

Guidelines for preclinical evaluation and clinical trials in osteoporosis



World Health Organization
Geneva
1998

WHO Library Cataloguing in Publication Data

Guidelines for preclinical evaluation and clinical trials in osteoporosis.

1. Osteoporosis – drug therapy
2. Legislation, Drug
3. Clinical trials – methods
4. Evidence-based medicine – methods
5. Guidelines

ISBN 92 4 154522 4

(NLM classification: WE 250)

The World Health Organization welcomes requests for permission to reproduce or translate its publications, in part or in full. Applications and enquiries should be addressed to the Office of Publications, World Health Organization, Geneva, Switzerland, which will be glad to provide the latest information on any changes made to the text, plans for new editions, and reprints and translations already available.

© **World Health Organization 1998**

Publications of the World Health Organization enjoy copyright protection in accordance with the provisions of Protocol 2 of the Universal Copyright Convention. All rights reserved.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the World Health Organization concerning the legal status of any country, territory, city or area or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

TYPESET AND PRINTED IN SWITZERLAND
98/11991 – Strategic/PCL – 15000

Contents

1. Introduction	1
2. Background	3
3. Definition of osteoporosis and related terms	5
3.1 Conceptual definition of osteoporosis	5
3.2 Operational definition of osteoporosis	5
3.3 Intervention thresholds	6
3.4 Intervention categories	6
4. Preclinical studies	7
4.1 Introduction and aims	7
4.2 Studies <i>in vitro</i>	8
4.2.1 Studies on bone resorption	8
4.2.2 Studies on bone formation	9
4.3 Animal models	9
4.3.1 Animal species	9
4.3.2 Experimental techniques	10
4.3.3 End-points	10
4.3.4 Statistics	11
4.3.5 Ethical considerations	11
4.4 Specific recommendations for animal models	11
4.4.1 The adult oophorectomized rat	11
4.4.2 Other rat models	12
4.4.3 Mouse models	12
4.4.4 Non-rodent models	12
4.5 Study design	13
4.5.1 Intervention schedule	13
4.5.2 Dosage	13
4.5.3 Duration of treatment	14

4.6	End-points	14
4.6.1	Assessment of bone mass and bone mineral	14
4.6.2	Bone histology and microarchitecture	14
4.6.3	Biomechanical testing of bone strength	15
4.6.4	Biochemical indices of skeletal metabolism	17
4.6.5	Fracture healing	18
4.7	Pharmacokinetics	18
4.8	Studies of long-term bone safety	19
4.9	Other toxicological studies	19
4.10	Minimal preclinical requirements for clinical evaluation in phase III	20
5.	Clinical trials in osteoporosis	21
5.1	Introduction	21
5.2	Bone mass and fracture	22
5.3	Techniques used in the assessment of osteoporosis and response to interventions	23
5.3.1	Single-energy absorptiometry	24
5.3.2	Dual-energy absorptiometry	24
5.3.3	Other techniques	25
5.3.4	Limitations of techniques	25
5.3.5	Sites of bone mineral measurements	26
5.4	Fracture assessment	27
5.4.1	Fractures as entry criteria and end-points	27
5.4.2	Vertebral fractures	27
5.5	Height	29
5.6	Quality of life	29
5.7	Other measurements	30
5.7.1	Bone biopsy	30
5.7.2	Biochemical indices of bone turnover	31
5.8	General principles of trial design	31
5.8.1	Population to be studied	32
5.8.2	Treatment regimen	33
5.8.3	Study design	33
5.8.4	Duration of studies	34
5.8.5	Sample size	34
5.8.6	Frequency of measurement	35
5.8.7	Safety	36
5.9	Non-pharmacological interventions	36
5.10	Good clinical practice	36

6. Phase I and phase II studies	38
6.1 Phase I studies	38
6.2 Phase II studies	38
6.3 Basic conditions	39
7. Phase III studies in severe (established) osteoporosis	41
7.1 Introduction	41
7.2 Population studied	41
7.3 Basic conditions	41
7.4 End-points	42
7.4.1 Reduction in the rate of fracture	42
7.4.2 Increase in bone mass	43
7.4.3 Prevention of bone loss or reduction in the rate of bone loss	43
7.5 Study design	44
7.5.1 Duration	44
7.5.2 Site of measurement	44
7.6 Other types of studies in severe (established) osteoporosis	44
7.6.1 Reduction of extraskkeletal risks in osteoporosis	44
7.6.2 Symptomatic end-points	44
7.7 Safety and clinical pharmacology	44
8. Phase III studies in osteoporosis without fragility fractures	46
8.1 Introduction	46
8.2 Population studied	46
8.3 Basic conditions	46
8.4 End-points	47
8.5 Study design	47
8.6 Safety and clinical pharmacology	47
9. Phase III studies in low bone mass (osteopenia)	49
9.1 Introduction	49
9.2 Population studied	49
9.3 Basic conditions	50
9.4 End-points	50
9.5 Study design	50
9.6 Duration of studies, including interruption of treatment	51
9.7 Safety	51
10. Phase III clinical trials in subjects with normal bone mass	52
10.1 Introduction	52
10.2 Population studied	52

10.3 Basic conditions	53
10.4 End-points	53
10.5 Study design	53
10.6 Duration of studies, including interruption of treatment	53
10.7 Safety	54
11. Phase IV studies	55
11.1 Introduction	55
11.2 Aims	55
11.2.1 Efficacy with continued treatment	55
11.2.2 Efficacy after discontinuation of treatment	55
11.2.3 Need for retreatment	56
11.2.4 Safety and other end-points	56
11.3 Type of study	56
11.3.1 Uncontrolled trials	56
11.3.2 Cohort study	57
11.3.3 Case-control study	57
11.3.4 Randomized trials	57
11.4 Sample size and statistical considerations	57
12. Glossary of terms	58
References	65
Acknowledgements	67
Annex. Recommendation for future guideline development	68

1.

Introduction

Osteoporosis and associated fractures increase markedly with age and are a major cause of mortality and morbidity — and thus of medical expense — throughout the world. The frequency of osteoporotic fractures is much higher in women than in men because of their generally greater life-expectancy and because of postmenopausal hormonal changes, and it has been estimated that the number of women over the age of 65 years will increase from 188 million in 1990 to 325 million in 2015.

Because of the increasing importance of osteoporosis, consensus is needed on the principles and methods applied in clinical trials of the efficacy and safety of drug treatments for the disease. A number of scientific groups and drug regulatory authorities have produced guideline documents and position papers on the criteria to be applied in investigations and evaluations of new treatments. The focus of these documents has varied, but none has provided a comprehensive statement of guiding principles for the design, implementation, and interpretation of either preclinical testing or clinical trials in osteoporosis.

To address this need, WHO's Division of Drug Management and Policies established a working group in 1995; the group included drug regulators and scientists and was charged with the development of international guidelines. It was recognized that such guidelines should:

- comprise evidence-based principles rather than detailed prescriptions for the conduct of studies;
- be sufficiently broad in scope to encompass the various types of osteoporosis, and their prevention and treatment (including non-pharmacological interventions);
- be relevant to all those who are involved in the development and evaluation of interventions, e.g. the pharmaceutical industry, drug regulatory authorities, and individual scientists.

The group identified a need for studies in different populations (men, children, premenopausal women, etc.). A recommendation for the development of further guidelines is given in the annex to this publication.

In developing the present guidelines as a practical tool for use by WHO Member States, every effort has been made to ensure their compatibility with

existing national and other provisions and to provide a framework on which national guidelines may subsequently be based. The guidelines have been reviewed by internationally accredited experts, by drug regulatory authorities, and by the pharmaceutical industry.

2.

Background

Osteoporosis is a major health-care problem of ageing communities yet, until recently, it has received relatively little attention. Its clinical significance lies in the fractures that arise, most commonly of the forearm, the vertebral bodies, and the hip. Of these, the most serious is hip fracture because of the high morbidity and associated mortality — and consequent high cost to health services (1). Vertebral, forearm, and upper humeral fractures also cause considerable morbidity and, since they tend to occur in younger individuals than hip fractures, have significant long-term impact on quality of life. Fractures at many other sites are also more frequent in osteoporosis than in health, but these are less important.

In most countries the incidence of osteoporosis is about 2–4 times higher in women than in men; this is related to the loss of bone that occurs as a result of estrogen deficiency after the menopause. In addition, women generally live longer than men, so that the frequency of osteoporotic fractures in most communities is at least 3 times higher in women than in men.

The absolute risk of osteoporosis varies between countries, and is highest in North America and western European countries. At the age of 50 years the remaining life-time risk of hip fracture in white women from the United States is 17.5%, and similar levels of risk are found in most Western European countries. Estimates for other types of osteoporotic fracture, such as vertebral and Colles fracture, are nearly as high, so that the combined total risk for fracture due to osteoporosis is about 35%. These figures are conservative since they take account only of the vertebral fractures that come to clinical attention and no account of osteoporotic fractures at other sites. The risks for osteoporotic fractures in both men and women are lower in other geographical areas, including Africa and Asia, but are likely to increase markedly in the future.

Between 1.3 and 1.6 million hip fractures were estimated to have occurred worldwide in 1990. These numbers will rise significantly as the number of elderly people increases because of continued improvements in life-expectancy. Between 1990 and 2025 the number of men aged 50 years or more will increase by 150% in Europe and by more than 200% in all other regions. The greatest increments will occur in Africa, Asia, and Latin America. For women the predicted increases are 131–140% in Europe, 183% in North America, and

more than 200% in all other regions. Asia will have the highest absolute increase in the elderly because of the large present population of the region. These demographic changes suggest that, by the year 2025, the number of hip fractures will have increased to 4 million per year worldwide, with the largest number occurring in Asia.

It is very likely that these predictions represent an underestimate since the age- and sex-specific incidence of osteoporotic fractures is increasing in both men and women, particularly in Asia. With even modest assumptions concerning the secular trend, the number of hip fractures in 2025 could be as high as 6 million, so that osteoporosis has a substantial and ever-increasing economic significance.

Treatment and prevention of osteoporosis — and even attitudes to the problem — vary widely from country to country, despite both the importance of the disease and improved knowledge of its causes. This is largely the result of a lack of awareness among medical practitioners and health ministries of the objectives and effects of interventions in osteoporosis. The rapidity with which new treatments and technologies have developed has proved a further obstacle to a cohesive and rational approach to the problem. There is thus a pressing need for consensus, based on the most scientifically appropriate information, concerning the manner of assessing the efficacy of therapeutic regimens.

Several national and international guidelines and position papers have been developed in recent years (2–8). While these have differed in their emphasis, they have focused principally on women with postmenopausal osteoporosis and have not considered other forms of osteoporosis. Differences in substance relate to uncertainties in the relationship between bone mineral density and fracture outcome. Considerable recent experience in osteoporosis studies using both bone mineral assessments and fractures has helped to clarify this relationship. Use of animal models of bone strength, rather than of osteoporosis, has also resolved some of the uncertainties (9), allowing more cohesive views and strategies to be developed.

3.

Definition of osteoporosis and related terms

3.1 Conceptual definition of osteoporosis

Various definitions of osteoporosis have been offered to describe the outcome of events (fragility fractures), the process giving rise to porous bones, or the resultant diminution in bone mass. The following definition (10) is now generally accepted:

“A disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk.”

While recognizing the multifactorial nature of the events that give rise to fractures, this definition also embraces the concept that low bone mineral density is an important component of fracture risk. Since bone mineral content or density can be determined with precision and accuracy, its measurement forms the basis for an operational definition of osteoporosis.

3.2 Operational definition of osteoporosis

The relationship between bone mineral density and fracture risk is continuous, and an estimate of bone mineral density therefore provides an effective assessment of fracture risk. It is thus possible to choose a value for bone mineral density that defines the presence of osteoporosis.

Bone mineral content or density in young healthy women (“peak bone mass”) is normally distributed, irrespective of the measurement technique used. By virtue of this normal distribution, bone density values in individuals may be expressed in relation to a reference population in standard deviation (SD) units; this reduces the problems associated with calibration differences between instruments. Use of standard deviation units in relation to the young healthy population is referred to as the t-score, where the mean is ascribed a value of zero.

The World Health Organization has proposed two diagnostic thresholds of bone mineral density for Caucasian women, based on the distribution of skeletal mass in young healthy individuals; these thresholds permit the establishment of four general diagnostic categories (1):

- *Normal.* A value for bone mineral within 1 SD of the young adult reference mean (t-score ≥ -1.0).

- *Low bone mass (osteopenia).* A value for bone mineral more than 1 SD below the young adult mean but less than 2.5 SD below this value ($-1.0 > \text{t-score} > -2.5$).
- *Osteoporosis.* A value for bone mineral 2.5 SD or more below the young adult mean ($\text{t-score} \leq -2.5$).
- *Severe (established) osteoporosis.* A value for bone mineral 2.5 SD or more below the young adult mean in the presence of one or more fragility fractures.

Suitable diagnostic cut-off values for non-Caucasian women and for men are less secure. It has been suggested (1) that a similar absolute value for bone mineral density to that used in women can be taken as a cut-off point for the diagnosis of osteoporosis in men — that is, a value 2.5 SD below the average for adult premenopausal women. It should be recognized that cut-off values are arbitrary and may differ according to sites measured, age, and type of equipment.

3.3 Intervention thresholds

An intervention threshold depends upon absolute risk and upon the risks and benefits of a particular treatment, and should be distinguished from a diagnostic threshold.

3.4 Intervention categories

Four intervention categories are defined below. Although such definitions and thresholds are inevitably arbitrary, they enable the objectives and categories of treatment to be more closely defined.

- *Prevention:* defined as an intervention in individuals with normal skeletal status or as a global intervention in the community. Examples include regimens that modulate peak bone mass and contribute to its preservation, and agents that prevent bone loss or preserve skeletal architecture in individuals at risk.
- *Prevention of osteoporosis in patients with osteopenia:* defined as intervention in patients with low bone mass (osteopenia), but without osteoporosis. Its aims are to prevent osteoporosis and reduce the probability of future fractures.
- *Treatment of osteoporosis:* defined as intervention in patients with osteoporosis. Its aims are to reduce the probability of fractures.
- *Treatment of severe (established) osteoporosis:* defined as intervention in patients who have sustained one or more fragility fractures associated with osteoporosis. Its aims are to reduce both the risk of further fractures and the morbidity associated with fracture.

Alternative intervention categories may be based on stratification of other risk factors for fractures, for example biochemical markers of bone turnover, hip axis length, family history of fractures, and the risk of falls.

4.

Preclinical studies

4.1 Introduction and aims

Preclinical evaluation is an essential component of the development of pharmaceutical interventions for osteoporosis. A similar systematic approach can also be used, at least in part, for interventions of a nutritional or mechanical nature. The general aims of preclinical evaluation fall into two categories — those relating to the general description, metabolism, pharmacodynamics, safety, and toxicity of any new agent, and those relating specifically to bone metabolism.

The aims of preclinical studies specifically relating to osteoporosis are:

- (a) To establish the relationship between the effects of an intervention on bone mass and on bone strength. In particular, studies should identify whether the use of an agent capable of restoring or preserving bone mass is associated with the formation of new bone tissue with normal architecture, and — particularly — with a commensurate increase in bone strength.
- (b) To elucidate the mechanism of action of a pharmaceutical agent and thus provide the rationale for its use in humans.
- (c) To demonstrate effects in animal studies of osteoporosis.
- (d) To establish the effects of long-term exposure to the agent on the quality of skeletal tissue.
- (e) To examine the effects of intervention on fracture repair.

The results of preclinical evaluation will determine whether a new agent should be tested in humans and, if so, the population and the end-points required for phase I, phase II, and phase III of the evaluation process.

Consensus is still lacking within the scientific community and drug regulatory authorities on a completely satisfactory experimental model (or models) for human osteoporosis because any existing models that display all the typical features of osteoporosis still fail to display the **same** pattern of spontaneous fragility fractures. However, several animal models of osteoporosis as defined in section 3.1 are available or under development.

Features of osteoporosis include low bone mass, microarchitectural deterioration of bone tissue, and increased bone fragility. Animal models are

available for estrogen-deficiency osteoporosis and for other less common forms of the disorder (e.g. due to corticosteroids or immobilization); indeed, depending on the species and type of intervention, animal models can reproduce more or less all the characteristics of osteoporosis.

The primary objective of preclinical testing is to investigate the pharmacodynamic effects and, in particular, the biomechanical consequences of a putative therapeutic intervention. Precise and accurate techniques, often comparable to those used in clinical evaluation, are available for assessing drug-related effects at clinically appropriate sites. These techniques include the measurement of bone mass and bone mineral density, evaluation of microarchitecture, and determination of static and dynamic histomorphometric variables, biochemical indices of skeletal turnover, and, most importantly, bone strength under well controlled conditions.

For some therapeutic interventions, there is increasing evidence that the results of preclinical studies can predict whether changes in bone mass will be associated with modifications in bone fragility — and therefore in fracture rate — in osteoporotic patients. For other therapeutic agents, however, the predictive value of these models is less secure.

The crucial importance of preclinical evaluation is underscored by the limited number of validated non-invasive techniques for assessing bone strength in humans. Assessment is best achieved by measurement of fracture rates. However, evaluation of fracture rates in phase III studies poses a number of difficulties relating to the multifactorial nature of fracture, the large sample sizes required (particularly for hip fracture), and uncertainties concerning the definition of incident vertebral fractures (i.e. fractures occurring during the study). It is therefore recommended that preclinical and clinical programmes for assessing the efficacy of therapeutic interventions be complementary. A comprehensive and adequate preclinical programme is therefore expected to reduce the requirements of clinical studies. Factors that may influence the clinical programme include the appropriate choice of preclinical models, and also the findings relating to the mechanism of action and general properties of the agent tested (see section 5.2).

4.2 Studies *in vitro*

Studies *in vitro* should complement a programme of *in-vivo* investigations with the aim of identifying or specifying the mode of action of the test intervention. There are several *in-vitro* models that can be used to assess the effects of interventions on bone resorption and, to a lesser extent, on bone formation.

4.2.1 Studies on bone resorption

Bone tissue culture systems using fetal mouse calvaria or long bones can provide information on the effects and dose characteristics of agents that inhibit bone resorption. These systems provide the opportunity to establish the

correspondence between characteristics determined *in vivo* and *in vitro*. They also allow assessment of the effects of the way in which interventions affect systemic and local factors that directly influence bone resorption. At the cellular level, appropriate experimental systems are also available that permit the effects of agents on osteoclast formation and activity to be determined.

4.2.2 Studies on bone formation

Tissue and cell culture systems can be used to explore the mechanism of action of agents shown to stimulate bone formation *in vivo*. Various transformed or non-transformed animal or human osteoblast-like cells can be used to test effects of the agent on proliferation, differentiation, and cell phenotype, and to identify cellular transduction signalling mechanisms. Generally speaking, *in-vitro* studies of bone formation are less well characterized in terms of their applicability to *in-vivo* situations than is the case for agents acting primarily on bone resorption. Moreover, there are no widely validated *in-vitro* systems for examining the effects of therapeutic agents on formation and mineralization of lamellar bone.

4.3 Animal models

Animal models should be appropriate for the intended clinical application. For drugs intended to prevent or reverse postmenopausal osteoporosis, a model should have at least the following characteristics:

- increased bone turnover after oophorectomy;
- bone loss leading to an osteoporotic state that is not spontaneously reversible;
- bone loss affecting both cortical and cancellous tissue at relevant skeletal sites such as the vertebral body, femoral neck, metaphyses and diaphyses of long bones; and
- increased skeletal fragility.

4.3.1 Animal species

Preclinical studies may be carried out in various species including mice, rats, rabbits, mini-pigs, sheep, dogs, and monkeys. The choice of species depends partly on the histodynamic characteristics of basal bone remodelling patterns and their comparability to those observed in humans, and partly on their responsivity to calciotropic and osteotropic stimuli or to the interventions, and the extent to which they are predictive of the response in humans. In addition, for some forms of osteoporosis, such as those caused by corticosteroids and immobilization, satisfactory models are more species-specific than is the case for hypogonadal states. Other factors that determine the choice of model include the consistency and rapidity of bone loss at various skeletal sites,

availability of animals and the ease and safety of their handling, and cultural sensitivities, which may preclude the use of particular species in some countries.

4.3.2 Experimental techniques

Various techniques may be used to induce changes in bone mass in experimental animals. These include oophorectomy, orchidectomy, corticosteroid use, manipulation of dietary calcium intake in lactating and non-lactating animals, immobilization by using plaster-casts, hemicordectomy, and sciatic denervation.

Use of these experimental animal models for documentation of drug effects is an important component of the initial assessment of efficacy and safety. For example, the results of several therapeutic interventions in the oophorectomized adult rat mimic observations made in postmenopausal osteoporotic women.

Newer models of osteoporosis, though less well characterized, are worthy of consideration. Some rodent strains, for instance, have low peak bone mass or display spontaneous signs of osteoporosis associated with accelerated ageing.

Apart from models of human osteoporosis, several other experimental *in-vivo* systems are appropriate for selecting molecules of potential interest and for dose-finding. This has been particularly well validated in establishing the potency of anti-resorbing drugs and their propensity to inhibit the mineralization of bone. Such models are also suitable for determining the potency of therapeutic interventions in animals treated with agents that enhance bone resorption.

Intercurrent fractures are common in patients with osteoporosis, and their repair may be affected by interventions. For this reason experimental models of fracture healing are considered to be important for preclinical evaluation.

4.3.3 End-points

A series of relevant end-points can be evaluated using either invasive or non-invasive techniques. These techniques include histomorphometry, *ex-vivo* culture of bone-forming or bone-resorbing cells, the chemistry and biochemistry of bone tissue, assessment of biochemical indices of skeletal turnover in blood and urine, assessment of metabolic balance of calcium combined with radioactive calcium kinetics, radiogrammetry of bone radiographs, neutron activation for whole-body calcium, single- and dual-energy absorptiometry, quantitative computed tomography, and dual-energy X-ray absorptiometry.

The change in resistance to mechanical deformation induced by treatment is a most important end-point and a requirement in the preclinical development programme (see section 4.8).

The choice of end-point to be measured will depend upon the stage of drug development. For example, screening tests designed to determine the anti-

resorptive potency of new compounds require different end-points from investigations of the effects of intervention on the number and activity of bone-resorbing cells. The distinction between end-points should be considered in the evaluation of stimulators of bone formation.

4.3.4 Statistics

The statistical approach to preclinical studies should be as rigorous as that used in clinical trials (see section 5.8.5). The number of animals in experimental groups should take account of the expected difference between treatments as well as the precision of the techniques selected.

4.3.5 Ethical considerations

Experimental procedure should comply with the regulations of local and national ethics committees, which may differ from country to country.

4.4 Specific recommendations for animal models

4.4.1 The adult oophorectomized rat

For the evaluation of agents intended to prevent and/or treat postmenopausal osteoporosis, the adult oophorectomized rat is a convenient and reliable experimental model. Like humans, rats have cancellous bone that undergoes remodelling once longitudinal growth has essentially ceased. In adult rats, oophorectomy is followed by an increase in bone turnover associated with bone loss and a permanent deficit of bone mass at several skeletal sites, including the vertebral bodies, the proximal femur, and the metaphyses of long bones such as the distal femur and proximal tibia. The microarchitectural alterations in cancellous bone are similar to those observed in postmenopausal and age-dependent osteoporosis. They include osteoclastic perforation of trabecular elements, which produces architectural discontinuities of cancellous bone and a consequent discrepancy between the amount of bone lost and the decrease in bone strength following oophorectomy.

Mechanical competence can be tested at several skeletal sites where bone mass can be measured either *in vivo* or *ex vivo*. In addition to the diaphyses of long bones, measurement sites include vertebral bodies and the metaphyses of several long bones. However, the differences in anatomy and structure of the proximal femur in rats and humans make this site problematic.

Many anti-osteoporosis interventions have been shown to exert similar effects on bone mass in rats to those observed in postmenopausal women. Some have also been shown to have similar effects on bone strength in adult oophorectomized rats and in humans.

The adult oophorectomized rat model does, however, suffer from a number of limitations that must be taken into account both in the design of

studies and in the interpretation of results obtained. Control of food intake (“paired feeding”) of experimental groups may be appropriate, since oophorectomy in the rat increases the intake of nutrients and may thereby influence skeletal metabolism. It should be recognized that cortical bone remodelling rates are spontaneously low under normal dietary conditions. The loading of parts of the rat skeleton differs from that in humans, and this is particularly relevant to the vertebral bodies. In adult rats the closure of epiphyseal plates is incomplete at some skeletal sites. The fact that adult rats maintain some degree of bone modelling should not necessarily be considered to be a drawback of the model, since enlargement of the external envelope is also observed at several sites of the adult human skeleton. The process by which bones enlarge, however, may differ between the species.

4.4.2 Other rat models

Adult or aged rats can be used as models of male hypogonadism, osteoporosis induced by nutritional deficiencies such as calcium deprivation, and immobilization osteoporosis. However, no rat model of corticosteroid-induced osteoporosis has yet been validated.

Young growing rats are useful in the study of factors that influence accrual of bone and hence affect peak bone mass. They may also be appropriate in studies examining the potential for interventions to affect bone growth in adults. Young rats may also provide information on the short-term effect of interventions on bone resorption, on calcium fluxes and balance, and on secretion rates and plasma levels of calciotropic hormones.

4.4.3 Mouse models

Adult oophorectomized mice can be used in models of postmenopausal osteoporosis, and models using congenitally osteoporotic mice are also available. Transgenic mice have been used in studies of the physiological or pathophysiological action of several agents on bone metabolism and may prove useful in elucidating the mechanism of action of interventions.

4.4.4 Non-rodent models

To complete the information gained from investigations in rodent species, particularly oophorectomized rats, a non-rodent model should additionally be used. This model should not only confirm the data obtained in rodents, but also provide information not available in those species. An important criterion for selection of such a model should be more extensive cortical bone remodelling than occurs in rodents. The model should be validated by the demonstration that osteoporosis, as defined in section 3.1, can be induced at relevant skeletal sites by manoeuvres that mimic the pathophysiological condition of interest, such as hypogonadism, calcium deprivation, and immobilization. Of the many

species available (primates, mini-pigs, sheep, ferrets, and dogs), the non-human primate shares many relevant features with humans, including menstrual cycles in some species, and reproductive hormone patterns similar to those observed in premenopausal women. These animals also develop high bone turnover in osteoporosis, particularly at the spine following oophorectomy.

In rabbits, dogs, and sheep, corticosteroids have been shown to induce changes in indices of bone formation, with or without alterations in indices of bone resorption, that are similar to the characteristics of corticosteroid-induced osteoporosis in humans. Despite the complexity, lack of homogeneity, and controversial pathophysiology of human glucocorticoid-induced osteoporosis, these species are likely to prove useful in the assessment of agents being considered for prevention or treatment of the condition.

4.5 Study design

4.5.1 Intervention schedule

Interventions designed for prevention are normally started immediately after an osteoporosis-inducing manoeuvre. In contrast, agents intended to treat the established disorder should not be administered until the osteoporotic state has been expressed.

The treatment schedule should simulate that proposed for use in the clinical development stage. For example, agents intended for intermittent, combined, sequential, or cyclic administration should be tested under comparable conditions as far as possible. Moreover, since the toxicity of a drug may vary markedly according to the route and mode of its administration, this should be taken into account in preclinical studies. Ideally, the preclinical evaluation should establish a rationale for the type of regimen to be used later in humans, since this may reduce the requirements of the clinical evaluation programme.

4.5.2 Dosage

Initial studies should employ full dose ranges and should consist of relatively short-term experiments using appropriate screening models with end-points such as changes in bone mass, determined by a non-invasive technique such as dual-energy absorptiometry or quantitative computed tomography. It is appropriate to include pharmacodynamic measurements such as calcium balance or biochemical estimates of bone resorption and bone formation.

For long-term studies, end-points should include bone mass, skeletal architecture, histomorphometric microstructural analysis, and determinants of bone strength. Three dose levels are recommended: the first should be the dose expected to elicit a minimal (or half-maximal) response; the second the dose that is thought to be optimally effective; and the third — for safety reasons — a dose that is significantly higher than the optimum effective dose.

4.5.3 Duration of treatment

The duration of treatment should be sufficient to take account of the relatively long turnover time for cancellous and cortical bone. However, there is no reliable experimental evidence that permits the exposure duration to be specified for all species or all interventions. Nevertheless, the lifespan of bone-remodelling units is shorter in small animals such as the rat than in larger species; animal studies should therefore be longer, the larger the species studied. It is arbitrarily proposed that, in monkeys, the agent should be applied for a period at least half of that selected for phase III evaluation, and in rats for a period four times shorter. An alternative approach in setting the duration of animal studies is to establish that appropriate steady-state dynamics — in terms of treatment-induced changes in skeletal mass — have been achieved in the species concerned.

4.6 End-points

4.6.1 Assessment of bone mass and bone mineral

Many techniques can be used either *in vivo* or *ex vivo* to assess bone mass and bone mineral density in both small and large animals. Non-invasive techniques include single- and dual-photon absorptiometry (SPA, DPA), single- or dual-energy X-ray absorptiometry (SXA, DXA), and peripheral quantitative computed tomography (pQCT). Whatever technique is selected, it should be validated in terms of both precision and accuracy for the animal species chosen for the evaluation. Sites selected for assessment should be not only clinically relevant but also appropriate for comparisons to be made between bone mineral, bone histomorphometry, and bone strength (see sections 4.7 and 4.8).

It is important to bear in mind that different techniques yield different types of information. Using pQCT it is possible to analyse cancellous and cortical mineral separately and also to assess cortical dimensions. Both DXA and pQCT can be used to monitor drug effects on bone shape and geometry, and to relate drug-induced changes to histomorphometric and biochemical tests.

4.6.2 Bone histology and microarchitecture

Bone histomorphometry is an essential preclinical tool for elucidating the mechanism of action of new therapeutic agents on bone remodelling and for assessing certain important components of bone quality in animals exposed to long-term treatments. Histologically, several components of bone quality can be assessed, including microarchitecture, texture of bone matrix (reflected in the presence of lamellar or woven bone), and the adequacy of mineralization. Abnormalities in any of these factors can compromise bone strength.

Histomorphometry is normally performed on undecalcified sections so that the effects of agents on cancellous and cortical bone can be assessed at both

cellular and tissue levels. The effects on bone remodelling should be evaluated at both cortical and cancellous sites. In young and adult rats, both modelling and remodelling may be present, and it is important to distinguish between the two processes. Modelling is the process responsible for the growth and shaping of bone, in which there is no local coupling of bone formation and bone resorption. Bone remodelling, however, is characterized by spatial and temporal coupling of bone formation and resorption.

Tetracycline and other fluorochromes are taken up in bone at the site of calcification, and the use of double labelling can provide dynamic information on bone mineralization and bone formation and allows the dimension of time to be introduced into quantitative analyses.

The choice of skeletal site depends upon the animal species, but the tibial and femoral metaphyses, the femur, the vertebrae, and the iliac crest are all useful.

Samples for histomorphometry should be handled in a laboratory with expertise in embedding, sectioning, and staining, and in the use of digitizing image-analysis systems.

In the rat, measurements of cancellous bone can be made at the proximal tibial metaphysis or distal femur at a distance of 1–4 mm from the growth plate. In primates, pigs, or sheep, measurements on vertebrae are appropriate, and the iliac crest is also a useful site in sheep and dogs. Cortical bone measurements can be made at the tibia or femoral shaft in rats or on the ribs in dogs. Measurements of both cortical and cancellous bone can be made at the femoral neck. With some interventions it may be important to assess the epiphyseal and metaphyseal cortical regions separately.

Assessments in cancellous bone should include bone volume, various static parameters of bone formation (osteoid surface, volume and width, osteoblast surface) and bone resorption (eroded surface, osteoclast surface, osteoclast number), microarchitecture, and connectivity parameters (trabecular number, width and separation, node number, free ends or fractal analysis, and dynamic indices including mineral apposition rate (from interlabelling width), mineralizing surfaces (double-labelled surfaces plus half the single-labelled surfaces), bone formation rate, activation frequency). Variables should be reported according to internationally recognized guidelines (11).

On cortical bone, measurements should include bone area, porosity, and width, mineral apposition rate, and eroded surface. Histological specimens may also be suitable for microradiographic investigations and in some instances for analytical measurements of biochemical or mineral constituents.

4.6.3 Biomechanical testing of bone strength

Several tests can be used for the evaluation of the structural or material properties of bones from large and small animals. Care should be taken to preserve and test the specimens in standardized conditions.

Methods of testing

Changes in the mechanical properties of the diaphyses of the long bones can be tested using three or four point-bending or torsion tests. The advantage of point-bending tests is that it is possible to choose the precise location on the long bone where the mechanical properties will be tested. Moreover, data produced using this type of test generally show minimal dispersion and yield means with relatively low standard deviations and standard errors. It is therefore possible to achieve sufficient statistical power with fewer animals. The advantage of a torsion test, on the other hand, is that it permits the bone to break at its weakest point; thus, the site at which the mechanical properties are measured is not predetermined, and the bone behaves more in accordance with a pathophysiological state. Compression, or a combination of bending and compression tests (producing a load–deflection curve), can be used to assess the mechanical resistance of the vertebral bodies and other sites. The structural and material properties of bone can be determined directly by analysing the load–deformation relationship (e.g. stress–strain curve) provided that the loaded surface can be accurately measured. This analysis provides information on the following properties:

- *Stiffness.* This is a measure of the ability of bone to accept and transmit mechanical loads. It is dependent on the cross-sectional diameter of the bone such that its value is proportional to the fourth power of the radius.
- *Failure load* (ultimate or breaking strength). This measurement corresponds to the maximal load that bone can sustain. For the bending test, maximal load is usually expressed as the bending moment.
- *Energy absorption.* This is the measure of bone toughness and corresponds to the area under the load–deflection curve up to the maximal load.

Skeletal sites

The relative contribution of cancellous and cortical tissue to overall bone strength varies markedly with skeletal site. The two types of bone tissue can be differentially influenced, not only by the nature of the technique used to induce osteoporosis, but also by the agent being tested. Tests should therefore be performed at several skeletal sites, of which the vertebral bodies and the mid-shaft regions of long bones are among the most important. Testing at other sites may be appropriate, particularly sites where there has been densitometric and/or histomorphometric evaluation. At spinal sites it may also be appropriate to undertake tests both on whole vertebrae and on cores of cancellous bone with or without removal of the end-plates.

The results of biomechanical testing should take into account changes in bone size and shape. A combined evaluation of the geometric and mechanical alterations permits changes in the intrinsic properties of bone during the treatment period to be distinguished from modifications resulting from changes

in the size and shape of the skeleton. This evaluation is particularly important in the case of agents that accumulate in bone and may alter the quality of the organic or crystal phase. The distinction can be made in bending and torsional tests on long-bone diaphyses by measurement of cross-sectional moment of inertia (a measure of the distribution of bone tissue around the central or neutral axis); at the proximal femur, by contrast, both the axis length and the cross-sectional moment of inertia at the femoral neck should be measured by an image-analysis system. Calculation of the moment of inertia from cylindrical or overall approximation of the cross-sectional geometry of the long bones should be avoided because of the intrinsic inaccuracy of the method.

For intact vertebral bodies, information on ultimate strength and intrinsic stiffness should come from compressive tests undertaken in a cephalo-caudal axis. Data should be analysed in relation to the effects of the intervention on bone mass, microarchitecture, and the material properties of skeletal tissue.

Some fractures in humans result from an accumulation of “microdamage” in cortical or cancellous bone caused by cyclical loading at forces below the fracture threshold. Such microdamage involves the accumulation of cracks (in cortical bone) or microfractures (in individual trabeculae) over time. When the number of cracks or microfractures becomes critical, a complete fracture may occur, sometimes with little or no trauma. Modelling “stress fractures” or “fatigue fractures” is difficult and may require a specially designed apparatus to fix an animal’s limb in such a way that it can be subjected to repetitive impact loading; alternatively, a continuous treadmill can be used, which imposes continual cyclical loads on the animal’s lower extremities.

4.6.4 Biochemical indices of skeletal metabolism

Indices of bone turnover

Indices of bone turnover are a valuable adjunct in short-term studies designed to determine the full dose range of a test agent. In long-term studies, indices of bone turnover should be considered as secondary end-points that may provide supporting evidence for the mechanism of action of a test agent. Osteodensitometric, histomorphometric, and biomechanical measurements are considered to be appropriate primary end-points.

Preference should be given to those indices of skeletal metabolism whose validity — both for the particular model and for the therapeutic intervention being tested — is supported by experimental and clinical evidence. Suggested bone-resorption markers validated in the context of many animal models include the pyridinoline cross-links and associated peptides, hydroxyproline, and urinary excretion of molecules previously incorporated into bone such as tetracycline, calcium isotopes, or strontium. The most widely validated indices of bone formation, though not in all species, include bone-specific alkaline phosphatase and osteocalcin. Fasting urinary excretion of calcium remains a useful variable in estimating net bone resorption.

Isotope techniques

Studies on the kinetics of radioactive calcium, tetracycline, or strontium, coupled with balance measurements, can provide useful information on the effects of the agent on calcium deposition and release. Like markers of bone turnover, they provide an assessment of the integrated effects on the skeleton.

4.6.5 Fracture healing

Bone repair following fracture involves several processes, comprising formation of the callus that bridges both bone ends, and the mineralization of the callus and its internal remodelling at the fracture site. These processes also take place, at least in part, during the healing of cortical or trabecular microfractures. Agents that affect bone metabolism may influence these processes, and it is therefore recommended that appropriate tests be undertaken on the fracture healing process, particularly where clinical studies using fracture as an end-point are not envisaged.

Models of fracture

Standardized closed fracture models of the rat femur or tibia, in which the bone fragments are supported by intramedullary fixation, may be appropriate for assessing the initial formation of callus and subsequent union.

In the long bones of dogs or sheep, an osteotomy model supported by plate and screw fixation may be of value for assessing remodelling at fracture sites. This procedure modifies callus formation, however, and is complicated by variable times for wound healing.

No well validated models are available for evaluating the healing of cancellous bone fractures such as those of the vertebral bodies.

End-points

Both radiographic and histological examinations are recommended for the evaluation of the healing process. In addition, mechanical tests should be undertaken at a suitable stage to determine the quality of fracture union (see section 4.7).

4.7 Pharmacokinetics

As for any new agent, appropriate experiments should be undertaken in at least two species in order to acquire appropriate pharmacokinetic information. For orally administered agents, these should include estimating bioavailability by determining the intestinal absorption rate with and without concomitant intake of food. Where possible, conventional measurement of the area under the

plasma concentration curve should be undertaken after oral and intravenous administration.

Pharmacokinetics and distribution should include the time (t_{\max}) between oral administration under standard conditions, such as after an overnight fast, and the appearance of maximal plasma concentrations (c_{\max}). Standard pharmacokinetic parameters, such as half-life, plasma levels following repeated administration, the distribution volume or plasma clearance, and protein binding, should be evaluated where possible. It may be necessary to explore long- and short-term accumulation at various skeletal sites and subsequent release, particularly in the case of agents that are incorporated into the skeleton. In such cases, steady-state pharmacokinetic measurements are problematic and may not be possible.

For each agent, routes of excretion should be documented and, where appropriate, the metabolism characterized.

4.8 Studies of long-term bone safety

In addition to conventional toxicological studies (see section 4.9), it is suggested that skeletal end-points might be incorporated into chronic toxicity. These end-points should be aimed at assessing the long-term effects of the agent on bone quality and might also include pharmacodynamic responses. The most critical end-points for bone quality are impairment of growth, mineralization defects, reductions in bone mass, and changes in bone strength. Most of the techniques outlined so far, including quantitative bone histology with fluorochrome prelabelling and tests for bone strength, can usefully be incorporated into toxicological programmes.

4.9 Other toxicological studies

The regulations governing experimental toxicological documentation of a new pharmaceutical product may vary from country to country. Within the International Conference on Harmonisation (ICH), some drug regulatory authorities (of European Union countries, Japan, and the USA) have achieved, or are working towards, harmonization of their requirements for these studies. The following ICH guidelines on experimental safety studies have been finalized and can be obtained from the ICH Secretariat.¹

- *Guideline on the need for carcinogenicity studies of pharmaceuticals* (S1A).
- *Carcinogenicity: use of two rodent species* (S1B).
- *Dose selection for carcinogenicity studies of pharmaceuticals* (S1C).
- *Genotoxicity: standard backing tests* (S2B).
- *Toxicokinetics: guidance on the assessment of systemic exposure in toxicity studies* (S3A).

¹ International Conference on Harmonisation, 30 rue de St Jean, 1203 Geneva 18, Switzerland; telephone: (+41)22 340 1200; fax: (+41)22 340 1380.

- *Pharmacokinetics: guidance for repeated dose tissue distribution studies (S3B).*
- *Single dose and repeat dose toxicity test (S4).*
- *Detection of toxicity to reproduction for medicinal products (S5A).*
- *Reproductive toxicology: male fertility studies (S5B).*
- *Safety studies for biotechnology products (S6).*

4.10 Minimal preclinical requirements for clinical evaluation in phase III

Much evidence indicates that adequate preclinical evaluation in terms of pharmacodynamics and bone quality can forecast the success or failure of drugs in preventing or treating osteoporosis. A well documented preclinical evaluation is therefore generally appropriate before a phase III clinical programme is undertaken, particularly where fracture studies are not envisaged. Evaluations completed before the start of clinical studies should include pharmacodynamic and bone tolerance data obtained by osteodensitometric, histomorphometric, and biomechanical methods. Information on the mechanism of action of the agent is also important, particularly where this will influence the choice of endpoints in clinical trials.

5.

Clinical trials in osteoporosis

5.1 Introduction

The clinical consequences of osteoporosis are fracture and any subsequent complications. The aim of all interventions, whether applied as preventive measures or in the management of the established disorder, is therefore to reduce the risk of fragility fractures.

These guidelines deal principally with pharmacological interventions in osteoporosis but are nevertheless appropriate for the assessment of other, non-pharmacological treatment modalities targeted to populations or to high-risk groups. Examples of non-pharmacological interventions include physical exercise, cessation of tobacco smoking, modifications of diet or other aspects of lifestyle, and the use of hip protectors in the elderly.

Prospective studies indicate that measurements of the amount of bone present can be used as a proxy for fracture risk in healthy female populations and untreated women with osteoporosis. For each standard deviation difference in bone mineral density, the risk of an osteoporotic fracture changes by a factor of 1.5–3. Higher gradients of risk are observed when predictions of specific fractures (e.g. of the hip) are made from measurements of bone mineral density at the relevant sites. The relative and absolute risks are well characterized in postmenopausal Caucasian women, but the absolute risks, in particular, are less secure in other races and in men.

Use of bone density measurements to predict fracture risk in female populations is at least as valid as the use of blood pressure measurements to predict stroke. Not all hypertensive patients suffer strokes, however, and not all strokes are due to hypertension; similarly, a low bone density does not mean that fracture is inevitable — only that the probability is high. This probabilistic association between bone mass and fracture poses many problems for the assessment of drug efficacy in individuals. In the individual, efficacy against fractures can be assessed only indirectly, either from changes in bone mineral density (or other surrogate of fracture risk), or on the basis of clinical research which assesses the effects of interventions on the fracture risk in populations. Practising physicians and other health-care professionals therefore have to be guided by the results of clinical and preclinical research rather than by personal experience.

5.2 Bone mass and fracture

With advancing age and progressive bone loss, the importance of skeletal factors other than bone density and of extraskkeletal factors for fracture risk increases in both men and women. Additional skeletal factors that contribute to bone strength *in vivo* include the shape of particular bones and the discontinuity of trabecular architecture caused by trabecular plate perforations. Extraskkeletal factors include both the greater likelihood of falls and the age-related deterioration of neuromuscular protective mechanisms. The relative importance of these additional factors increases with age (and varies with ethnicity); thus, for a given bone density, absolute fracture will be higher in elderly individuals than in other age groups. In other words, considerations of age (and sex and ethnicity) capture an element of risk that is not captured by the assessment of bone density. For these reasons, interpretation of intervention-induced changes in skeletal mass is most secure in studies of prevention of osteoporosis; greater reservation is called for in studies conducted late in the natural history of the disorder, for example in studies on hip fracture risk. Notwithstanding, the predictive value of bone density measurements for *relative* fracture risk in untreated patients does not change with age.

A major problem with the use of bone mineral density to predict fracture risk has been the view that a *change* in density induced by an intervention may not be associated with a comparable *change* in the risk of fracture. Interventions may have markedly different effects on fracture risk despite having similar effects on bone mineral density. This phenomenon is attributable to changes in the material properties or “quality” of bone. Moreover, many forms of osteoporosis, particularly postmenopausal osteoporosis, are associated not only with bone loss but also with subsequent destruction of skeletal architecture which alters the connectivity of trabecular elements and results in the structure of cancellous bone being weakened out of proportion to the amount of bone lost. Stimulators of bone formation may, for example, restore skeletal mass, but have little if any effect in restoring the trabecular connectivity of cancellous bone. As a result, skeletal strength would improve less than the change in bone mineral density would suggest. This should not be taken to mean that prevention of bone loss is not of therapeutic value, even in patients with advanced bone disease, but rather that the restoration of skeletal mass to normal may not restore fracture risk to the same level as current therapeutic regimens. Since the risk of many osteoporotic fractures increases exponentially with age, even a slight reduction in the age-related increment in risk is likely to have a large effect on fracture frequency.

A change in bone mineral density may be most accurate as an indicator of change in fracture risk early in the course of osteoporosis, when skeletal architecture is reasonably intact. In postmenopausal osteoporosis, the more time that has elapsed since the menopause, the less justifiable is the assumption that treatment-induced changes in bone mineral content or density at one site reflect a proportional change in fracture risk at another.

A second problem relates to the effects of interventions on the quality of bone. The use of some bone-active agents may impair the mineralization of new bone or result in new bone being woven rather than lamellar in structure. Examples include high doses of etidronate and fluoride, both of which may result in increased numbers of fractures despite the preservation or augmentation of bone mass. Another current concern is that excessive attenuation of bone turnover by inhibitors of bone resorption may adversely affect the self-repairing capacity of the skeleton, despite the maintenance of bone mass. These problems can often be put in perspective by adequate preclinical assessment (see Section 4) or by appropriate histological, pharmacodynamic, and kinetic studies in the clinical development phase (see below).

It is thus apparent that different approaches to phase III (and even phase II) studies are appropriate for different classes of agent and perhaps for different ages at which intervention is contemplated. Where a new agent has the same mechanism of action and pharmacodynamic profile as an existing agent of known efficacy in reducing fracture risk, phase III studies on fracture may not be required if animal studies have already shown effects comparable to the reference agent. In addition, where the mechanism of action and pharmacodynamics of well-characterized classes of compounds are known, preclinical evaluation shows no adverse effects on the material and structural properties of bone, and no qualitative abnormalities are found, bone mineral density measurements may provide adequate evidence for efficacy in osteoporosis.

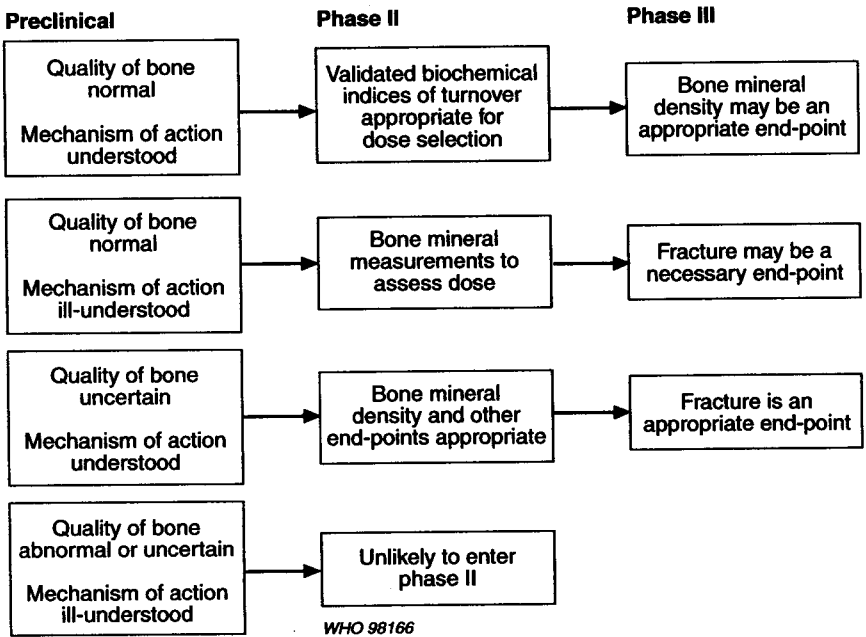
However, if preclinical studies show that effects on the quality of bone are in any way uncertain, fracture becomes the only appropriate primary end-point. A summary of strategic routes for clinical development is provided in Fig. 1.

5.3 Techniques used in the assessment of osteoporosis and response to interventions

Measurement of the bone mineral content of the skeleton has a central place in osteoporosis research and the assessment of intervention in phase III studies. The past two decades have seen the development of a variety of techniques that are — in principle — usable in clinical trials of osteoporosis, provided that their characteristics (precision, accuracy, and interpretation) are taken into account in the study design, calculation of sample size, analyses, and interpretation. These techniques include single- and dual-energy absorptiometry (isotope-based and X-ray based), quantitative computed tomography (QCT), neutron activation analysis, quantitative radiography, ultrasound attenuation, and ultrasound velocity and reflection.

The two techniques most widely utilized to assess bone mass are single- and dual-energy absorptiometry, which variously assess mineral content of the whole skeleton or particular sites, particularly those sites at risk of osteoporotic fracture such as the wrist, spine, and hip.

Fig. 1 Strategic routes for clinical development



5.3.1 Single-energy absorptiometry

Single-photon absorptiometry (SPA) measures bone mineral at peripheral (appendicular) sites such as the heel and the wrist and was the first method developed for the purpose. Most commonly, ^{125}I is used as the photon-emitting source. The amount of bone mineral in the tissue traversed by a well collimated gamma-ray beam is derived from its attenuation through bone and soft tissue relative to that through soft tissue alone. Single-energy X-ray absorptiometry (SXA) has recently become available for scanning the wrist. It uses an X-ray source and avoids the need for isotopes.

5.3.2 Dual-energy absorptiometry

Sites such as the spine and hip present particular problems and cannot be accurately assessed by SPA or SXA. The development of dual-photon and dual-energy absorptiometry (DPA, DXA) has resolved many of the problems, and these techniques permit measurement of bone mineral at these sites and elsewhere.

5.3.3 Other techniques

The use of other techniques developed to measure bone mineral is generally restricted to clinical research; with the exception of QCT, their value for measuring treatment-induced changes is uncertain. These limitations currently extend to the use of ultrasound measurements, both for trial inclusion and to measure the primary end-point of intervention. Standard skeletal radiographs provide very inaccurate assessments of the amount of bone mineral. Available quantitative radiographic techniques have been shown to be accurate, precise, and responsive to some interventions as regards measurements of cortical width, but they do not capture changes in the porosity of cortical bone.

5.3.4 Limitations of techniques

Bone mineral content (in g or g/cm) or “density” (in g/cm²) may be determined at focal axial or appendicular skeletal sites (e.g. forearm, heel, spine) or in the total skeleton. There are limitations to all techniques, relating to the type of information obtained, and to sites of measurement and their relevance to fracture, which hamper interpretation of apparent changes. The most important of these, from the point of view of clinical trials, are problems of *precision* and *accuracy*.

Precision, or repeatability, refers to the agreement between a set of replicate measurements. Measurement of bone loss over a reasonable period of time is difficult with available techniques, because very small changes must be detected (about 50 mg of calcium daily, from a total calcium pool of 750 000 mg). Postmenopausal bone loss over 1 year is approximately 1–3%, and this figure is similar to the precision error of many of the available methods. Precision is thus of major importance in the design of clinical trials in osteoporosis.

In order to limit the numbers of patients required for a study, care should be taken to use techniques in the most precise manner possible. For example, determinations of combined cortical width by metacarpal morphometry are more reproducible for sequential observations than determinations of area.

For results to be valid and conclusive, it is important to assess the precision of the equipment concerned in the hands of the intended operators. The long-term *in-vivo* precision should be calculated on the basis of repeated measurements over a period of years in individuals belonging to the population to be studied. Methods should not be changed during a study.

Accuracy is the nearness of a test result to the true value. That is, it represents the extent of agreement between a test result and the “true” value of the parameter being measured. In terms of bone mineral measurements, it reflects the ability to estimate the true amount of skeletal calcium assessed by cadaver experiments. Apart from QCT, none of the commonly used techniques for determining bone density actually measures true density; typically they give values for bone mineral content per unit area. In addition, most techniques are

affected to a varying extent by soft tissue composition, which makes their accuracy *in vivo* less than that *in vitro*. Accuracy is important in the stratification of patients into groups at risk, and may be of major importance for clinical trials if bone mineral content or density is to be a criterion for inclusion or exclusion. Errors of classification in the elderly may occur because of aortic calcification, spinal osteoarthritis, or prevalent (existing) fractures at the lumbar spine. Affected vertebrae should be disregarded for the purposes of entry criteria. It may be feasible to measure bone mineral at another site (e.g. the hip or forearm) for inclusion purposes, even when the primary site for investigation is to be the spine.

Errors of accuracy may also arise during the course of the study itself. Systematic changes in accuracy may occur with interventions (for instance, those that change lean body mass or fat composition) or with changes in isotope source or detection equipment. Incident fractures may have spurious effects on bone mineral measurements and affected vertebrae should be excluded from analyses both before and after fracture. The same technique may yield different results in different laboratories because of systematic biases (in standards, edge detection, etc.), and centralized quality control and the appropriate use of standards are thus important aspects of trial design. Ideally, the same make and model of equipment should be used for bone mineral assessments in all parts of multicentre studies, since the population mean and variance of bone mineral measurements may differ between instruments. Where this is not feasible, analytical power should be sufficient to allow the effect of changes in equipment to be evaluated. The same instrument should be used for the same individuals throughout the trial; in the rare instances when an instrument must be changed, the appropriateness of the change should be validated.

5.3.5 Sites of bone mineral measurements

The sites selected for bone mineral measurements will depend upon the objectives of intervention. Sites of biological relevance include the distal forearm, the hip, and the spine, but measurements made at other sites may also be used to provide supporting information on efficacy or safety or where their validity has been demonstrated.

In general, measurements should be carried out at two or more skeletal sites, one of which should be an axial site. Measurements at the proximal femur have a high priority because of the clinical significance of this site. Changes at sites comprising mainly cancellous or mainly cortical bone may need to be separately documented. In some instances, where interventions may have variable effects at different sites, whole-body measurements may be appropriate. Selection of sites and/or techniques may also be influenced by the likelihood of accuracy errors that may be induced by treatment affecting soft-tissue composition.

Use of bone mineral density measurements is subject to a number of additional considerations that relate to the design of clinical trials and the duration of observation (see section 5.8).

5.4 Fracture assessment

5.4.1 Fractures as entry criteria and end-points

Fractures may be assessed as entry criteria or be used as an end-point. All studies in osteoporosis should document prior fractures according to site at study entry. All subsequent fractures should also be documented, even where fracture is not an end-point of the particular study. Time to fracture should also be documented where possible and may provide a primary end-point, for example in studies of hip fracture. In assessing fractures it is important to distinguish between high- and low-energy-induced events; stress fractures should be documented separately, with the method of their detection.

Most fractures are clinically obvious and readily documented. In the case of vertebral fracture, however, problems arise in defining the presence or absence of fracture, and the changes that occur in vertebral shape.

5.4.2 Vertebral fractures

There is no “gold standard” for the definition of vertebral fracture, but vertebral wedge and endplate deformities, which occur with greater frequency than complete crush fractures, are commonly documented in clinical trials. The advantage of examining these deformities is that fewer patients may be required for studies of fracture; however, wedge and endplate deformities have less clinical significance than symptomatic fractures.

Quantitative morphometric approaches may be used both to assess vertebral fractures for entry criteria and to assess incident fractures. Several algorithms have been proposed (12) which may be applied to standardized lateral radiographs of the spine or to absorptiometric images. However, interpretation of changes in shape of the vertebral bodies may be difficult for the following reasons:

- Not all vertebral deformities are due to osteoporosis, and it may therefore be necessary to use the services of a centre with appropriate experience for the triage of radiographic abnormalities.
- There are intrinsic statistical problems in assessing prevalent and incident fractures, resulting in part from variations in normal vertebral heights at different vertebral levels (13). Algorithms should therefore take account of this variability by means of level-specific criteria or stringent criteria for an incident fracture (e.g. a 20% decrease in anterior or central height). A further problem is the false-positive rate, which is high for many approaches because of the large number of measurements made (14).

For example, it seems reasonable to characterize a vertebral deformity as being present if the difference in height (posterior, central, or anterior) is two standard deviations below that expected, or below that of adjacent undeformed vertebrae. If the anterior, posterior, and central heights of T4 to L5 (14 vertebrae, 42 sites) are measured in 100 healthy subjects (4200 measurements), the use of 2 standard deviations to characterize a deformity will result in 2.2% of the measurements (92 measurements in only 100 subjects) being spuriously attributed to a vertebral deformity. A more stringent cut-off, say 3 standard deviations, will capture fewer false-positives (0.1% of the vertebrae, or 4 false-positives); this is an acceptable rate, but there is the added problem that real fractures are lost as well as false-positives.

This false-positive rate profoundly affects the apparent prevalence and incidence of vertebral fracture. The extent to which this clouds interpretation of clinical trials in osteoporosis depends on the frequency of false-positives compared with real events. A relatively high false-positive rate will reduce the apparent efficacy of an effective intervention. Moreover, an unequal distribution of false-positive events between treatment areas will cause bias in this estimate of efficacy when the incidence of false-positives is high or that of true-positives is low.

There is no universal agreement on the method of reporting the incidence of vertebral fractures. Since vertebral fractures are commonly diagnosed by an abnormal vertebral shape, an abnormality in a single X-ray may reflect a long-standing (prevalent) fracture or a recent event (incident fracture). Different criteria are commonly used to differentiate an incident fracture from a prevalent fracture. Reporting changes in vertebral deformity scores in terms of the number of newly deformed vertebrae or the number of patients with new deformities yields conflicting results. Consider, for example, a drug trial in which 14 new vertebral fractures are observed in the placebo group and 7 in the treatment group; apparent drug efficacy is 50%. In reality, 14 new fractures may have occurred in 5 patients in the one group and 7 new fractures in 5 patients in the other; drug efficacy is then apparently zero.

Since the apparent outcome of studies of vertebral fracture depends critically on the methodology used to assess vertebral fracture and to express the results, such studies should provide detailed information on the criteria used to define incident and prevalent fractures in trial populations, plus justification for the methodology. Since the precision of quantitative methods varies, as does — perhaps — the pathophysiological process, recording of new vertebral fractures should make the distinction between fractures occurring *de novo* in previously undeformed vertebrae and progressive deformities in those previously characterized as fractured. It is also important to take account of progressive deformities that initially just fail to reach the diagnostic criteria but just exceed them on subsequent measurement. Clinical trials that report repeated spinal radiography should document the numbers of patients with new fractures over a given time interval as well as the cumulative number of fractures and/or their

severity. Both end-points are clinically relevant, but the primary end-point should be specified.

Techniques for radiographic or absorptiometric imaging should be standardized in an acceptable manner. Images should be read with the investigator blinded to the treatment allocation — but not necessarily to other information such as patient identity or the temporal sequence of tests, to avoid errors of classification — preferably at a centre specialized in the assessment of vertebral morphometry. In multicentre studies, it is advisable for radiographs to be read at a central location, with a central laboratory designated to oversee quality control.

Studies of vertebral fracture frequency using imaging techniques (radiography or absorptiometry) should report the following information separately and at each time interval studied:

- the number of patients who develop at least one new radiographic deformity in the relevant time interval;
- the number of patients in whom established deformities, ascribed to osteoporosis, deteriorate;
- the number of new vertebral deformities;
- the number of progressive vertebral deformities.

The primary end-point will depend upon the hypothesis to be tested (e.g. the time to fracture, the proportion of individuals sustaining fracture).

5.5 Height

In randomized controlled trials of severe (established) vertebral osteoporosis, differences in height between test groups may be taken as a useful index of antifracture efficacy. Some data have been published that show the relation between height loss and vertebral fractures; subject to further confirmation, measurement of height could become a validated primary end-point. The most reproducible method uses the Harpenden stadiometer but with all techniques adequate quality control, with the daily use of height standards, is an important component and should be documented.

5.6 Quality of life

One of the ultimate aims of studies of vertebral and other osteoporotic fractures is to reduce the associated morbidity and thereby improve quality of life. Although there are validated instruments for measuring both morbidity from vertebral fracture and quality of life, none has been shown to be sufficiently sensitive for documenting changes in trials of efficacy. Several instruments are being developed, and the use of both generic and specific instruments is therefore important, but not a requirement, in trials of efficacy.

5.7 Other measurements

The techniques outlined above should be distinguished from estimates of safety and from measurements that provide information on the mechanism of action of the agent. The latter include indices of skeletal metabolism (alkaline phosphatase, osteocalcin, urinary calcium/creatinine ratio, etc.) and major calcium-regulating hormones (e.g. parathyroid hormone, calcitriol, calcitonin). The choice of tests to provide pharmacodynamic information should be related to the mechanism of action of the agent under evaluation. This type of test may also be useful in phase II studies to provide information on dose-responses: the changes that occur over short periods are large compared with the effects on skeletal mass. These tests differ from measurements that might be made to characterize the population, particularly at study entry, such as bone biopsy, serum calcidiol to exclude osteomalacia, nutritional history, etc. Measurements of body composition, such as fat and lean body mass, may be appropriate where interventions are known or suspected to cause changes that may affect the accuracy of mineral measurements.

5.7.1 Bone biopsy

Bone biopsies are necessarily taken from restricted skeletal sites; they are usually transiliac. For many of the quantitative measurements, the coefficient of variation is high, and for this reason the technique is no longer considered suitable for the estimation of bone mass. However, bone biopsy may be a useful adjunct in the assessment of safety and in pharmacodynamic determinations. Prelabelling of bone with tetracycline permits mineral apposition and bone formation rates to be estimated and, when coupled with measurements of osteoid seam thickness, the presence or absence of osteomalacia to be ascertained. These measurements are useful when preclinical studies suggest that an agent under investigation may impair bone turnover or mineralization. In studies of prevention, it may be valuable to look at trabecular disposition and connectivity. Bone biopsy may also be used to detect the absence or presence of woven bone and for other histological estimates of safety.

Bone biopsies are essential where preclinical evaluation has raised the possibility that the quality or mechanical competence of bone may be inadequate or where the mechanism of action of the agent is poorly understood. Biopsies obtained for either of these reasons should normally be taken at the end of the period of exposure. In some circumstances, however, such as cyclic or intermittent exposure to an agent, evaluation of an agent that is retained in the skeleton, or early evaluation of the quality of bone, biopsy at different time-points during exposure may be informative. Paired biopsies are not necessary provided that the sample size is adequate and appropriately documented reference ranges are available. A placebo group in the study provides an appropriate comparison group. Phase II and/or phase III studies may be

appropriate stages of drug development at which to perform biopsies if required.

5.7.2 Biochemical indices of bone turnover

Indirect indices of skeletal metabolism are valuable for the assessment of patients and sometimes for elucidating the mechanism of action of an agent and evaluating treatment. Their use has been best characterized with inhibitors of bone resorption in postmenopausal osteoporosis. In contrast to focal assessments of disease by bone histology, regional assessments of bone mass, or radiography, biochemical indices provide an integrated assessment of the activity of the underlying disorder. Disease activity may be assessed in terms of calcium metabolism, collagen turnover, and indices of the functional activity of bone cells themselves.

Indices of calcium metabolism include serum calcium adjusted for fluctuations in albumin (or ionized calcium) and the fasting urinary excretion of calcium.

Alkaline phosphatase activity, derived in part from osteoblasts, is the most frequently used biochemical marker of skeletal disease and provides an index of bone formation. The bone-derived fraction of alkaline phosphatase is more specific than total alkaline phosphatase. Serum osteocalcin is thought to be synthesized exclusively by osteoblasts and odontoblasts and is therefore a more sensitive index of increases in bone turnover than total alkaline phosphatase. Care is required in interpreting values in the presence of renal failure, since the kidney is a site of metabolism of osteocalcin. The relationship between changes in these indices and changes in bone formation has been best documented in the context of postmenopausal osteoporosis following treatment with inhibitors of bone resorption. Interpretation of responses is more problematic with many other agents.

Of the various indices of bone resorption available, the most commonly used is urinary excretion of hydroxyproline. The collection of samples under fasting conditions and the expression of the measurement as a ratio to creatinine excretion obviates the need for 24-hour urine collections and dietary restrictions. More specific indices of bone resorption include the urinary pyridinoline cross-links and related peptides. All markers of bone resorption have been best validated in the context of inhibitors of bone resorption.

5.8 General principles of trial design

Irrespective of the techniques to be used and the way in which data are interpreted, a number of important general principles apply to the design of studies of efficacy.

5.8.1 Population to be studied

It is important that sampling for studies draws from an appropriate population. For example, trials of preventive efficacy should not be undertaken in patients with severe osteoporosis. Similarly, the finding that bone loss at the menopause can be prevented should not be taken to imply that subsequent fracture rates would decrease in a population with severe osteoporosis; this may be the case, but testing in an appropriate population would be required.

Where appropriate, confirmatory trials should be undertaken in study populations of different ethnic compositions for the following reasons:

- causes and mechanisms of bone loss may vary between ethnic groups;
- bone mass and densitometric definitions of osteopenia will also vary;
- fracture incidence varies widely with geographical location and ethnic group;
- environmental conditions may vary widely for different ethnic groups.

The target population should be defined according to sex, age, ethnic composition, nutritional status, physiological state and — where appropriate — cause of osteoporosis or osteopenia. In postmenopausal women the date and type (natural or secondary) of menopause should be documented. In glucocorticoid-induced bone loss, the characterization should include the dose and duration of glucocorticoid therapy. Similar characterization is needed for other causes of osteoporosis.

Characterization of the study population with respect to other recognized risk factors for fracture is also desirable, since this information may help to define responsive or unresponsive sub-populations. Examples include smoking, a family history of hip fracture, and hip axis length. Consideration should be given to determination of genetic markers of osteopenia or fracture risk, since experimental data suggest that genotype may influence responsiveness to interventions.

Inclusion and exclusion criteria should be defined, but care should be taken not to make the study inappropriate for the population for whom treatment is intended. These criteria should be selected with due consideration to the likely effects and side-effects of the intervention. They may also be used to optimize the power of a study, for example by avoiding the heterogeneity of bone loss (e.g. by excluding women at the time of oophorectomy), of causes (e.g. glucocorticoid-induced osteoporosis), significant diseases that might impair the interpretation of results (e.g. osteomalacia), prior exposure to bone-active agents, and concurrent medication that influences bone metabolism. Stratification should be considered as an approach to avoid confounding due to common risk factors. Similarly, diseases such as previous hyperthyroidism, alcoholism, and treated hyperparathyroidism, and corticosteroid exposure may be accommodated by pre-hoc stratification or by post-hoc analysis where this is made explicit in the study protocol.

Although inclusion and exclusion criteria may increase the homogeneity of study populations, there is a risk of a biased population being studied, inappropriate to the final use of the intervention. Restrictive criteria may mean that the ultimate use of an approved agent is limited to the specific population studied or that labelling of an agent with regard to its safety must be modified.

Prior treatment of osteoporosis or osteopenia and the use of bone-active agents may or may not be exclusion criteria, but clear reasons for the enrolment strategy should always be given in the study design.

5.8.2 Treatment regimen

When a combination of treatments is given, studies should normally evaluate each of the components of that treatment unless previous studies have provided a rationale or adequate justification for combined or sequential therapies.

5.8.3 Study design

The aims of the study should be defined clearly and the end-points described. Criteria for efficacy should be specified in the initial protocol. Where multiple techniques are used, the primary indicator of efficacy should be specified. End-points or criteria of efficacy should be distinguished from measurements of safety and measurements that provide information on the mechanism of action of the intervention.

Pivotal — rather than supporting or circumstantial — evidence for efficacy should be based on prospective studies regardless of whether the study envisaged is clinical or (in phase IV) epidemiological.

Clinical studies of efficacy should be randomized, double-blind, and controlled. There may well be problems with this kind of trial design in long-term studies where the drop-out rate could be appreciable. In randomized controlled studies, the control group may be randomized to placebo or to a proven active agent. In the latter case, the active agent(s) should already have been adequately characterized by placebo-controlled studies. Stratification may be necessary in studies of mixed populations, such as men and women, or patients with osteoporosis of different etiologies, or where different outcomes are likely. Where less stringent entry criteria are used, population pharmacodynamics may be an appropriate adjunct in assessing the heterogeneity of responses. Care should be taken to avoid unblinding during the assessment of variables to be used as end-points, e.g. biochemical markers.

In most study designs, the duration of exposure of controls should be identical to the duration of exposure to the agent under investigation since there may be time-dependent differences in rates of change of bone mass and of fracture frequency.

Cross-over studies are not generally possible and are likely to be inappropriate because of the long-term effects of intervention on bone turnover.

5.8.4 Duration of studies

The detection of bone loss in individuals depends upon the rate of bone loss, the duration of study, and the reproducibility of the techniques applied. Recent technical developments have significantly improved the precision of measurements, including those at sites of clinical relevance. In clinical trials in which populations are studied, apparent variability in the rates of bone loss and the ability to detect differences in loss depend on the heterogeneity of bone loss, in both untreated and treated populations. All the therapeutic interventions thus far investigated, with the possible exception of fluoride, initially induce transient changes in bone mass, with rates of change varying until a new steady state is achieved. Theoretical and experimental evidence suggests that achievement of the new steady state may take 2 years or more in postmenopausal osteoporosis. In addition, rates of fracture also change with time, and long-term observations are required to document the therapeutic effect of interventions on fracture.

The duration of study will depend upon the phase of development, the nature of the intervention, and the claim to be made for the activity of the agent. When treatment is of limited duration, it is important to study offset of activity in addition to the effects of an intervention, since both may affect the ultimate therapeutic benefits. For this reason, studies of the offset of effect of interventions are an important component of clinical development. Recommendations on the duration of exposure to an effective agent should take account of the pharmacodynamics of its offset of action.

5.8.5 Sample size

The factors that determine sample size for the study are:

- The nature of the distribution of the measurement to be used as an endpoint.
- The difference between treatment groups that is considered to be clinically significant (for example, a specified difference in fracture frequency or a given difference in bone mineral density at the end of treatment) compared with another treatment or placebo. The acceptable difference will vary according to the characteristics, including unwanted effects, of the intervention used, and should be specified. In studies of equivalence of effect, the clinically irrelevant difference is used for sample size calculations.
- The duration of study.
- The expected number and probable timing of drop-outs.
- The heterogeneity of change (standardized difference) in both groups.

The null hypothesis to be tested will affect the sample size. The study should have sufficient power (i.e. be of sufficient size) to reject the null hypothesis, if it is wrong, in favour of the alternative hypothesis. The power should be at least 80%. The probability of erroneously rejecting the null hypothesis, where in fact it is true, should be set at $p = 0.05$ or less. Care should be taken to adjust the probability if several primary end-points have been assessed and tested.

Study size will also depend to an extent on the precision of the techniques selected. Assessment of precision should not be based solely on the coefficient of variation obtained in phantom studies or in normal subjects, but should be a precision estimate appropriate for the population under study.

The features that determine power should be evaluated at the start of the study. In some instances, information (for example on heterogeneity of response to treatment) may not be available, and appropriate provision should then be made in the protocol for assessing heterogeneity during the conduct of the study without breaking the blind, so that a larger sample can be recruited if necessary. Such an analysis should not be used to stop the trial, but only to increase the numbers or the duration of treatment. The random code should be opened only in planned interim analysis or in cases of emergency.

The hypotheses tested may include a test agent being equivalent to, or better or worse than, an established treatment. Where a test treatment is considered to be less effective than an established treatment, use of a placebo group, or other direct means of demonstrating that the treatment is effective, is strongly recommended.

Primary analyses should normally be based on intention to treat. Drop-outs during the course of long-term studies pose major problems for analysis; they may destroy the randomization applied to the various groups of the study and decrease the power of intention-to-treat analyses. Where entry is stratified according to sex or types of osteoporosis, the power of the study should be sufficient to allow analysis of separate strata. Reasons for withdrawal should be reported together with an analysis of the way in which drop-out patients differ from those who complete the trial (in the whole study population and in the various groups). Where possible, patients withdrawing should be followed up; where appropriate, they may rejoin the study.

5.8.6 Frequency of measurement

Ideally, measurements might be made at the start and at the end of the study period, but repeated determinations are suggested for the following purposes:

- to ensure that systematic problems, particularly in bone mass measurements (for example, changes in accuracy due to source changes), do not occur;
- to document transients and new steady states;
- to compare rates of change;
- to improve the efficiency of intention-to-treat analyses.

In practice, bone mineral measurements should be undertaken 6-monthly when studies of 2 years' duration are envisaged and at least once yearly in studies lasting 3 years or more. There are advantages in obtaining two baseline measurements. Vertebral radiographs should be assessed annually.

5.8.7 Safety

Due account should be taken of the hazards involved in using any of the various techniques. For example, single- and dual-photon absorptiometry and dual-energy X-ray analysis may be appropriate for use in children, but computerized axial tomography delivers a higher radiation dose and the frequency of testing should be reduced.

Together with standard pharmacokinetic studies, population pharmacokinetics are often an appropriate investigational aspect of safety.

Kidney and liver function tests and serum calcium, phosphate, and parathyroid hormone are appropriate tests of safety in most situations, in addition to the classical evaluation of haematological indices. Bone biopsy may also be an appropriate index of safety in certain circumstances.

In view of the long duration of clinical trials involving bone-active interventions, interim data analyses for safety should be planned. In general, safety end-points should be reviewed at least annually by a monitoring committee. Blinding of subjects and investigators should be preserved.

5.9 Non-pharmacological interventions

Non-pharmacological interventions are not a direct consideration for the regulatory agencies involved with drug studies but in some instances may involve other regulatory agencies. It is appropriate, however, that such interventions are assessed and judged with the same rigour as pharmacological agents. Important aspects include the definition of basal conditions, such as the nutritional state of the subjects studied, and the design and analysis of trials. Adverse events are commonly under-reported and compliance and drop-out rates often ignored. In the case of some modalities, such as exercise, dose-response studies and blinding to intervention may not be feasible; this type of difficulty should be taken into account when claims for efficacy are judged. Random allocation should be undertaken where feasible, and the possibility of cross-contamination and other biases (resulting, for example, from the changes in dietary habits induced by exercise regimens) should be monitored.

5.10 Good clinical practice

Studies involving human subjects should be conducted according to the guidelines for good clinical practice for trials on pharmaceutical products (15). The International Conference on Harmonisation has developed the following

harmonized guidelines on developing documentation of efficacy and safety in clinical trials:

- *The extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life-threatening conditions (E1).*
- *Dose–response information to support drug registration (E4).*
- *Studies in support of special populations: geriatrics (E7).*
- *Timing of preclinical studies in relation to clinical trials (M3).*

6.

Phase I and phase II studies

6.1 Phase I studies

In phase I studies, the putative agent is administered at increasing dose levels, first in single, and thereafter in multiple, dosage. Pharmacokinetic, pharmacodynamic, and safety studies should be planned with respect to the findings in animals. Advanced scaling-up techniques (e.g. allometry) should be applied when doses are extrapolated from animals to humans.

In many circumstances, phase I studies will be performed in healthy subjects, but even early phase I studies may be appropriate in patients with osteoporosis, for example, where the agent has an extremely long half-life in the tissues of animals. Phase I studies both in healthy subjects and in patients allow not only traditional safety evaluations of the agent (full blood count, liver enzymes, parameters of renal function, etc.) but also preliminary pharmacodynamic studies, often based on the measurement of biochemical and hormonal indices of bone metabolism (see section 6.2).

In pharmacokinetic studies, the concentration–time course should be described, allowing the relevant pharmacokinetic characteristics such as clearance, volume of distribution, and half-life to be derived. The excretion pathways, including metabolism, should be elucidated; these data should make it possible to predict the influence of pathological changes on metabolism and routes of excretion.

6.2 Phase II studies

Phase II studies are intended to explore a range of doses in order to determine the minimal dose level at which there are detectable pharmacological effects appropriate to the intended therapeutic action, and which is at the same time well tolerated. They are designed to support the dosage recommendations for approval.

The study design should be appropriate for the aim of the study. Addressing the issue of dosage recommendations adequately requires double-blind, placebo-controlled, parallel group studies of long enough duration to demonstrate the effect of the drug on the primary end-point. Where

combination treatments are envisaged, such as the combination of hormone replacement therapy and a test agent, an untreated control group may not be necessary, but all groups should receive the established treatment. The duration of exposure should be sufficient to allow identification of an appropriate dose or doses for phase III investigation and will depend upon the end-point and sample size utilized. The patient population for phase II should be similar to the intended population for phase III studies. Information on pharmacokinetics of special subgroups of patients, such as the elderly, should be gathered at this stage of the development.

Bone mineral density is an accepted primary end-point for phase II in most, but not all, situations. Measurements should be made at clinically relevant sites (see section 5.3), and studies should be of sufficient duration, with sufficient numbers of individuals, to permit dose–response characteristics to be established.

Biochemical indices of bone or calcium turnover may sometimes be considered as primary end-points. They include urinary pyridinoline cross-links and related peptides, bone-derived alkaline phosphatase, and osteocalcin. The use of biochemical indices alone as end-points in phase II will depend upon the agent under investigation. Where mechanisms of action are thoroughly understood, such as in the case of some inhibitors of bone resorption, and where the quality of bone in preclinical studies is not adversely affected, an exposure of 3–6 months may be appropriate. For other agents, particularly stimulators of bone formation, the interpretation of biochemical indices is less secure, and phase II studies should appropriately include other end-points detailed in this section.

In some circumstances, preclinical findings will demand that studies involve more invasive measurements of bone architecture and quality after prolonged treatment.

Bone mineral measurements may need to be supported by additional end-points in some cases, for example where bone size is altered, where body fat and lean body mass are affected, or where the agent affects the accuracy of measurement.

Pharmacodynamic assessments, e.g. intestinal absorption of calcium, concentrations of vitamin D metabolites and parathyroid hormone, may be undertaken to elucidate the pharmacodynamic action of the agent. The safety of the agent should be explored as described in section 6.1 and requires special attention if the dosage range is wide, particularly since the duration of treatment will be long compared with phase I studies.

6.3 Basic conditions

Care should be taken to exclude patients with nutritional disorders such as osteomalacia unless the agent being studied is intended to correct these deficiencies or to counteract their skeletal effects. Otherwise patients should

have normal nutrition with respect to calcium and vitamin D. In some situations, a “run-in” phase to correct nutritional deficiencies may be considered; in others, calcium and/or vitamin D supplements may be given with test medication where considered appropriate, and, if efficacy is proved, the applicant may be required to recommend that the agent be used in conjunction with calcium and/or vitamin D.

There may be situations in which phase II studies provide a direct platform for combined phase II and phase III studies. For example, exposure for 3 months to 1 year (depending on the end-point and mechanism of action of the agent) may permit one or more doses to be identified and study of cohorts given such doses may be continued under phase III.

7.

Phase III studies in severe (established) osteoporosis

7.1 Introduction

A claim for efficacy should be based on the results of at least two well designed and controlled studies. Multicentre studies are appropriate and it may be feasible for results to be combined if the protocol provides for this. Multiple dose levels are not a requirement unless:

- a combined phase II/phase III approach is used; or
- justification for the proposed dose used has not already been otherwise established.

7.2 Population studied

Patients enrolled in phase III studies of severe osteoporosis will have had one or more prior fragility fractures. Where different types of fracture are used as entry criteria it is desirable to stratify individuals according to the type of fracture. When both men and women are studied, either they should be stratified by sex or other balancing procedures should be adopted. Similar considerations apply to different etiologies of osteoporosis (postmenopausal, corticosteroid-induced, etc.). Restrictive inclusion and exclusion criteria are likely to influence the claims that can be made for an intervention (see section 5.8.1).

Where agents for severe osteoporosis are likely to be used in the very elderly, patients from this section of the population should normally form a substantial component of the clinical trial patient base. The number of exposures of elderly patients to compounds intended for use in adults should be appropriate for a separate assessment of the effects of age and co-morbidity such as impaired renal function.

7.3 Basic conditions

Care should be taken to exclude patients with nutritional disorders such as osteomalacia unless the agent being studied is intended to correct these deficiencies or to counteract their skeletal effects. Otherwise patients should have normal nutrition with respect to calcium and vitamin D. In some

situations, a “run-in” phase to correct nutritional deficiencies may be considered; in others, calcium and/or vitamin D supplements may be given with test medication where considered appropriate, and, if efficacy is proved, the applicant may be required to recommend that the agent be used in conjunction with calcium and/or vitamin D.

7.4 End-points

Before the start of a clinical trial, all end-points — both primary and secondary — that are used in assessment of efficacy and safety must be defined and described in the study protocol.

The end-points required to substantiate claims for efficacy depend upon the preclinical evaluation and the class of compound being tested. Where preclinical studies have shown no abnormalities in the quality of bone, and the mechanism of action of the agent is well understood, fracture studies may not be required (see Fig. 1). However, where fracture studies are not undertaken, the absence of information on fracture will affect the labelling of the agent and the claims made for its therapeutic effects.

The ideal treatment should reduce a patient’s risk of fracture to that of an individual with normal bone mass. Notwithstanding, prevention of a progressive increase in fracture risk and lessening of morbidity are also worthwhile aims of treatment. The assessment of fracture rates is not without problems and the feasibility of such studies depends on the type of fracture (e.g. hip or spine) and, in the spine, on the definition of fracture and refracture. For these reasons, the aims of treatment may be limited by the feasibility of measuring ideal end-points. This in turn will influence the claims that can be made for a treatment. The following types of study may be used to assess benefit or efficacy:

- reduction in the rate of fracture;
- an increase in bone mass;
- prevention of bone loss or reduction in the rate of bone loss.

7.4.1 Reduction in the rate of fracture

At the population level, the ideal index of efficacy for an intervention is a reduction in the expected fracture rate. In certain circumstances, however, this ideal may be difficult to achieve. The incidence of hip fracture rises steeply only above the age of 70 years, a period during which deaths from other causes become increasingly frequent. An intervention study in patients who have already sustained one hip fracture may not therefore feasibly adopt a second hip fracture as an end-point because of the sample size requirement. The incidence of vertebral fractures in younger patients who have already sustained one such fracture represents a more feasible option.

In the case of vertebral osteoporosis, small deformations can be measured with sufficient frequency to make studies of fracture feasible. As discussed in section 5.4, minor deformities may be of less clinical relevance than overt vertebral collapse. Evidence should be provided that these vertebral deformities constitute an adequate proxy at least for the direction of effects on overt vertebral fracture or on height or quality of life. In addition, because the risk of vertebral fracture decreases with time, studies of controls must be of a similar duration to studies in treated patients.

The appropriate documentation of fractures is detailed in section 5.4.

7.4.2 Increase in bone mass

Studies intended to show an increase in bone mass should be of sufficient duration to reveal the effects of intervention independently of any transient state. For inhibitors of bone turnover, this may require exposure for 4 or 5 years, part of which may be in phase IV studies. Where no transient state is evident, studies should document continuous and constant increments in bone mass on at least three consecutive occasions. Claims concerning the restoration of bone mass should be substantiated by appropriate evidence at clinically relevant sites.

In severe osteoporosis the architecture of cancellous bone is abnormal, and an increase in bone mass may not lead to a proportional reduction in fracture frequency. This consideration may be less relevant, however, when osteoporosis is due to some secondary cause such as glucocorticoid toxicity, when trabecular connectivity of cancellous bone is relatively well preserved. In postmenopausal osteoporosis increases in bone mass do not necessarily restore skeletal strength, and claims should be reinforced by preclinical and, where appropriate, clinical evidence that the quality of bone formed is normal or that fracture frequency is reduced.

7.4.3 Prevention of bone loss or reduction in the rate of bone loss

The term “prevention of bone loss” is intended also to include the small increments in skeletal mass associated with the use of inhibitors of bone turnover. For some patients the prevention of further loss of bone or the attenuation of loss may be the only practical or acceptable method of treatment. To a large extent, the maintenance of the status quo is the least satisfactory therapeutic option, as it implies an unchanging risk of fracture. The benefits of this approach should be maximized by applying them as early as possible in the natural history of the disorder. Where treatments reduce bone turnover, evidence of therapeutic benefits should be supplemented with evidence that the reduction in turnover is not abnormally suppressed or associated with an increased risk of fracture.

7.5 Study design

7.5.1 Duration

Trials should be of sufficient duration to avoid the misinterpretation of transient states. A *minimum* of 3 years is recommended for phase III studies, but it should be noted that duration of study is not necessarily synonymous with duration of intervention. Where transient states are documented, longer-term studies are also needed for treatments that are intended for longer-term use (5–10 years) but may be undertaken in phase IV. It is also important to document the rate of offset of action in order to determine whether catch-up bone loss occurs. It is recommended that the offset of effect should be studied for at least 1 year in phase III and possibly longer in phase IV.

7.5.2 Site of measurement

Bone mass should be measured at at least two sites, which should be appropriate for the population under study (see section 5.3).

7.6 Other types of studies in severe (established) osteoporosis

7.6.1 Reduction of extraskeletal risks in osteoporosis

Studies that examine the effects of interventions on extraskeletal risks such as falls or severity of falls should be as rigorously controlled as studies of pharmacological intervention. Double-blind studies are not always feasible but every effort should be made to ensure adequate randomization and blinding of the analysis from the treatment allocation. Fracture is the appropriate end-point.

7.6.2 Symptomatic end-points

Since analgesics and nonsteroidal anti-inflammatory drugs are used to alleviate pain and other related symptoms of osteoporosis, the extent and frequency of drug intake may be used as a secondary end-point. However, several bone-active agents may reduce the morbidity associated with osteoporotic fractures by reducing either fracture rates or the complications of incident fractures, and these changes may be appropriate secondary end-points. For some interventions, assessment of neuromuscular function may be an appropriate secondary end-point.

7.7 Safety and clinical pharmacology

The use of pharmacological agents is expected to be high among the elderly (aged 65 years or more). In such circumstances assessment of subpopulations,

7. PHASE III STUDIES IN SEVERE (ESTABLISHED) OSTEOPOROSIS

e.g. those aged over 80 years, is appropriate if such studies have not been undertaken in phase II. Agents that are subject to hepatic metabolism and/or renal excretion may pose special problems in the elderly, and appropriate pharmacodynamic and kinetic analysis may be appropriate in the context of phase III studies.

Population kinetics and dynamics can be used for screening the effect of a large number of variables not otherwise studied individually in traditional drug–disease or drug–drug interactions. In accord with the recommendations of the International Conference on Harmonisation, exposure should be studied in 300–600 patients, giving 95% confidence of observing adverse reactions at a frequency of 0.5–1.0% uniquely in the intervention groups and not in the background population. Total exposure should exceed 1500 patient-years.

8.

Phase III studies in osteoporosis without fragility fractures

8.1 Introduction

A claim for efficacy should be based on at least two independent, well controlled studies or multicentre studies (see section 7.1). Multiple dose levels are not normally necessary where justification for the proposed dose shown has already been otherwise established.

8.2 Population studied

Populations may comprise otherwise healthy men and/or women with low bone density (below an appropriate cut-off value) but without fragility fractures. For women, the cut-off is a bone mineral value at the site of interest 2.5 SD or more below the mean of the young healthy adult population. For men, the same absolute cut-off may be used or a bone mineral value that gives a fracture probability in excess of an acceptable norm based on bone mineral density with or without other risk factors. All criteria should be defined.

It is recommended that trials in osteoporosis should be appropriate to the claim e.g. targeted at the time of the menopause or at the elderly.

In trials in the elderly, there should be evidence that patients are of adequate nutritional status; patients with vitamin D deficiency should not be included. Where broad age ranges are envisaged, population pharmacodynamics and/or pharmacokinetics may be appropriate to elucidate the interactions of age and vitamin D status.

8.3 Basic conditions

Care should be taken to exclude patients with nutritional disorders such as osteomalacia unless the agent being studied is intended to correct these deficiencies or to counteract their skeletal effects. Otherwise patients should have normal nutrition with respect to calcium and vitamin D. In some situations, a “run-in” phase to correct nutritional deficiencies may be considered; in others, calcium and/or vitamin D supplements may be given with test medication where considered appropriate, and, if efficacy is proved,

the applicant may be required to recommend that the agent be used in conjunction with calcium and/or vitamin D.

8.4 End-points

The primary end-points are those listed in section 7.4. End-points other than fracture may be more appropriate where clinical studies of fracture have already been undertaken in severe (established) osteoporosis and outcome has been positive. As for studies in severe (established) osteoporosis, the selection of end-points used will influence the indication for which a licence may be granted.

It is essential to measure bone mineral at sites of clinical relevance. Bone mass measurements should be made at at least two independent sites (see section 5.3).

Although postmenopausal osteoporosis involves the whole skeleton, rates of loss may differ markedly from site to site, partly because of differences in turnover. Treatment may have different, or even opposite, effects at different sites. In the immediate postmenopausal period fractures are unlikely, and the prevention or reduction of bone loss at specific sites is an appropriate end-point for efficacy.

Fractures should be documented as outlined in section 5.4.

8.5 Study design

Where studies have been previously undertaken in severe (established) osteoporosis with a positive outcome, and the time-course of onset of effect has been well characterized, the duration of exposure may be less than 3 years. However, where the duration of treatment is required to exceed this value, long durations of trial exposure become necessary.

Trials should be of sufficient duration to avoid the misinterpretation of transient states. A duration of 3 years or more is suggested for phase III studies. Where transient states are documented, longer-term studies are also needed in phase IV where treatments are to be of longer duration (5–10 years). It is also important to document the rate of offset of action and to determine whether catch-up bone loss occurs; for rapidly effective anabolic regimens, this may be done during phase III, otherwise in phase IV. Other aspects of study design are discussed in section 5.8.

8.6 Safety and clinical pharmacology

The use of pharmacological agents is expected to be high among the elderly (aged 65 years or more). In such circumstances assessment of subpopulations, e.g. those aged over 80 years, is appropriate if such studies have not been undertaken in phase II. Agents that are subject to hepatic metabolism and/or renal excretion may pose special problems in the elderly, and appropriate

pharmacodynamic and kinetic analysis may be appropriate in the context of phase III studies.

Population kinetics and dynamics can be used for screening the effect of a large number of variables not otherwise studied individually in traditional drug–disease or drug–drug interactions. In accord with the recommendations of the International Conference on Harmonisation, exposure should be studied in 300–600 patients, giving 95% confidence of observing adverse reactions at a frequency of 0.5–1.0% uniquely in the intervention groups and not in the background population. Total exposure should exceed 1500 patient-years.

9.

Phase III studies in low bone mass (osteopenia)

9.1 Introduction

Patients with low bone mass (osteopenia) (1) (see section 3.2) generally have a less severe or immediate clinical problem than those with osteoporosis or severe osteoporosis. Many healthy older women will have bone mass in the osteopenic range; for example, bone mineral density will be more than 1 SD below the young adult mean in 50% of women by age 60 years. In individuals below the age of skeletal maturity, osteopenia might be defined as a Z-score of -1 or less; that is, 1 or more SD below the age- and sex-matched reference range. The implications of these considerations are:

- the target population for interventions in this category may be large;
- patients in this category will be asymptomatic and will generally have lower fracture risk than those with osteoporosis or severe osteoporosis;
- diagnostic thresholds may differ from intervention thresholds.

9.2 Population studied

The target population for the study intervention should be defined as discussed in section 5.8.1 and according to physiological status. Menopausal status should be carefully defined, and it may be appropriate to stratify study populations according to a range of the time since menopause, which may have an effect on the rate of change in bone mass.

Appropriate consideration should be given to limiting the target population for study, for instance by including only osteoporosis of specified secondary cause (corticosteroid-induced, immobilization-induced, etc.) or by defining intervention thresholds in terms of bone mineral density more restrictively than diagnostic thresholds for osteopenia.

Clinical trials in osteopenia should be conducted in both women and men, although studies in women have a higher priority because of the higher incidence of osteoporotic fractures. Men may be included in some studies, such as of glucocorticoid-induced osteopenia, if the study population is stratified by sex and power is sufficient for subgroup analysis; otherwise they should be studied in separate trials (see sections 5.8.1 and 7.2).

The nutritional status of both the study sample and the target population should be defined and should be taken into account when selecting a potential target population for an intervention. For example, it would be scientifically and ethically inappropriate to evaluate drug therapy for osteopenia in a malnourished population.

9.3 Basic conditions

In general, trial subjects should be representative of the environmental and behavioural characteristics of the target population (see section 5.8.1). This is especially pertinent for lifestyle factors thought to influence bone mass or fracture risk. Where individuals deviate grossly from the norm in these characteristics, for example by virtue of a high level of weight-bearing physical activity or extreme dietary faddism, moderation of the behaviour, stratification, or exclusion from the trial may be considered.

Other nutritional considerations are covered in section 7.3.

9.4 End-points

Changes in bone mass will generally be the major end-point in therapeutic trials for osteopenia. If preclinical evaluation demands other end-points, these should normally have been included in separate trials in patients with osteoporosis or severe osteoporosis. Bone mass should normally be assessed at clinically relevant skeletal sites (see section 5.3), but measurements at additional sites may be desirable for evaluating safety (see section 9.7).

Although the use of fracture data as an efficacy end-point is desirable, fracture incidence will be low in most osteopenic target populations, to the extent that study size could become prohibitive. In some circumstances, however, fracture risk may be relatively high and anti-fracture efficacy may then be a primary end-point. Examples are: trials of fracture prevention in patients who are osteopenic at the time of heart or liver transplantation; and trials of hip fracture prevention in elderly populations who have a high underlying risk of hip fracture.

As is the case for studies in osteoporosis the end-points selected will influence the claim that can be made for a treatment.

A number of secondary end-points may also be specified for clinical trials in osteopenia (see sections 5.7 and 7.4). Depending on the data available from preclinical studies and from trials in osteoporosis or severe osteoporosis, bone histology may be a desirable secondary end-point, although careful justification would be needed for this invasive investigation in asymptomatic subjects.

9.5 Study design

The randomized, controlled, double-blind trial is usually the method of choice for efficacy studies of medium duration (up to 3 years of intervention). Where

an established therapy is available for the target population, a comparative study may be performed, with control subjects receiving the established treatment. If a combination of interventions is to be evaluated, a factorial design is usually optimal. Open extension of randomized, controlled trials, or other cohort designs, should also be considered, either in phase III (see section 9.6) or in phase IV.

9.6 Duration of studies, including interruption of treatment

The recommended duration of studies is discussed in section 7.5. Longer-term studies may be undertaken if required by the aims of the study. It is emphasized that, in an osteopenic population, the objectives of intervention are the long-term enhancement or preservation of bone mass and the prevention of fractures. Given the chronicity of osteopenia and increased fracture risk, studies of more than 3 years' duration should be undertaken to assess both persistence of efficacy and safety of long-term exposure and of repeated treatments. These considerations are particularly important where interventions are targeted at relatively young adult populations. In many instances long-term investigations will be phase IV studies. Sometimes open extension of randomized controlled trials, or other cohort designs, will be feasible in phase III to permit observations to be made over an adequate period. Where treatments of limited duration are envisaged, the offset of therapeutic effect on cessation of treatment should be studied in phase III or phase IV (see section 5.8.4).

9.7 Safety

Before an intervention is evaluated in an asymptomatic osteopenic population, there should be substantial evidence of its safety, possibly from phase III studies for other indications (e.g. osteoporosis). Clinical trials in osteopenia raise particular issues in addition to the general conditions (discussed in section 5.8.7). Trials in osteopenia, where bone mass is the primary end-point, allow effects of clinically significant magnitude to be detected reliably with a relatively modest number of subjects. The safety of interventions evaluated for osteopenic populations is of particular importance, so that the study size should also provide sufficient statistical power to detect a small excess of clinically significant adverse events attributable to the study intervention. Some agents have the potential to affect the distribution of bone between cortical and cancellous sites, or between the axial and the appendicular skeleton, and regional or whole-body estimates of bone mineral may be appropriate measurements of both safety and efficacy.

10.

Phase III clinical trials in subjects with normal bone mass

10.1 Introduction

Adult subjects with normal bone mass have bone mass measurements within 1 SD of the young adult reference mean (1). Their baseline risk of fragility fractures is low. The objective of clinical trials in these subjects is to assess the safety and efficacy of interventions:

- in preventing the bone loss and increased fracture risk that would otherwise be expected to occur;
- in augmenting bone mass in a study sample representing a general population in which an increase in mean bone mass may reduce overall fracture incidence;
- in mitigating the effect of a risk factor for fracture which is independent of bone mass.

It is likely that most trials pursuing the first objective will involve pharmacotherapy, most pursuing the second objective will involve non-pharmacological interventions, and those pursuing the third objective may evaluate a wide range of interventions.

It is noteworthy that target populations for interventions in this category may be very large, even extending to the general population. At the other extreme, a small patient population may be targeted because of its exposure to a specific (and often powerful) risk factor for bone loss.

10.2 Population studied

In general the same considerations as in sections 5.8.1 and 9.2 will apply. Where the objective is prevention of bone loss, the population studied should be clearly recognized as being at risk of bone loss over the period of study. Risk should be defined carefully. For agents with adverse effects on skeletal metabolism, both the dose of the agent and the duration of exposure must be defined at study entry.

10.3 Basic conditions

Care should be taken to exclude patients with nutritional disorders such as osteomalacia unless the agent being studied is intended to correct these deficiencies or to counteract their skeletal effects. Otherwise patients should have normal nutrition with respect to calcium and vitamin D. In some situations, a “run-in” phase to correct nutritional deficiencies may be considered; in others, calcium and/or vitamin D supplements may be given with test medication where considered appropriate, and, if efficacy is proved, the applicant may be required to recommend that the agent be used in conjunction with calcium and/or vitamin D.

10.4 End-points

Similar considerations to those in sections 7.4 and 9.4 will apply. Where the study sample is derived from a normal healthy population, the development, validation, and use of the most minimally hazardous end-points is to be encouraged. For example, techniques that allow prediction of fracture risk (or change in fracture risk) without exposing subjects to ionizing radiation are highly desirable (see section 10.7).

10.5 Study design

The considerations raised in sections 7.5 and 9.5 apply. Where trials are undertaken in healthy populations without coexisting morbidity, the duration of studies should be clinically appropriate to permit an evaluation of the place of the intervention in the community. This may demand trials of more than 3 years' duration, as well as investigation of population kinetics and dynamics, to identify those most likely to benefit. In a normal, healthy population, the risk-benefit ratio of intervention must be highly favourable, and this has implications for sample size.

10.6 Duration of studies, including interruption of treatment

The considerations raised in sections 7.5 and 9.6 apply. Although healthy populations demand long-term studies, exposure of some target populations to a specific risk for bone loss may be for a defined, limited period, and study duration should then take account of the duration of risk exposure.

10.7 Safety

The considerations raised in sections 5.8.7 and 9.7 apply. Because the risk–benefit ratio of interventions evaluated in a normal, healthy population must be highly favourable, clinical trials in this category must be of a design, size, and duration that allow adverse effects to be detected at very low incidence rates.

11.

Phase IV studies

11.1 Introduction

Phase IV studies in this field are particularly important because some agents used in the prevention and treatment of osteoporosis may need to be used for long periods, and perhaps for life. Where such long-term treatment is envisaged, studies are needed to demonstrate the continuing beneficial effects of the agent used, and its safety, throughout the treatment period. It may be appropriate, therefore, for regulatory authorities to consider granting marketing authorization on the basis of limited exposures and extending the permitted exposure when appropriate phase IV studies have been completed.

11.2 Aims

11.2.1 Efficacy with continued treatment

The scope of phase IV studies may include several aspects of safety and efficacy. For many agents, steady-state responses to intervention are unlikely to have been completely characterized at the time of authorization because of the slow turnover of bone mineral. In such cases, long-term studies may be necessary to determine whether there is continued effect, a stabilization of bone mineral, or a loss of effect. Adequate characterization of continued effect or stabilization should normally be based on three consecutive measurements showing unaltered rates of change of bone mineral (see section 7.4.2).

11.2.2 Efficacy after discontinuation of treatment

It is usually necessary to characterize the offset of effect of an agent when treatment is stopped. This will normally have been evaluated in phase III (see section 7.5), but longer-term studies may be required in phase IV, particularly in the case of agents that are retained in the skeleton with a prolonged half-life of elimination (which might contribute to a persistence of effect).

11.2.3 Need for retreatment

Some treatment regimens may be devised for a limited period of exposure, with authorization being granted on that basis. It may be necessary to devise phase IV studies to identify patients or situations in which retreatment is considered to be appropriate.

11.2.4 Safety and other end-points

Treatment and prevention of osteoporosis will usually require treatments of prolonged duration. Although there will normally have been several hundred patient-years of exposure before marketing authorization is granted on the basis of efficacy, even serious adverse reactions may remain undetected if they occur with an incidence less than several per cent. The potential for adverse effects and drug interactions is also high where the population is elderly, with appreciable co-existing morbidity and multiple drug exposure. A major aim of all phase IV studies will therefore be the detection of side-effects, including fractures and cancer.

Long-term studies of safety may also be advisable after stopping treatment with test agents that are retained in the skeleton and exhibit a slow terminal elimination phase.

Phase IV studies may also provide an opportunity to study the risk–benefit ratio of an intervention in population settings and the impact on health economics of targeting those at greatest risk or identifying those likely to derive the greatest benefit. They may also allow the modification of any labelling that is particularly restrictive as a result of uncertainties at the time of authorization.

Irrespective of the study design, all phase IV studies should document fracture events (see section 5.4), and a prospective plan for statistical analysis should be developed.

11.3 Type of study

11.3.1 Uncontrolled trials

It is a common misperception that phase IV studies are uncontrolled trials. Generally, however, this is the case only in dealing with safety issues, when randomization is impossible; the main aim is then to observe a large number of patients exposed to the treatment, in order to estimate the incidence of adverse reactions. Even in these circumstances it may be useful to observe untreated controls, or controls given a different treatment, which will facilitate the assessment of drug-related adverse events. As with all studies, uncontrolled trials should be performed according to a predefined study protocol that specifies inclusion criteria, the type of data to be collected, and the duration of the study.

Uncontrolled studies may also be appropriate to address long-term efficacy and the effects of stopping treatment, and to clarify risk–benefit assessments.

11.3.2 Cohort study

A cohort study in phase IV is aimed at detecting specified adverse reactions to a drug and estimating their incidence. The observed cohort comprises patients treated with the agent and controls who are untreated or differently treated. The hypothesis to be tested should be explicit in the study protocol, which should also provide a clear statement of inclusion and exclusion criteria. The non-randomized nature of this type of prospective trial is its main drawback, making it less suitable for assessment of the efficacy than a randomized controlled trial.

11.3.3 Case–control study

A retrospective case–control study may be used to determine whether or not rare adverse events are drug-related, and may be particularly useful for examining the consequences of long-term exposures and any adverse reactions arising after exposure. The study protocol should describe the hypothesis to be tested, the method by which cases and the controls are selected, the technique of data acquisition, and the co-variables recorded for risk adjustment.

Since the selection of controls is prone to bias, the results obtained with this type of study should be interpreted cautiously.

11.3.4 Randomized trials

Although randomized controlled trials can be used to address all the questions that remain when marketing authorization is granted, they may sometimes be difficult or impossible to perform. There are particular difficulties in using placebo controls because of the perception of patients and physicians that it may be unethical to withhold treatment. The use of alternative treatments instead of a placebo may solve the problem.

11.4 Sample size and statistical considerations

The number of patients admitted into a trial depends on the various statistical considerations outlined previously (see section 5.8.5). For uncontrolled studies there may be no formal way to calculate the sample size because of the lack of hypothesis. However, the incidence of adverse reactions in the trial should be reported, together with the estimates of power.

12.

Glossary of terms

The definitions given below apply specifically to the terms used in these guidelines; they may have different meanings in other contexts.

animal models

Animal models are simplified but imperfect representations of more complex systems found in humans. They include:

- induced or experimental systems that attempt to reproduce conditions found in humans;
- spontaneous or natural models that are recognized as being similar to some condition in humans;
- negative or non-reactive models that are the normal counterparts of a disease model; and
- orphan models that are animal diseases for which no human or animal counterpart is known (16).

biomechanical testing

The performance of load–deformation tests in which the bone is loaded in a manner appropriate to its function (tension, compression, torsion, bending). The purpose is to determine the mechanical properties of the bone. A load–deformation curve may be converted to a stress–strain curve, allowing measurement of bone material properties by separating geometric effects. Specific parameters have been defined by Martin & Durr (17).

bone formation

Part of the remodelling process in adults, which involves chemotaxis, proliferation, and differentiation of osteoblasts, followed by formation of mineralized bone and cessation of osteoblast activity. Additionally, the term describes intramembranous, endochondral, and appositional bone formation in the growing individual (18).

bone mineral content (BMC)

The total amount of bone mineral measured within a defined region of interest.

bone mineral density (BMD)

The total amount of bone mineral measured within a defined region of interest, divided by the projected area of bone (areal BMD) or the determined volume within which that bone mineral is located (volumetric BMD).

bone resorption

The removal of bone largely, and possibly exclusively, by osteoclasts. It is the initial step in the remodelling process and involves osteoclast activation and the removal of bone by the action of acid and enzymes secreted by osteoclasts (18).

bone turnover

The levels of bone formation and resorption, most of which are the result of remodelling in the adult.

cortical porosity

The amount of space within Haversian canals, determined by the balance between bone resorption and formation. In adults, increased bone turnover is associated with increased cortical porosity.

coupling (of bone remodelling)

The attraction of osteoblasts to sites of prior bone resorption during the remodelling sequence.

dual-energy X-ray absorptiometry (DXA)

A photon-absorption technique for measuring BMC and areal BMD using an X-ray generator as a source of photons which can be quantified by using two energy levels (1).

fragility fracture

A fracture caused by injury that would be insufficient to fracture normal bone; the result of reduced compressive and/or torsional strength of bone.

hip axis length

A linear measurement along the femoral neck axis as defined by DXA analysis software, from below the lateral aspect of the greater trochanter to the inner pelvic brim. It is reported to be an independent predictor of hip fracture risk in women (19).

histomorphometry of bone

Measurement of morphological components of bone by light microscopy (static histomorphometry), sometimes using fluorescent labels — administered at timed intervals — which are incorporated into forming bone and provide information about the rate of bone formation and, by derivation, about other parameters of bone metabolism (dynamic histomorphometry) (20). Specific parameters and recommended notation have been defined by Parfitt et al. (11).

lamellar bone

Cortical or cancellous bone showing a clear lamellar pattern under polarized light, caused by birefringence due to the arrangement of collagen fibrils in alternating orientations (20).

microarchitectural deterioration of bone tissue

Changes in microscopic structure of bone leading to reduced mechanical strength. In cancellous bone these changes include trabecular thinning, perforation, and transection, and in cortical bone increased cortical porosity. Microdamage or fatigue fractures may also reduce mechanical strength (1).

mineralization defect

Impairment in the mineralization of newly-formed bone matrix, leading to a reduced mineralization rate and an increase in non-mineralized osteoid tissue (18).

modelling (of bone)

The process of unapposed bone formation or bone resorption by which the overall shape of bone is changed in response to physiological and/or mechanical influences (20).

neutron activation analysis

Analysis of the effects of irradiating an area of bone with neutrons, which incites many of its constituent elements to become radioactive. Calcium can be

identified and quantified by its characteristic gamma-ray emissions after neutron activation (*1*).

osteoid seam width

A measurement made on bone biopsies to determine the presence or absence of osteomalacia.

osteomalacia

A disorder characterized by impaired mineralization of newly-formed bone matrix. *See* mineralization defect.

osteopenia

In the context of this document, a value for BMD or BMC more than 1 SD, but less than 2.5 SD, below the young adult mean (*1*). The term is often also used to describe radiographic or absorptiometric appearances of low BMC of any cause.

osteoporosis

In the context of this document, a value for BMD or BMC 2.5 SD or more below the young adult mean value (*1*).

peak bone mass

The maximum value for BMC or BMD achieved early in adult life.

periosteal apposition

Bone formation due to apposition at periosteal surfaces that results in bone growth.

peripheral quantitative computed tomography (pQCT)

A quantitative computed tomography technique applied at appendicular sites, for example the forearm, to measure total, cancellous, and cortical BMC/BMD and dimensions.

phase I studies

The first trials of a new active agent or new formulation in humans, often carried out in healthy volunteers. Their purpose is to establish a preliminary evaluation of the safety, and the pharmacokinetic and where possible, pharmacodynamic profile of the active ingredient in humans.

phase II studies

Trials performed in a limited number of subjects and often, at a later stage, of a comparative design (e.g. placebo control). Their purpose is to demonstrate therapeutic activity and to assess the short-term safety of the active ingredient in patients suffering from a disease or condition for which the active ingredient is intended.

phase III studies

Trials in large patient groups with the purpose of determining the short- and long-term safety and efficacy of the formulation(s) of the active ingredient, and of assessing its overall and relative therapeutic value.

phase IV studies

Trials carried out on the basis of the product characteristics for which marketing authorization was granted, normally in the form of post-marketing surveillance or assessment of therapeutic value or treatment strategies.

population pharmacodynamics

The study of the biochemical and physiological effects and mechanism of action of drugs in an exposed population, and the specific statistical analysis of the data.

population pharmacokinetics

The study of the absorption, distribution, biotransformation, and excretion of drugs in an exposed population, and the specific statistical analysis of the data.

postmenopausal

After the onset of permanent loss of reproductive ovarian function. In women with an intact uterus, this state is diagnosed clinically after 1 year or more of amenorrhoea where no cause, other than menopause, is apparent. In clinical trials it should be confirmed by measurement of gonadotropin and estrogen levels.

quality of bone

Microstructural, ultrastructural, and compositional characteristics of bone that influence its biomechanical material properties independently of mass or density.

quantitative computed tomography (QCT)

An extension of computed tomography imaging whereby the image is quantified under appropriate conditions to give a measure of volumetric BMD (g/cm^3 or g/ml) (1).

quantitative radiography

A technique of radiographic photodensitometry that depends on measuring the optical density of X-ray films of the bones.

radiogrammetry

A technique using linear measurements of cortical bone on X-ray films to measure cortical width, from which cortical area can be derived (1).

remodelling (of bone)

The local and discrete process in the adult skeleton by which bone is continually resorbed and reformed by the coordinated actions of osteoclasts and osteoblasts on trabecular surfaces and in Haversian systems (18). There is a spatial and temporal coupling of bone resorption, followed by formation and mineralization of bone.

severe (established) osteoporosis

A value for BMC or BMD 2.5 SD or more below the young adult mean in the presence of one or more fragility fractures (1).

t-score

The value for BMD or BMC expressed as the number of SD by which an individual result differs from the mean value for young adults of the same sex (1).

trabecular connectivity

The architectural continuity of cancellous bone, a determinant of its mechanical strength, which is diminished by the transection of trabeculae due to osteoclastic resorption (18, 20).

ultrasonic attenuation

A transmission ultrasound technique using broadband ultrasound. Attenuation in bone is plotted against frequency and the slope of the linear portion of the plot is used to characterize bone (1).

woven bone

Bone in which collagen fibrils viewed under polarized light are arranged in a disorganized orientation that compromises the strength of the bone. This structure occurs during the formation of primary bone and in states with a high bone formation rate (20).

Z-score

The value for BMD or BMC expressed as the number of SD by which an individual result differs from the mean value for individuals of the same age and sex (1).

References

1. *Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group.* Geneva, World Health Organization, 1994 (WHO Technical Report Series, No. 843).
2. Avouac B, Caulin F. Clinical pharmacology of osteoporosis and recommendations for clinical studies. *Recent trends in clinical pharmacology*, 1991, 205:3–29.
3. Kanis JA, Caulin F, Russell RGG. Problems in the design of clinical trials in osteoporosis. In: St J Dixon A, Russell RGG, Stamp TCB, eds. *Osteoporosis: a multidisciplinary problem.* New York, Grune & Stratton, 1983:205–222 (Royal Society of Medicine, International Congress Symposium Series).
4. Kanis JA, Geusens P, Christiansen C. Guidelines for clinical trials in osteoporosis. A position paper of the European Foundation for Osteoporosis and Bone Disease. *Osteoporosis international*, 1991, 1:182–188.
5. *Guidelines for preclinical and clinical evaluation of agents used in the prevention or treatment of postmenopausal osteoporosis.* Rockville, MD, Division of Metabolism and Endocrine Drug Products, Food and Drug Administration, 1994.
6. FDA HRT Working Group. Guidance for clinical evaluation of combination estrogen/progestin-containing drug products used for hormone replacement therapy of postmenopausal women. *Menopause*, 1995, 2:131–136.
7. Reginster JY et al. Recommendations for the registration of new chemical entities used in the prevention and treatment of osteoporosis. *Calcified tissue international*, 1995, 57:247–250.
8. Ishii J. Summary of the draft Japanese guidelines for the evaluation of agents for treatment osteoporosis. In: *Proceedings of the 11th Education Seminar of the Japan Osteoporosis Foundation.* Tokyo, Lifescience Publishing, 1995: 14–18.

9. *Note for guidance on involutional osteoporosis in women*. London, Committee for Proprietary Medicinal Products, September 1997 (CPMP/EWP/552/95).
10. Bonjour JP, Ammann P, Rizzoli R. Importance of preclinical studies in the development of drugs for the treatment of osteoporosis. *Osteoporosis international* (in press).
11. Parfitt AM et al. Bone histomorphometry: standardization of nomenclature, symbols, and units. *Journal of bone and mineral research*, 1987, 2:595–610.
12. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. *American journal of medicine*, 1993, 94:646–650.
13. Assessing vertebral fractures. National Osteoporosis Foundation Working Group on Vertebral Fractures. *Journal of bone and mineral research*, 1995, 10:518–523.
14. Kanis JA. What constitutes evidence for drug efficacy in osteoporosis? *Drugs and aging*, 1993, 3:391–399.
15. *The use of essential drugs. Seventh report of the WHO Expert Committee*. Geneva, World Health Organization, 1997 (WHO Technical Report Series, No. 867).
16. Davidson MK, Lindsey JR, Davis JK. Requirements and selection of an animal model. *Israeli journal of medical science*, 1987, 23:551–555.
17. Martin RB, Durr DB. *Structure, function and adaptation of compact bone*. New York, Raven Press, 1989.
18. Mundy GR. *Bone remodelling and its disorders*. London, Martin Dunitz, 1995.
19. Faulkner KG et al. Simple measurement of femoral geometry predicts hip fracture: the study of osteoporotic fractures. *Journal of bone and mineral research*, 1993, 9:1211–1217.
20. Eriksen EF, Axelrod DW, Melsen F. *Bone histomorphometry*. New York, Raven Press, 1994.

Acknowledgements

WHO is grateful to Dr T. Einhorn, Mount Sinai Medical Center, New York, and to Dr H. Bone, Henry Ford Hospital, Detroit, for their review of this document in its early stages.

The writing group comprised Professor J.-P. Bonjour (Geneva), Dr M. Couper (WHO, Geneva), Dr S. Dutta (Rockville, MD), Dr G. Fracchia (Brussels), Professor U. Gundert-Remy (Göttingen), Dr J. Idänpään-Heikkilä (WHO, Geneva), Mr M. Ikeda (Tokyo), Dr A. Kalache (WHO, Geneva), Professor J.A. Kanis (Sheffield), Dr N. Khaltsev (WHO, Geneva), Dr L. Lutwak (Rockville, MD), Dr O. Meirik (WHO, Geneva), Professor P. Meunier (Lyon), Dr T. Nakamura (Kitakyushu, Japan) and Dr J.D. Wark (Melbourne).

The text of these guidelines was discussed and reviewed in a consultation on 19–21 May 1997 in Geneva. The participants were: Professor F. Ben Khalifa (Tunis), Professor J.-P. Bonjour (Geneva), Professor I. Darmansjah (Jakarta), Dr V. Fuchs (Berlin), Professor U. Gundert-Remy (Berlin), Professor J.A. Kanis (Sheffield), Dr L. Lutwak (Rockville, MD), Dr T. Nakamura (Kitakyushu), Dr J.A. Sisson de Castro (Porto Alegre), Professor J.D. Wark (Melbourne), Dr F. Caullin (Antony), Dr A. Santora (Tahway, NJ), Dr T. Sernau (Mannheim); Secretariat: Dr J. Idänpään-Heikkilä, Dr M.R. Couper, Dr A. Kalache, Dr O. Meirik, Dr N. Khaltsev (WHO, Geneva).

Supported with the technical and financial assistance of the European Foundation for Osteoporosis and Bone Disease.

Annex

Recommendation for future guideline development

There is a clear need for prospective research, including clinical trials, to evaluate the efficacy and safety of interventions applied during growth and development to augment peak bone mass and stable adult bone mass, and to improve long-term fracture risk. It might be envisaged that most of these interventions would fall within the intervention categories “prevention” and “prevention of osteoporosis in subjects with osteopenia” (see section 3.4). Current research in these populations involves non-pharmacological interventions such as modulation of calcium nutrition and of physical activity. These types of studies have important implications for the formulation of evidence-based public health recommendations. Many of the principles for the evaluation of anti-osteoporosis measures described in these guidelines are applicable to studies in childhood and adolescence. However, it was considered that research involving these populations should not be dealt with explicitly here; a number of issues specific to this age group need to be addressed and are beyond the scope of these guidelines. The development of appropriate, validated, short-term surrogate end-points for bone loss in later life is a critical goal, and it is strongly recommended that guidelines for osteoporosis-related research during childhood and adolescence be developed.



WHO LIBRARY



00061159