased spinal BMD at 24 months.[122] Other studies, however, found free crosslinks to be unresponsive to bisphosphonate therapy.[24,123] However, a recent study investigating the newly developed serum assays for CTx provides encouraging data, showing large changes with bisphosphonate therapy and a low coefficient of variation.[26]

Similar findings in bisphosphonate-treated Paget's disease show the greatest suppression in NTx and CTx from baseline compared with other markers of bone resorption, such as OHP, ICTP, and TRACP.[124] Diagnostically, ALP is most useful marker for assessing activity of Paget's disease, with bone-specific ALP becoming a more sensitive marker of residual activity if TALP is within the normal range.[125]

At the menopause, bone-specific ALP seems to be the most responsive bone-formation marker and NTx seems to be the most responsive resorption marker,[120,122,126,127] responding to a 37–52% increase in formation and a 79–97% increase in resorption at this time. Of the newer formation markers, PICP levels change in parallel to osteocalcin and bone-specific ALP levels in subjects on HRT and steroids,[33,128] although PICP has less discriminatory power.[89] Bone-specific ALP and osteocalcin, however, are discordant in a number of situations, such as Paget's disease, renal failure, metastatic bone disease, and osteomalacia.[115,129] This could be owing to osteocalcin being a protein, integral to bone matrix, and responsive to some extent to resorption, in addition to formation.

Many unresolved issues surround bone-marker measurement. At present, assay heterogeneity, lack of standardization, external QA programmes, and relatively poor assay precision linked to high intraindividual variation has limited their widespread adoption in routine clinical practice. The 95% confidence intervals are wide, limiting the ability to classify subjects into subgroups according to turnover. Formation markers tend to have tighter CVs than resorption markers, probably because most are measured in serum, but this is assay-dependent. There is a need for definitive assessment of assay standardization and minimization of variability before assessment of clinical value can be determined in longitudinal studies.

References

1. Calvo, M.S., Eyre, D.R. and Gundberg, C.M. (1996). Molecular basis and clinical application of biological markers of bone turnover. *Endocr Rev* **17:**333–68.
2. Reid, K.B.M. (1974). A collagen-like amino acid sequence in a polypeptide chain of human C1q (a subcomponent of the first component of complement). *Biochem J* **141:**189–203.
3. Deacon, A.C., Hulme, P., Hesp, R. *et al*. (1987). Estimation of whole body bone resorption rate: a comparison of urinary total hydroxyproline excretion with two radioisotopic tracer methods in osteoporosis. *Clin Chim Acta* **166:**297–306.
4. Reed, P., Holbrook, I.B., Gardener, M.L.G. *et al*. (1991). Simple, optimized liquid-chromatographic method for measuring total hydroxyproline in urine evaluated. *Clin Chem* **37:**285–90.
5. Bettica, P., Moro, L., Robins, S.P. *et al*. (1992). Bone resorption markers galactosyl hydroxylysine, pyridinium crosslinks and hydroxyproline compared. *Clin Chem* **38:**2313–18.

6. Moro, L., Gazzarrini, C., Crivellari, D. *et al*. (1993). Biochemical markers for detecting bone metastases in patients with breast cancer. *Clin Chem* **39:**131–4.
7. Seibel, M.J., Robins, S.P. and Bilezikian, J.P. (1992). Urinary pyridinium crosslinks of collagen—specific markers of bone resorption in metabolic bone disease. *Trends Endocrinol Metab* **3:**263–70.
8. James, I. and Crowley, C. (1993). Assay of pyridinium crosslinks in serum using narrow bore ion paired reversed phase high performance liquid chromatography. *J Chromatogr* **612:**41–8.
9. Ogawa, T., Ono, T., Tsuda, M. *et al*. (1982). A novel fluor in insoluble collagen: a crosslinking moiety in collagen molecule. *Biochem Biophys Res Commun* **107:** 1252–7.
10. Uebelhart, D., Gineyts, E., Chapuy, M.-C. *et al*. (1990). Urinary excretion of pyridinium crosslinks: a new marker of bone resorption in metabolic bone disease. *Bone Miner* **8:**87–96.
11. Robins, S.P. (1995). Collagen crosslinks in metabolic bone disease. *Acta Orthop Scand* **66(Suppl 266):**171–5.
12. Eyre, D.R., Koob, T.J. and Van Ness, K.P. (1984). Quantitation of hydroxypyridinium crosslinks in collagen by high performance liquid chromatography. *Anal Biochem* **137:**380–8.
13. Kollerup, G., Thamsborg, G., Bathia, H. *et al*. (1992). Quantitation of urinary hydroxypyridinium cross-links from collagen by high performance liquid chromatography. *Scand J Clin Lab Invest* **52:**657–62.
14. Robins, S.P., Woitge, H.W., Hesley, R.P. *et al*. (1994). Direct, enzyme linked immunoassay for urinary deoxypyridinoline as a specific marker for measuring bone resorption. *J Bone Miner Res* **9:**1643–9.
15. Abbiati, G., Rigoldi, M., Frignani, S. *et al*. (1994). Determination of pyridinium crosslinks in plasma and serum by high performance liquid chromatography. *J Chromatogr* **656:**303–10.
16. Robins, S.P., Duncan, A., Wilson, N. *et al*. (1996). Standardization of pyridinium crosslinks, pyridinoline and deoxypyridinoline, for use as biochemical markers of collagen degradation. *Clin Chem* **42:**1621–6.
17. Beardsworth, L.J., Eyre, D.R. and Dickson, I.R. (1990). Changes with age in the urinary excretion of lysyl and hydroxylysylpyridinoline, two new markers of bone collagen turnover. *J Bone Miner Res* **5:**671–6.
18. Gerrits, M.I., Thijssen, J.H.H. and van Rijn, H.J.M. (1995). Determination of pyridinoline and deoxypyridinoline in urine, with special attention to retaining their stability. *Clin Chem* **41:**571–4.
19. McLaren, A.M., Isdale, A.H., Whiting, P.H. *et al*. (1993). Physiological variations in the urinary excretion of pyridinium crosslinks of collagen. *Br J Rheum* **32:**307–12.
20. Bjorgaas, M., Haug, E. and Johnsen, H.J. (1999). The urinary excretion of deoxypyridinium cross-links is higher in diabetic than nondiabetic adolescents. *Calcif Tissue Int* **65:**121–4.
21. Garnero, P., Gineyts, E., Arbault, P. *et al*. (1995). Different effects of bisphosphonate and oestrogen therapy on free and peptide bound bone crosslinks excretion. *J Bone Miner Res* **10:**641–9.
22. Kamel, S., Brazier, M., Rogez, J.C. *et al*. (1996). Different responses of free and peptide bound crosslinks to vitamin D and calcium supplementation in elderly women with vitamin D insufficiency. *J Clin Endocrinol Metab* **81:**3717–21.

23. Hanson, D.A., Weis, M.E., Bollen, A. *et al*. (1992). A specific immunoassay for monitoring human bone resorption: quantitation of type 1 collagen crosslinked N-telopeptides in urine. *J Bone Miner Res* **7:**1251–8.
24. Melkko, J., Niemi, S., Risteli, J. *et al*. (1990). Radioimmunoassay of the carboxyterminal propeptide of human type 1 procollagen. *Clin Chem* **36:**1328–32.
25. Bonde, M., Qvist, P., Fledelius, C. *et al*. (1994). Immunoassay for quantifying type 1 collagen degradation products in urine evaluated. *Clin Chem* **40:**2022–5.
26. Rosen, H.N., Moses, A.C., Garber, J. *et al*. (2000). Serum CTX: a new marker of bone resorption that shows treatment effect more often than other markers because of low coefficient of variability and large changes with bisphosphonate therapy. *Calcif Tissue Int* **66:**100–3.
27. Eriksen, E.F., Charles, P., Melsen, F. *et al*. (1993). Serum markers of type 1 collagen formation and degradation in metabolic bone disease: correlation with bone histomorphometry. *J Bone Miner Res* **8:**127–32.
28. Calvo, M.S., Eyre, D.R. and Gundberg, C.M. (1996). Molecular basis and clinical application of biological markers of bone turnover. *Endocr Rev* **17:**333–68.
29. Ebeling, P.R., Atley, L.M., Guthrie, J.R. *et al*. (1996). Bone turnover markers and bone density across the menopausal transition. *J Clin Endocrinol Metab* **81:**3366–71.
30. Kamel, S., Brazier, M., Rogez, J.C. *et al*. (1996). Different responses of free and peptide bound crosslinks to vitamin D and calcium supplementation in elderly women with vitamin D insufficiency. *J Clin Endocrinol Metab* **81:**3717–21.
31. Bonde, M., Fledelius, C., Qvist, P. *et al*. (1996). Coated-tube radioimmunoassay for C-telopeptides of type I collagen to assess bone resorption. *Clin Chem* **42:**1639–44.
32. Zanze, M., Souberbielle, J.C., Kindermans, C. *et al*. (1997). Procollagen propeptide and pyridinium cross-links as markers of type I collagen turnover: sex- and age-related changes in healthy children. *J Clin Endocrinol Metab* **82:**2971–9.
33. Price, C.P., Kirwan, A. and Vader, C. (1995). Tartrate resistant acid phosphatase as a marker of bone resorption. *Clin Chem* **41:**641–3.
34. Cheung, C.K., Panesar, N.S., Haines, C. *et al*. (1995). Immunoassay of a tartrate resistant acid phosphatase in serum. *Clin Chem* **41:**679–86.
35. Halleen, J.M., Hentunen, T.A., Karp, M. *et al*. (1998). Characterization of serum tartrate resistant acid phosphatase and development of a direct two site immunoassay. *J Bone Miner Res* **13:**683–7.
36. Weiss, M.J., Henthorn, P.S., Lafferty, M.A. *et al*. (1986). Isolation and characterization of a cDNA encoding a human liver/bone/kidney-type alkaline phosphatase. *Proc Natl Acad Sci USA* **83:**7182–6.
37. Weiss, M.J., Ray, K., Henthorn, P.S. *et al*. (1988). Structure of the human liver/bone/kidney alkaline phosphatase gene. *J Biol Chem* **263:**12002–10.
38. Whyte, M.P. (1994). Hypophosphatasia and the role of alkaline phosphatase in skeletal mineralization. *Endocr Rev* **15:**439–61.
39. Farley, J.R., Hall, S.L., Ilacas, D. *et al*. (1994). Quantification of skeletal alkaline phosphatase in osteoporotic serum by wheat germ agglutinin precipitation, heat inactivation, and a two-site immunoradiometric assay. *Clin Chem* **40:**1749–56.
40. van Hoof, V. and De Broe, M.E. (1994). Interpretation and clinical significance of alkaline phosphatase isoenzyme patterns. *Crit Rev Clin Lab Sci* **31:**197–293.
41. Moss, D.W. (1982). Alkaline phosphatase isoenzymes. *Clin Chem* **28:**2007–16.
42. Price, C.P. (1993). Multiple forms of human serum alkaline phosphatase: detection and quantitation. *Ann Clin Biochem* **30:**355–72.

43. Rosalki, S.B. and Ying Foo, A. (1984). Two new methods for separating and quantifying bone and liver alkaline phosphatase isoenzymes in plasma. *Clin Chem* **30:**1182–6.
44. Moss, D.W. and Edwards, R.K. (1984). Improved electrophoretic resolution of bone and liver alkaline phosphatases resulting from partial digestion with neuraminidase. *Clin Chim Acta* **143:**177–82.
45. Mattiazzo, M. and Ramasamy, I. (1993). Wheat germ lectin affinity electrophoresis of serum alkaline phosphatase with commercially available agarose gels. *Clin Chem* **39:**1404–7.
46. Kuwana, T., Sugita, O. and Yakata, M. (1988). Reference limits of bone and liver alkaline phosphatase isoenzymes in the serum of healthy subjects according to age and sex as determined by wheat germ lectin affinity electrophoresis. *Clin Chim Acta* **173:**273–80.
47. Sorensen, S. (1988). Wheat germ agglutinin method for measuring bone and liver isoenzymes of alkaline phosphatase assessed in postmenopausal osteoporosis. *Clin Chem* **34:**1636–40.
48. Behr, W. and Barnert, J. (1986). Quantification of bone alkaline phosphatase in serum by precipitation with wheat germ lectin: A simplified method and its clinical plausibility. *Clin Chem* **32:**1960–6.
49. Farley, J.R., Hall, S.L., Ilacas, D. *et al*. (1994). Quantification of skeletal alkaline phosphatase in osteoporotic serum by wheat germ agglutinin precipitation, heat inactivation, and a two-site immunoradiometric assay. *Clin Chem* **40:**1749–56.
50. Burlina, A., Plebani, M., Secchiero, S. *et al*. (1991). Precipitation method for separating and quantifying bone and liver alkaline phosphatase isoenzymes. *Clin Biochem* **24:**417–23.
51. Rosalki, S.B., Ying Foo, A., Burlina, A. *et al*. (1993). Multicentre evaluation of Iso—ALP test kit for measurement of bone alkaline phosphatase activity in serum and plasma. *Clin Chem* **39:**648–52.
52. Lawson, G.M., Katzmann, J.A., Kimlinger, T.K. *et al*. (1985). Isolation and preliminary characterization of a monoclonal antibody that interacts preferentially with the liver isoenzyme of human alkaline phosphatase. *Clin Chem* **31:**381–5.
53. Hill, C.S. and Wolfert, R.L. (1989). The preparation of monoclonal antibodies which react preferentially with human bone alkaline phosphatase and not liver alkaline phosphatase. *Clin Chim Acta* **186:**315–20.
54. Garnero, P. and Delmas, P.D. (1993). Assessment of the serum levels of bone alkaline phosphatase with a new immunoradiometric assay in patients with metabolic bone disease. *J Clin Endocrinol Metab* **77:**1046–53.
55. Panigrahi, K., Delmas, P.D., Singer, F.R. *et al*. (1994). Characteristics of a two Site immunoradiometric assay for human skeletal alkaline phosphatase serum. *Clin Chem* **40:**822–8.
56. England, T.E., Samsoondar, J. and Maw, G. (1994). Evaluation of the Hybritech Tandem-R Ostase immunoradiometric assay for skeletal alkaline phosphatase. *Clin Biochem* **27:**187–9.
57. Van Hoof, V.O., Martin, M., Blockx, P. *et al*. (1995). Immunoradiometric method and electrophoretic system compared for quantifying bone alkaline phosphatase in serum. *Clin Chem* **41:**853–7.
58. Crofton, P.M. (1992). Wheat germ lectin affinity electrophoresis for alkaline phosphatase isoforms in children: age dependent reference ranges in liver and bone disease. *Clin Chem* **38:**663–70.

59. Okesina, A.B., Donaldson, D., Lascelles, P.T. *et al*. (1995). Effect of gestational age on levels of serum alkaline phosphatase isoenzymes in healthy pregnant women. *Int J Gynaecol Obstet* **48:**25–9.
60. Melton, L.J., Khosla, S., Atkinson, E.J. *et al*. (1997). Relationship of bone turnover to bone density and fractures. *J Bone Miner Res* **12:**1083–91.
61. Bouman, A.A., Scheffer, P.G., Ooms, M.E. *et al*. (1995). Two bone alkaline phosphatase assays compared with osteocalcin as a marker of bone formation in healthy elderly women. *Clin Chem* **41:**196–9.
62. Hauschka, P.V. and Carr, S.A. (1982). Calcium-dependent alpha helical structure in osteocalcin. *Biochemistry* **21:**2538–47.
63. Koshihara, Y. and Hoshi, K. (1997). Vitamin K_2 enhances osteocalcin accumulation in the extracellular matrix of human osteoblasts *in vitro*. *J Bone Miner Res* **12:** 431–8.
64. Liu, G. and Peacock, M. (1998). Age-related changes in serum uncarboxylated osteocalcin and its relationships with bone density, bone quality, and hip fracture. *Calcif Tissue Int* **62:**286–9.
65. Hart, J.P., Shearer, M.J., Klenerman, L. *et al*. (1985). Electrochemical detection of depressed circulating levels of vitamin K_1 in osteoporosis. *J Clin Endocrinol Metab* **60:**1268–9.
66. Puchacz, E., Lian, J.B., Stein, G.S. *et al*. (1989). Chromosomal localization of the human osteocalcin gene. *Endocrinology* **124:**2648–50.
67. Price, P.A., Williamson, M.K. and Lothringer, J.W. (1981). Origin of the vitamin K-dependent bone protein found in plasma and its clearance by kidney and bone. *J Biol Chem* **256:**12760–6.
68. Farrugia, W. and Melick, R.A. (1986). Metabolism of osteocalcin. *Calcif Tissue Int* **39:**234–8.
69. Garnero, P., Grimaux, M., Seguin, P. *et al*. (1994). Characterization of immunoreactive forms of human osteocalcin generated *in vivo* and *in vitro. J Bone Miner Res* **9:**255–64.
70. Van Der Eems, K.L., Brown, R.D. and Gundberg, C.M. (1985). Circulating levels of 1, 25 dihydroxyvitamin D, alkaline phosphatase, hydroxyproline and osteocalcin associated with antler growth in white-tailed deer. *Acta Endocrinol* **118:** 407–14.
71. Demiaux, B., Arlot, M.E., Chapuy, M.C. *et al*. (1992). Serum osteocalcin is increased in patients with osteomalacia: correlations with biochemical and histomorphometric findings. *J Clin Endocrinol Metab* **74:**1146–51.
72. Ducy, P., Desbois, C., Boyce, B.F. *et al*. (1996). Increased bone formation in osteocalcin deficient mice. *Nature* **382:**448–52.
73. Masters, P.W., Jones, R.G., Purves, D.A. *et al*. (1994). Commercial assays for serum osteocalcin give clinically discordant results. *Clin Chem* **40:**358–63.
74. Lee, A.J., Hodges, S. and Eastell, R. (2000). Measurement of osteocalcin. *Ann Clin Biochem* **37:**432–46.
75. Gundberg, C.M. and Weinstein, R.S. (1986). Multiple immunoreactive forms of osteocalcin in uraemic serum. *J Clin Invest* **77:**1762–7.
76. Klein, G., Wadlington, E.L., Collins, E.D. *et al*. (1984). Calcitonin levels in sera of infants and children: relations to age and periods of bone growth. *Calcif Tissue Int* **36:**635–8.
77. Tarallo, P., Henny, J., Fournier, B. *et al*. (1990). Plasma osteocalcin: biological variations and reference limits. *Scand J Clin Lab Invest* **50:**649–55.

78. Nielsen, H.K., Brixen, K., Kassem, M. *et al*. (1992). Inhibition of the morning cortisol peak abolishes the expected morning decrease in serum osteocalcin in normal males: evidence of a controlling effect of serum cortisol on the circadian rhythm in serum osteocalcin. *J Clin Endocrinol Metab* **74:**1410–14.
79. Nielsen, H.K., Laurberg, P., Brixen, K. *et al*. (1991). Relations between diurnal variations in serum osteocalcin, cortisol, parathyroid hormone and ionized calcium in normal individuals. *Acta Endocrinol* **124:**391–8.
80. Morrison, N.A., Shine, J., Fragonas, J.-C. *et al*. (1989). 1,25-dihydroxyvitamin D-responsive element and glucocorticoid repression in the osteocalcin gene. *Science* **246:**1158–61.
81. Wenz, I., Reissmann, R. and Bornig, H. (1991). Determination of osteocalcin in serum by an ultramicro-ELISA with alkaline phosphatase as marker enzyme. *Biomed Biochem Acta* **50:**145–9.
82. Kuronen, I., Kokko, H. and Parivianen, M. (1993). Production of monoclonal and polyclonal antibodies against human osteocalcin sequences and development of a two site ELISA for intact human osteocalcin. *J Immunol Methods* **163:**233–40.
83. Monaghan, D.A., Power, M.J. and Fottrell, P.F. (1993). Sandwich enzyme immunoassay of osteocalcin in serum with use of an antibody against human osteocalcin. *Clin Chem* **39:**942–7.
84. Parivianen, M., Kuronen, I., Kokko, H. *et al*. (1994). Two-site enzyme immunoassay for measuring intact human osteocalcin in serum. *J Bone Miner Res* **9:**347–54.
85. Rosenquist, C., Bonde, M., Fledelius, C. *et al*. (1994). A simple enzyme-linked immunosorbent assay of human osteocalcin. *Clin Chem* **40:**1258–64.
86. Kao, P.C., Riggs, B.L. and Schryver, P.G. (1993). Development of an osteocalcin chemiluminoimmunoassay. *Clin Chem* **39:**1369–74.
87. Power, M.J., Gosling, J.P. and Fottrell, P.F. (1989). Radioimmunoassay of osteocalcin with polyclonal and monoclonal antibodies. *Clin Chem* **35:**1408–15.
88. Delmas, P.D., Christiansen, C., Mann, K.G. *et al*. (1990). Bone gla protein (osteocalcin) assay standardization report. J *Bone Miner Res* **5:**5–11.
89. Diego, E.M.D., Guerrero, R. and de la Piedra, C. (1994). Six osteocalcin assays compared. *Clin Chem* **40:**2071–7.
90. Colford, J., Sailer, D. and Longman, C. (1997). Five osteocalcin assays compared: tracer specificity, fragment interference, and calibration. *Clin Chem* **43:**1240–1.
91. Power, M.J., O'Dwyer, B., Breen, E. *et al*. (1991). Osteocalcin concentrations in plasma prepared with different anticoagulants. *Clin Chem* **37:**281–4.
92. Melkko, J., Kauppila, S., Niemi, S. *et al*. (1996). Immunoassay for intact amino terminal propeptide of human type 1 procollagen. *Clin Chem* **42:**947–54.
93. Olsen, B.R., Guzman, N.A., Engel, J. *et al*. (1977). Purification and characterization of a peptide from the carboxy-terminal region of chick tendon procollagen type I. *Biochemistry* **16:**3030–6.
94. Smedsrod, B., Melkko, J., Risteli, L. *et al*. (1990). Circulating C-terminal propeptide of type I procollagen is cleared mainly via the mannose receptor in liver endothelial cells. *Biochem J* **271:**345–50.
95. Sorva, A., Taähtelaä, R., Risteli, J. *et al*. (1994). Familial high serum concentrations of the carboxy-terminal propeptide of type 1 procollagen (P1CP). *Clin Chem* **40:**1591–3.
96. Melkko, J., Hellevik, T., Risteli, L. *et al*. (1994). Clearance of NH2-terminal propeptides of types I and III procollagen is a physiological function of the scavenger receptor in liver endothelial cells. *J Exp Med* **179:**405–12.

97. Savolainen, E.R., Goldberg, B., Leo, M.A. *et al*. (1984). Diagnostic value of serum procollagen peptide measurements in alcoholic liver disease. *Alcohol Clin Exp Res* **8:**384–9.
98. Pedersen, B.J. and Bonde, M. (1994). Purification of human procollagen type 1 carboxyl terminal propeptide cleaved as *in vivo* from procollagen and used to calibrate a radioimmunoassay of the propeptide. *Clin Chem* **40:**811–16.
99. Linkhart, S.G., Linkhart, T.A., Taylor, A.K. *et al*. (1993). Synthetic peptide based immunoassay for amino terminal propeptide of type 1 procollagen: application for evaluation of bone formation. *Clin Chem* **39:**2254–8.
100. Risteli, J., Niemi, S., Trivedi, P. *et al*. (1988). Rapid equilibrium radioimmunoassay for the amino terminal propeptide of human type III procollagen. *Clin Chem* **34:**715–18.
101. Simon, L.S., Krane, S.M., Wortman, P.D. *et al*. (1984). Serum levels of type I and III procollagen fragments in Paget's disease of bone. *J Clin Endocrinol Metab* **58:**110–20.
102. Taubman, M.B., Goldberg, B. and Sherr, C.J. (1974). Radioimmunoassay for human procollagen. *Science* **186:**1115–17.
103. Melkko, J., Niemi, S., Risteli, L. *et al*. (1990). Radioimmunoassay for the carboxyterminal propeptide of human type I procollagen. *Clin Chem* **36:**1328–32.
104. Davis, B.H. and Madri, J.A. (1987). An immunohistochemical and serum ELISA study of type I and type III procollagen aminopropeptides in primary biliary cirrhosis. *Am J Pathol* **128:**265–75.
105. Rasmusson, H.B., Teisner, B., Bangsgaard-Petersen, F. *et al*. (1992). Quantification of fetal antigen 2 (FA2) in supernatants of cultured osteoblasts, normal human serum, and serum from patients with chronic renal failure. *Nephrol Dial Transplant* **7:**902–7.
106. Ebeling, P.R., Peterson, J.M. and Riggs, B.L. (1992). Utility of type I procollagen propeptide assays for assessing abnormalities in metabolic bone diseases. *J Bone Miner Res* **7:**1243–50.
107. Tahtela, R., Turpeinen, M., Sorva, R. *et al*. (1997). The aminoterminal propeptide of type I procollagen: evaluation of a commercial radioimmunoassay kit and values in healthy subjects. *Clin Biochem* **30:**35–40.
108. Orum, O., Hansen, M., Jensen, C.H. *et al*. (1996). Procollagen type I N-terminal propeptide (PINP) as an indicator of type I collagen metabolism: ELISA development, reference interval, and hypovitaminosis D induced hyperparathyroidism. *Bone* **2:**157–63.
109. Linkhart, S.G., Linkhart, T.A., Taylor, A.K. *et al*. (1993). Synthetic peptide-based immunoassay for amino-terminal propeptide of type I procollagen: application for evaluation of bone formation. *Clin Chem* **39:**2254–8.
110. Crofton, P.M., Wade, J.C., Taylor, M.R. *et al*. (1997). Serum concentrations of carboxyl-terminal propeptide of type I procollagen, amino-terminal propeptide of type III procollagen, cross-linked carboxyl-terminal telopeptide of type I collagen, and their interrelationships in schoolchildren. *Clin Chem* **43:**1577–81.
111. Trivedi, P., Risteli, J., Risteli, J. *et al*. (1991). Serum concentrations of the type I and III procollagen propeptides as biochemical markers of growth velocity in healthy infants and children and in children with growth disorders. *Pediatr Res* **30:**276–80.
112. Lieuw-A-Fa, M., Sierra, R.I. and Specker, B.L. (1995). Carboxy-terminal propeptide of human type I collagen and pyridinium cross-links as markers of bone growth in infants 1 to 18 months of age. *J Bone Miner Res* **10:**849–53.

113. Christenson, R.H. (1997). Biochemical markers of bone metabolism: an overview. *Clin Biochem* **30:**573–93.
114. Blumsohn, A. and Eastell, R. (1997). The performance and utility of biochemical markers of bone turnover: do we know enough to use them in clinical practice? *Ann Clin Biochem* **34:**449–59.
115. Taylor, A.K., Lueken, S.A., Libanati, C. *et al*. (1994). Biochemical markers of bone turnover for the clinical assessment of bone metabolism. *Rheum Dis Clin North Am* **20:**589–607.
116. Blumsohn, A., Hannon, R.A. and Eastell, R. (1995). Biochemical assessment of skeletal activity. *Physical Med Rehab Clin North Am* **6:**483–505.
117. Eastell, R., Robins, S.P., Colwell, T. *et al*. (1993). Evaluation of bone turnover in type I osteoporosis using biochemical markers specific for both bone formation and bone resorption. *Osteoporos Int* **3:**255–60.
118. Rosalki, S.B. (1998). Biochemical markers of bone turnover. *Int J Clin Practice* **52:**255–6.
119. Demers, L.M. (1997). Clinical usefulness of markers of bone degradation and formation. *Scand J Clin Lab Invest* **227(Suppl.):**12–20.
120. Garnero, P., Hausherr, E., Chapuy, M.C. *et al*. (1996). Markers of bone resorption predict hip fracture in elderly women: the EPIDOS Prospective Study. *J Bone Miner Res* **11:**1531–8.
121. Seibel, M.J., Cosman, F., Shen, V. *et al*. (1993). Urinary hydroxypyridinium crosslinks of collagen as markers of bone resorption and oestrogen efficacy in postmenopausal osteoporosis. *J Bone Miner Res* **8:**881–9.
122. Garnero, P., Shih, W.J., Gineyts, E. *et al*. (1994). Comparison of new biochemical markers of bone turnover in late postmenopausal osteoporotic women in response to alendronate treatment. *J Clin Endocrinol Metab* **79:**1693–700.
123. Chesnut, C.H., 3rd, McClung, M.R., Ensrud, K.E. *et al*. (1995). Alendronate treatment of the postmenopausal osteoporotic woman: effect of multiple dosages on bone mass and bone remodelling. *Am J Med* **99:**144–52.
124. Blumsohn, A., Naylor, K.E., Assiri, A.M. *et al*. (1995). Different responses of biochemical markers of bone resorption to bisphosphonate therapy in Paget's disease. *Clin Chem* **41:**1592–8.
125. Delmas, P.D. (1999). Biochemical markers of bone turnover in Paget's disease of the bone. *J Bone Miner Res* **14:**66–9.
126. Garnero, P., Sornay-Rendu, E., Chapuy, M.C. *et al*. (1996). Increased bone turnover in late postmenopausal women is a major determinant of osteoporosis. *J Bone Miner Res* **11:**337–49.
127. Prestwood, K.M., Pilbeam, C.C., Burleson, J.A. *et al*. (1994). The short term effects of conjugated oestrogen on bone turnover in older women. *J Clin Endocrinol Metab* **79:**366–71.
128. Hassager, C., Jensen, L.T., Johansen, J.S. *et al*. (1991). The carboxy-terminal propeptide of type I procollagen in serum as a marker of bone formation: the effect of nandrolone decanoate and female sex hormones. *Metabolism* **40:**205–8.
129. Duda, R.J., O'Brien, J.F., Katzmann, J.A. *et al*. (1988). Concurrent assays of circulating bone GLA protein and bone alkaline phosphatase: effects of sex, age and metabolic bone disease. *J Clin Endocrinol Metab* **66:**951–7.

14

Future Therapies and Clinical Trials

Colin G. Miller

> Second to being right in this world is being totally wrong.
>
> T.H. Huxley

14.1. Introduction

The diagnosis and treatment of osteoporosis is still a relatively young science. Although Albright first described osteoporosis in 1947,[1] it is only since the late 1980s and early 1990s that we have had the instruments available for the diagnosis and then treatment of this disease. As for any medical area in its infancy, we can, therefore, expect a significant number of changes in the coming years, as a result of more comprehensive understanding of bone physiology and elucidation of the genetic factors leading to increased risk factors.

14.2. Future Therapies

When the first edition of this book was written, it was stated that "In the immediate future the therapeutic regimens that will be available to the prescribing physician will increase to include several different formulations of hormone replacement therapy (HRT), with several different modes of delivery, three bisphosphonates available in Europe (two in the USA), nasal calcitonin, and selective estrogen receptor modulators or SERMs." In the 4 years since the book was written, the therapeutic landscape has changed dramatically. HRT, which was once the first-line therapy for osteoporosis, is no longer used, because it was shown in the Women's Health Initiative study (2002) that women taking this drug had an increased risk of developing breast cancer.[2]

In 2002, the first recombinant parathyroid hormone (PTH) drug was introduced onto the market [PTH(1–34); Forteo® (Eli Lilly and Company, Indianapolis, USA)]. A second PTH [PTH(1–64); Preos® (NPS Pharmaceuticals, Parsipany, NJ, USA)] is approved in the European Union (EU) and under review by the US Food

and Drug Administration (FDA). Several more versions of PTH are currently in various stages of development, including forms using different methods of delivery, not just subcutaneous injection. There have been some safety issues, which on the whole have been resolved (i.e. osteosarcoma was reported in rats). It is also worth noting that, after a number of years of prescribing, this issue has not been reported in humans.

Modern medicine and treatment of disease is, generally, based on treating each disease specifically and separately. However, several new therapies (including SERMs) have provided the opportunity for prevention and/or treatment of more than one disease at a time. At the time of writing the first edition of this book, one SERM was on the market (raloxifene) and a second drug in this class was in phase III development. This latter compound, lasofoxifene, was rejected by the FDA for the indication of prevention of osteoporosis in 2005.

A novel area that has shown some major promise is the receptor for activation of NF-KB (RANK) ligand inhibitor denosumab. Phase II study results have demonstrated that twice-yearly subcutaneous injections of the compound demonstrated an increase in bone mineral density (BMD) at the lumbar spine of 5.1% after 24 months.[3]

The other major development is the bisphosphonate zoledronic acid, which has been show to increased BMD over a 12 month period following a single annual injection.[4] Although bisphosphonates have been well characterized, and are the leading treatment for osteoporosis, this finding will be an interesting development and could potentially be the first therapy, in any field (except for vaccines), for which a single annual injection provides treatment for 1 year.

14.3. Osteoporosis-Related Diseases

Other bone diseases will also appear on the "radar screen", developing the range of uses for these treatments. Already, we have seen male osteoporosis become a treatable disease, with alendronate (Fosamax®, Merck & Co. Inc., Whitehouse Station, NJ, USA). It was only a few years ago that the general opinion was that osteoporosis was a female disease. This is no longer the case. We are starting to see development in other therapeutic areas. Bone diseases in children have been an area of minimal concern, but with the FDA now promoting studies in children, we can expect to see diseases such as osteogenesis imperfecta being treated with the bisphosphonates. In later adolescence, amenorrhoea caused either by anorexia nervosa and related diseases or by extreme physical activity, for example Olympic female gymnasts, can result in decreased bone accretion in the formative years. If the level of bone deposition in these juveniles can be improved with the use of some of the new therapeutic agents, the incidence of fracture throughout an individual's life span can be reduced.

There are a number of other conditions that have a secondary effect of excessive bone loss that are currently untreated or undertreated. The use of steroids and bone loss is well documented, but not everyone is treated. Women taking luteinizing

hormone-releasing hormone (LHRH) analogues for endometriosis, bone loss after transplant owing to cisplatin use and other medications, plus subjects on chronic kidney dialysis also have bone-loss problems. As the number of subjects with transplants increases and this field of medicine develops, the effects of bone loss will become a major problem and not a side issue that can be relatively ignored.

14.4. Fixing the Fracture

No preventative treatment is currently efficacious in all subjects. Risk reduction is still the best outcome available, which means a significant number of individuals will still end up having a fracture. Fractures can be subdivided into two types, according to the current treatment and outcome: vertebral fractures and appendicular fractures.

14.4.1. Vertebral Fractures

Until recently, there was nothing available to treat a confirmed fracture of the vertebra, other than a brace and conservative noninvasive techniques. Vertebroplasty and kyphoplasty have started to come to the fore. At the time of writing, more studies are warranted to assess the long-term safety and efficacy of these techniques. Both techniques offer the solution of repairing fractured vertebrae. Vertebroplasty involves high-pressure infusion of state methyl methacrolate into the collapsed vertebra through a small hole drilled in both pedicles. Kyphoplasty is a refinement of this process, using two small angioplasty balloons that are inflated inside the accessed vertebra; the vertebra is, again, accessed through holes in the pedicles while the subject is in the recumbent position. The balloons move the trabecular bone to the cortical shell, providing a void where the cement can be injected under low pressure. The subject remains supine until the cement has hardened and can bear weight. Not only has the subject's vertebra been repaired and strengthened, but there is also a reduction in pain, which is thought to be caused by the exothermic curing process of the cement, which probably kills off the bone nociceptors.

It should be appreciated that most vertebrae do not undergo a complete collapse at one time or fracture in the classical sense of a "break" in the continuous structure of the bone, but rather undergo a series of deformities, producing a reduction in height. This leads to the challenge of diagnosing a mild fracture or deformity. Therefore, this has to be considered carefully when deciding to use this kind of approach.

Vertebroplasty and kyphoplasty only allow treatment of already deformed vertebrae. The remaining vertebrae are also at high risk of failure and fracture, so therapy must be initiated. The choice of therapies must be carefully considered to improve a subject's bone mass, although this will only be effective over a period of time, for example 1 year or more, with the current antiresorption therapies. This still might not prevent further fracture, but merely reduces the probability that more will occur. In the meantime, statistics have shown that the greatest risk of sustaining a

new fracture of the vertebrae occurs within 12 months of sustaining the first fracture. In future, based on the initial data from clinical trials, PTH will be the initial therapy of choice following vertebral collapse, because of its rapid mode of action.

Further studies must be performed to evaluate how severe the deformity must be before intervention of this kind is contemplated. Without this treatment, relatively small deformities would cause pain for a few weeks or months, but then would settle down. For many subjects, this would be relatively mild, and in many cases, is just considered a bout of back pain that does not require a visit to the physician. Even if it does, the impact of the fracture at this stage is not of sufficient morbidity for the subject to require surgery. However, it is arguably these kinds of situations in which this new treatment option might have the best outcome, because the immediately adjacent vertebrae are, perhaps, more likely to deform again. Furthermore, the immediately adjacent vertebrae could also then be treated in a preventative manner.

However, at this stage, the immediate use of this methodology must be around treating the most severely deformed vertebra, because of the invasive nature of the treatment. For subjects with multiple deformities, in which the lower rib is resting on the iliac crest, some surgeons will remove the lower ribs to improve the subject's comfort. If this can be avoided by the use of vertebral strengthening cement, this would be a valuable treatment option for these subjects.

14.4.2. Appendicular Fractures

Appendicular fractures are treated by reduction and fixing, to prevent further movement. For the femur, which is the most prevalent form of osteoporotic fracture in the elderly, this involves pinning and plating the fracture with a hemiarthroplasty or total arthroplasty. Arthroplasty is the most costly outcome not only for the medical community's budget, but also because of the comorbidity and mortality related to the fracture and treatment. The mortality rate following a fracture is 20% at 3 months and up to 35% within 1 year of the fracture.[2] There is an approximately 20% chance that a subject will go on to fracture the second hip within the next 2 years.

All the other types of appendicular fracture are not as costly, either because they are less prevalent or because they can be treated on an out-patient basis using standard methodologies. However, the bone is weak because of the nature of the disease and, therefore, might not heal rapidly or well.

In both these kinds of fracture, there is a need for bone-healing compounds to be developed. Bone morphogenic proteins (BMPs) were being evaluated a few years ago, but current research does not seem to bear out the initial promise. There are other slurries and cements that are being evaluated, but at the time of writing, nothing new seems to be coming to the forefront as the latest methodology for bone healing.

One novel compound that, at least at the time of writing, could hold some promise is the analogue of a naturally occurring somatokine. It is being used in subjects following hip fracture. The drug's primary role is in the cognitive functionality it maintains in the elderly, ensuring that far fewer of them suffer from increased

morbidity. The early studies are very promising. What is more, because of the biochemical pathway the drug affects, it has a positive effect on bone. It seems too good to be true that we can have a drug that can have a profound effect on subjects at both mental and bone levels, following such a negative event as a hip fracture.

14.5. Male Osteoporosis and Testosterone

Most of the research in the field of osteoporosis has been historically targeted at women. Women have more fractures per 1000 of the population than men. This is, in part, because of the higher bone mass that men have compared with women. However, when fracture risk per loss of 1 standard deviation (SD) in bone mass is evaluated, men have the same relative risk (RR) of sustaining a fracture as women for the same absolute BMD. Furthermore, the mortality rate among men who sustain a hip fracture is higher than that in women. Male osteoporosis is, therefore, now being approached as a serious medical disease and treatments for men are being identified.

Obviously, female hormone supplements and SERMs are not available to men, but bisphosphonates and calcitonins work as well for men as for women. Denosumab would be an ideal double therapy for men because it works directly on the RANK ligand pathway; PTH studies in men are also underway. Testosterone is also being evaluated for male osteoporosis because it apparently has a role in bone homeostasis, at least as a breakdown product.

Male osteoporosis will, therefore, become a significant disease in the eyes of the medical profession, in addition to the lay public. Osteoporosis is now a well-recognized disease among women, although there is still relatively little screening carried out for the disease. Male osteoporosis will have to go through the same slow acceptance process as that for women before a great deal of attention is paid at the general practitioner (GP) level.

14.6. Genetic Influence and Genetic Screening

During the course of writing this book, the human genetic code has been fully evaluated and the first studies have been presented evaluating the genetic profile of individuals with the disease, with the goal of providing genetic screening for the identification of individuals at risk of osteoporosis. Currently, there are studies underway using low BMD as the phenotype, in which the aim is to evaluate the possibility to diagnose the level of risk at birth. It is hoped that by knowing the genetics of the parents, an individual's lifetime risk of fracture could be evaluated. At least theoretically, children in this situation could be encouraged to exercise more regularly and have the appropriate amounts of calcium and other vital nutrients essential for bone development. However, this study presupposes that BMD is the primary cause of osteoporosis. Another study is also investigating the familial links of osteoporosis using twins. Fracture is the endpoint of this study, rather than BMD.

As for most therapeutic fields, we are still a long way off identifying genetic markers for osteoporosis. It is highly unlikely that there is only one or two genes that control bone metabolism, because it is such a complex organ. Therefore, it is going to be some time before genetics has a direct role in identifying people at risk of fracture. We might have better opportunity of identifying receptors or mediators that can be genetically switched on or off, at least in the medium term.

14.7. Differentiation Between those who have a Propensity to Fall and those with Low BMD

One area of differentiation that will become more apparent is the different causes of osteoporotic fracture, i.e. subjects who have low BMD versus those who have a propensity to fall and might have a relatively normal BMD. Currently, this differential diagnosis is not made, but with a variety of treatments and a better understanding of the individual risks, we can expect subjects to be treated very differently.

The currently available strategy for women with a high propensity to fall is to provide them with underwear that has built-in hip protectors. Studies have shown that these are successful in reducing fracture, at least in elderly women in residential care. These protectors, although carefully designed, unfortunately do have to have some bulk in them and, therefore, widen the hips. There are issues with women (or men) at any age having to wear something that is less than flattering. This ensures that compliance in wearing the underwear is $<100\%$, plus they offer no protection during times of undressing and moving around the bathroom.

Other than the hip protectors, very little is being done to evaluate those women with a high propensity to fall, and the cause of the disease in these subjects. It is apparently a cognitive functionality that is primarily affected, which could be successfully treated with somatokines, as discussed in Section 14.4.2. As for all the medications for treatment of osteoporosis, it is unlikely to completely prevent fracture from occurring, but will reduce the frequency of fracture within the population.

14.8. Monotherapy and Multiple Therapeutic Responses

Because the treatment of osteoporosis is still in its infancy, most studies have only looked at the effect of monotherapy. The first studies evaluating the combination of Fosamax® and PTH have now been completed.[5] There seems to be an additive effect, which is only to be expected because bone loss has multiple causes. Therefore, we can expect, in the future, a more rational approach regarding which therapies should be used for each situation, with the potential of a planned switching between therapies at different ages or BMD levels and also the use of combinations of therapies. Preliminary data for PTH and bisphosphonates looks very promising, if PTH is given first, followed by the bisphosphonate (and not the other way round).

14.9. Summary

In the past 6–10 years or so, there has been a rapid development in the diagnosis and treatment of osteoporosis. Because of the nature of the disease, clinical trials are long and complex, and yet there have been some remarkable strides in this therapeutic field although, as already stated, we are still in the infancy of treating this disease. Therefore, we can eagerly anticipate some further major developments in this field, many of which have been described.

The field of clinical trials in osteoporosis is changing and we have probably seen the last placebo-controlled trials started in this area. This provides a series of challenges in its own right, because the regulatory agencies cannot ethically expect subjects to participate in placebo-controlled trials. Therefore, as we have already started to see, the guidelines are being changed and will continue to evolve. The challenges associated with active comparator trials and the numbers of subject that will need to be enrolled will have to be carefully evaluated.

As for the near future, at least one more bisphosphonate will be developed. Zometa® (Novartis International AG, Basel, Switzerland) is the most potent of all the bisphosphonates and has recently gained marketing approval for use in bone metastases. As previously stated, it can be administered annually. This regimen would have the potential to maintain good subject contact, with annual visits to the physician's office by the subject, providing close to 100% compliance.

New administrative routes and formulations for calcitonin are currently being evaluated. Although the calcitonins have not demonstrated efficacy of the same level as the bisphosphonates, the new formulations could hold some advantage in monotherapy or combination therapy.

We can anticipate at least two formulations of PTH to be on the market in the next few years, probably for treatment of osteoporosis. There is one school of thought that PTH will become the therapy of choice for treatment of osteoporosis, and the other therapies discussed will become front-line prevention therapies. However, there is insufficient evidence to support this hypothesis at the time of writing.

One final area of change that might occur is the involvement of the orthopaedic community in the treatment of osteoporosis. Until now, orthopaedic specialists have, generally, not been involved in this disease because it has not required surgery. With the advent of vertebroplasty and kyphoplasty, orthopaedic specialists are beginning to consider this bone disease as one in which they should be involved. They might become the high prescribers of the future.

"Future gazing" is always a challenge. It is difficult to know where the next new development will come from and where it will take us in our understanding of bone physiology and treatment. Even during the writing of this book, the treatment potentials have changed several times. Hopefully, this chapter has provided a near-term perspective on the most probable scenarios, but if one consider the advances in any field within the past 10 years, it becomes very difficult to predict with any certainty the changes in the next 10 years. However, the quote by Huxley at the start of this chapter, "Second to being right in this world is being totally wrong", certainly proved correct for the authors' predictions in the first edition.

The field of osteoporosis, owing to its relative infancy, is still a challenging area, but one in which there is still a good deal of work to be completed by the scientist and physician, with some fascinating discoveries to be made.

References

1. Albright, F. (1947). *Ann Int Med* **27:** Osteoporosis. *Nutr Rev* 1989;47:85–6.
2. Lacey, J.V., Jr, Mink, P.J., Lubin, J.H. *et al*. (2002). Menopausal hormone replacement therapy and risk of ovarian cancer. *JAMA* **288:**334–41.
3. Mclung, M.R., Lewiecki, M., Cohen, S.B. *et al*. (2006). Denosumab in postmenoausal women with low bone mineral density. *New Engl J Med* **354:**821–31.
4. Reid, I.R., Brown, J.P., Burckhardt, P. *et al*. (2002). Intravenous Zoledronic Acid in Postmenopausal Women with Low Bone Mineral Density. *New Engl J Med* **346:**653–61.
5. Rittmaster, R.S., Bolognese, M. and Ettinger, M.P. (2000). Enhancement of bone mass in osteoporotic women with parathyroid hormone followed by alendronate. *J Clin Endocrinol Metab* **85:**2129–34.

Appendix A: Glossary

ACP	Acid phosphatase
ACRP	Association of Clinical Research Professionals (http://www.acrpnet.org/)
ACV%	Annualized coefficient of variation
ADR	Adverse drug reaction
AE	Adverse event
ALP	Alkaline phosphatase
ANOVA	Analysis of variance
AP	Antero-posterior
ASIS	Anterior superior iliac spine
BALP	Bone-specific alkaline phosphatase
BMC	Bone mineral content in grammes
BMD	Bone mineral density in grammes per square centimetre. Calculated as the bone mineral content of a bone divided by the area of the bone and as such it is not true density
BMI	Body-mass index
BMP	Bone morphogenic proteins
BUA	Broadband ultrasound attenuation measured in decibels per megahertz (dB MHz^{-1}); a measure of the attenuation of ultrasound through the heel
CLIA	Clinical Laboratory Improvement Amendments. An accrediting body for laboratories in the USA (http://www.clianet.org)
COREC	Central Office of Research Ethics Committees
CPA	Clinical Pathology Accreditation. An accrediting body for laboratories in the UK (http://www.cpa-uk.demon.co.uk/index.htm)
CPMP	The Committee for Proprietary Medicinal Products. The European Committee responsible for reviewing standards for Good Clinical Practice
CRA	Clinical research associate. A representative of a CRO responsible for liaison between the CRO and sites participating in the trial
CRO	Clinical research organization. A company responsible for administering the trial on behalf of a sponsor
CT	Computed tomography
CTA	Clinical trial agreement on study consent
CTx	C-terminal telopeptide or crosslaps
CUSUM	Cumulative sum charts
CV	Curriculum vitae
CV, CV%	Coefficient of variation
DPyd	Deoxypridinoline
DSMC	Data Safety Monitoring Committee
DXA	Dual-energy X-ray absorptiometry. The principal technique used for measuring BMD
ED	Effective dose measured in millisieverts (mSv). A measure of radiation dose that can be related directly to the risk of the radiation exposure
ELISA	Enzyme-linked immunosorbent assay

EQA	External quality assurance
ESD	Entrance skin dose measured in milligrays (mGy)
ESP	European spine phantom
FDA	US Food and Drug Administration. The US agency responsible for the licensing of pharmaceuticals. A good source of guidelines on the conduct of clinical trials is their website http://www.fda.gov/
GCP	Good clinical practice. The international standard for conduct of clinical trials
GHyL	Galactosyl hydroxylysine
GLP	Good laboratory practice
GP	General practitioner or primary care physician
GREES	The Group for the Respect of Ethics and Excellence in Science
HPLC	High-performance liquid chromatography
HRT	Hormone-replacement therapy or oestrogen-replacement therapy
ICH	International Committee on Harmonization. Produces international standards for the conduct of clinical trials (http://www.ifpma.org/ich1.html)
ICP-MS	Inductively coupled plasma mass spectrometry
ICR	Institute of Clinical Research (http://www/instituteofclinicalresearch.org/)
ICTP	Type I collagen C-terminal telopeptide
IDSC	International DXA Standardization Committee
IND	Investigational new drug
IQC	Internal quality control
IRB/IEC	Institutional Review Board or Independent Ethics Committee responsible for reviewing and approving the ethical basis of the trial at each participating center
IRMA	Immunoradiometric assay
ISCD	International Society for Clinical Densitometry (http://www.iscd.org/)
ISO	International Standards Organization (http://www.iso.ch/iso/en/ISOOnline.openerpage)
MRI	Magnetic resonance imaging
MXA	Morphometric X-ray analysis. A technique from radiographers for lateral scans of the spine using DXA that measures the height of vertebra, allowing the identification of vertebral fractures
NEQAS	UK National External Quality Assessment Schemes (http://www.ukneqas.org.uk/)
NHANES	The National Health and Nutrition Examination Survey; a survey conducted by the National Center for Health Statistics, part of the Centers for Disease Control and Prevention, US Public Health Service. This survey has been designed to collect information about the health and diet of people in the USA. NHANES is unique in that it combines a home interview with health tests. The tests included measuring BMD at the femur (http://www.cdc.gov/nchs/nhanes.htm)
NIBSC	National Institute for Biological Standards and Control (http://www/nibsc.ac.uk)
NIH	National Institutes of Health. An American agency who commission and fund health research. They have produced consensus guidelines on osteoporosis (http://www.nih.gov/)
NIST	National Institute of Standards and Technology (http://www.nist.gov/)

NME	New molecular entity. A new drug under development being tested in a clinical trial
NNT	Number needed to treat (to prevent one fracture)
NOF	National Osteoporosis Foundation (USA) (http://www/nof.org/)
NOS	National Osteoporosis Society (UK) (http://www/nos.org.uk/)
NTx	N-terminal telopeptide-linked crosslink
OA	Osteoarthritis
OHP	Hydroxyproline
PA	Postero-anterior
PDXA	Peripheral DXA measures BMD of forearm or heel
PQCT	Peripheral QCT—usually measures BMD of the forearm
PTH	Parathyroid hormone
Pyd	Pyridinoline
QA	Quality assurance. Review of the quality control information by sampling or audit
QC	Quality control. The ongoing sampling of a process to ensure that the quality of the process is within defined limits. For example, daily phantom measurements on DXA equipment
QCT	Quantitative computed tomography. A method of measuring bone density from a CT scan
QUALEFFO	Quality of Life Questionnaire of the European Foundation of Osteoporosis
QUS	Quantitative ultrasound. A technique that measures the attenuation of ultrasound (BUA) through the heel and the velocity or speed of sound (VOS or SOS). These measures are often combined to calculate the stiffness or Stiffness Index
RCT	Randomized, controlled trial
RIA	Radioimmunoassay
RIQAS	Randox International Quality Assessment Scheme (http://www.randox.com/riqas.htm)
RMS SD	Root mean squared standard deviation
ROC	Receiver operating characteristics
ROI	Region of interest. Area set on image to calculate BMD
SADR	Serious adverse drug reaction
SAE	Serious adverse event
sBMD	Standardized BMD in milligrammes per square centimetre
SCV%	Standardized coefficient of variation
SD	Standard deviation
SEE	Standard error of estimate
SERM	Selective oestrogen receptor modulator
SF-36	A validated health status questionnaire
SMO	Site management organization. A local research organization or hospital that recruits subjects and carries out the clinical trial on behalf of a CRO
SOCRA	Society of Clinical Research Associates (http://www.socra.org/)
SOP	Standard operating procedure
SOS	Speed of sound (ms^{-1}) measured using QUS in the heel
SSC	Study site coordinator
SSE	Residual sum of squares
TALP	Total alkaline phosphatase

tns gene	Tissue nonspecific gene
TRACP	Tartrate resistant acid phosphatase
VOS	Velocity of sound (ms^{-1}) measured using QUS in the heel
WEQAS	Wales External Assessment Scheme
WHO	World Health Organization (http://www.who.int/)

Appendix B: Sample Data from the Example Study Used in this Book

See table that begins overleaf.

The BMD measurements for the AP spine are shown in g/cm^2. The BUA is in dB MHz^{-1}. There are missing values in the BUA data because of equipment failure. Group 1 is the placebo-treated group and Group 2 is the active treatment group. There is additional demographic information, such as age, centre, and smoking habits. Smoking is coded as 1 = nonsmoker and 2 = smoker.

Age	Baseline BMD	BMD 1 year	BMD 2 years	BMD 3 years	BMD 4 years	Baseline BUA	BUA 1 year	BUA 2 years	BUA 3 years	BUA 4 years	Group	Centre	Smoking
50	1.246	1.285	1.285	1.325	1.325						2	1	2
48	0.98	0.966	1.01	1.033	1.015						2	2	2
48	1.068	1.081	1.052	1.076	1.09	115	114	110	108	104	2	1	2
49	0.956	0.964	0.973	0.988	1.013	119	118	117	118	119	2	2	1
52	1.001	1.019	1.051	1.087	1.093	99.1	95	96	98	99	2	1	2
57	0.844	0.866	0.886	0.909	0.912	111.5	107	114	115	112	2	2	2
50	0.825	0.927	0.908	0.948	0.979	118.7	124	122	125	130	2	1	2
51	1.108	1.129	1.128	1.14	1.125	140.5	137	141	140	139	2	2	1
45	1.053	1.069	1.042	1.12	1.098	106.9	107	108	107	106	2	1	1
57	0.861	0.895	0.929	0.977	0.936	110.9	110	108	107	108	2	2	2
51	1.018	1.026	1.038	1.041	1.053	116.5	129	120	116	118	2	1	2
58	0.883	0.904	0.912	0.961	0.972	134	128	124	120	121	2	2	1
51	0.961	0.96	0.991	1.022	1.022	105.2	101				2	1	2
50	0.911	0.983	1.013	1.018	1.015	120.5	121	122	123	118	2	2	2
53	0.911	0.926	0.938	0.942	0.944	132	137	130	125	120	2	1	1
52	0.952	1.02	0.97	1.053	1.031	114	111	113	111	110	2	2	2
52	1.068	1.078	1.134	1.142	1.148	120	120	120	121	118	2	1	2
52	0.856	0.875	0.89	0.912	0.902	104.4	103	102	101	99	2	2	1
51	0.808	0.82	0.793	0.8	0.811	145	125	125	127	129	2	1	2
56	0.927	0.965	0.968	0.985	1.017	125	121	114	110	110	2	2	1
53	1.185	1.235	1.236	1.297	1.338	112.3	112	110	108	111.2	2	1	1
54	1.005	1.062	1.056	1.096	1.084	125	125	124	123	122	2	2	1
56	0.73	0.817	0.818	0.819	0.823	111.1	117	114	113	112	2	1	1
49	0.789	0.814	0.814	0.839	0.839	101	99				2	2	1
50	1.019	1.059	1.068	1.11	1.105	110	107	109	105	113	2	1	2
48	0.974	0.986	0.969	1.006	1.016	114	112	114	114	112	2	2	1
47	0.974	1.028	0.979	1.031	1.021	119	114	116	118	118	2	1	2

59	0.778	0.836	0.844	0.853	0.874	106.5	106	105	105	106	2	2	1
58	0.887	0.95	0.951	0.973	0.959	105.6	110	113	110	115	2	1	1
49	0.854	0.917	0.946	0.916	0.943	103.2	106	106	99	103	2	2	2
56	0.735	0.737	0.711	0.68	0.689	99.7	95	96	92	91.6	1	1	1
50	0.789	0.728	0.697	0.685	0.68	120	113	115	114	101	1	2	2
55	0.797	0.783	0.774	0.767	0.762	107	105	89	94	101	1	1	2
50	0.803	0.793	0.764	0.758	0.722	109.8	104	96	101	106	1	2	2
45	0.822	0.781	0.823	0.84	0.813	110.8	109	101	101	103	1	1	2
48	0.829	0.833	0.839	0.811	0.809	108	105	100	101	90	1	2	1
53	0.83	0.825	0.81	0.773	0.854	115	110	112	106	109	1	1	1
51	0.861	0.861	0.835	0.858	0.876	113.2	113	110	107	91	1	2	2
54	0.884	0.856	0.836	0.845	0.826	107	106	101	100	109	1	1	1
53	0.905	0.921	0.922	0.955	0.931	109	108	104	102	101	1	2	2
50	0.907	0.864	0.877	0.867	0.888	140	135	106	101	101	1	1	1
57	0.935	0.923	0.912	0.911	0.908	112	109	114	112	110	1	2	1
53	0.913	0.898	0.895	0.887	0.924	106.8	105	101	90	103	1	1	2
50	0.926	0.941	0.876	0.861	0.855	111.1	109	113	110	111	1	2	1
52	0.943	0.912	0.897	0.941	0.945	110.8	113	109	112	116	1	1	2
55	0.967	0.933	0.94	0.943	0.927						1	2	2
58	0.976	0.95	0.965	0.963	0.955	109.9	105	106	105	99	1	1	1
49	0.975	0.969	0.969	0.973	1.013	135	132	129	131	123	1	2	2
55	0.952	0.94	0.926	0.938	0.917	103.4	97	100	97	89.9	1	1	1
47	0.988	0.966	0.895	0.901	0.889	110	107	104	97	103	1	2	1
50	0.996	0.93	0.921	0.918	0.936	124.8	109	119	115	114.9	1	1	2
51	1.004	1.017	0.969	1.002	0.969	120.5	123	118	115	117	1	2	1
47	1.01	0.972	0.948	0.923	0.886	110.1	109	110	109	104	1	1	2
48	1.036	1.036	1	1.016	0.974	114.3	111	109	100	108	1	2	2
53	1.056	1.109	1.042	1.071	1.06	110	102	104	103	101	1	1	1
49	1.06	1.025	0.982	0.97	0.927	106	105	97	93	98	1	2	1
49	1.068	1.017	0.978	1.022	1.002	119	116	121	115	111	1	1	1
54	1.08	1.069	1.07	1.085	1.099	125.8	126	118	104	123	1	2	1
52	1.116	1.073	1.051	1.045	1.091	120	114	114			1	1	2
54	1.163	1.169	1.141	1.102	1.102						1	2	2

Index