

**National Institute for Health and Clinical  
Excellence**

**Systematic reviews of clinical effectiveness  
prepared for the guideline**

**'Osteoporosis: assessment of fracture risk  
and the prevention of osteoporotic  
fractures in individuals at high risk'**

**September 2008**

**National Collaborating Centre for Nursing and  
Supportive Care**

# Contents

Contents .....	2
1 Review methods .....	3
1.1 Selection criteria .....	3
1.2 Search strategy .....	8
1.3 Data extraction .....	8
1.4 Appraisal of methodological quality .....	9
1.5 Data synthesis .....	12
1.6 Grading evidence .....	15
2 Pharmacological interventions to prevent osteoporotic fractures.....	18
2.1 Bisphosphonates .....	21
2.2 Strontium ranelate .....	85
2.3 Calcitonin and parathyroid hormone.....	91
2.4 Sex hormone therapies.....	121
2.5 Vitamin and mineral supplements.....	145
2.6 Cross review comparisons.....	193

## The following are available as separate files

References

Appendices

Appendix I: Pharmacological interventions, doses in the BNF as at March 2008

Appendix II: Search strategy

Appendix III: Included studies

Appendix IV: Excluded studies

Appendix V: Methodological quality of studies

Appendix VI: Grade evidence tables

# 1 Review methods

This section describes the methods of reviewing that were common to all reviews of intervention studies. Further specific details are given in the individual reviews.

## 1.1 Selection criteria

The following selection criteria were applied to studies to determine their suitability for inclusion in the reviews:

### 1.1.1 Types of studies

For intervention studies, the randomised clinical trial (RCT) was the primary trial design. Quasi-randomised studies could only be included if there was no other evidence (for example, allocation by alternation or date of birth).

Studies were excluded if there were fewer than 10 participants in each arm.

Studies were limited to the English language, with the exception of studies translated for Cochrane reviews or as directed by the GDG, but the date was not restricted.

### 1.1.2 Types of participants

Study participants were those at risk of osteoporotic fractures, who may or may not have had a previous fracture, including people with osteoporosis, osteopenia or normal bone mineral densities (BMDs). Also included were studies in people with glucocorticoid-induced osteoporosis and those in men, or men and women.

The World Health Organization (WHO) has established diagnostic criteria for osteoporosis based on the measurement of BMD (WHO 1994). Osteoporosis is defined as a BMD value that is 2.5 standard deviations (SD) or more below the mean BMD of the young adult population (specific for gender) at their peak bone mass; that is, a T-score of less than or equal to  $-2.5$  SD.

Osteopenia is defined as a T-score between  $-2.5$  and  $-1.0$  SD, and people with normal BMD are classified as those with a T-score within 1 SD of zero.

For the purposes of this guideline, for women, the T-score was calculated from BMD measurements using the Nhanes III reference: BMD at the total femur for non-Hispanic white women aged 20–29 years: BMD mean 0.956 g/cm<sup>2</sup>, SD 0.123 g/cm<sup>2</sup>. Other values from the same source were used for men.

$$\text{T-score} = [\text{BMD (at femoral neck)} - \text{BMD ref}] / \text{SD ref}$$

Population groups of postmenopausal women, men, men and women mixed, and people who have taken glucocorticoids were included in each intervention review. However, these groups were treated as separate subgroups for analysis purposes. Each review included all people, regardless of their T-score and previous fracture history; that is, studies in people with osteoporosis, osteopenia or with normal BMD were combined in each intervention review and not considered as subgroups. This ‘lumping’ of studies was on the basis of the Guideline Development Group’s (GDG’s) assumption that the relative risk (RR) for the prevention of fracture was independent of the absolute risk of fracture.

### **1.1.3 Types of intervention**

The interventions considered varied across reviews and are detailed in appendix I.

Interventions could be applied over different time periods and participants could be followed up after the trial period ceased. However, studies were excluded if the intervention was given for less than 12 months. Interim results were not extracted, and the main outcome measures were recorded at the end of the intervention period (or in follow-up periods). Studies were combined in analysis regardless of their duration of intervention.

Different doses, regimens and routes of delivery were also permitted, but the GDG decided that only those studies using the licensed dose (as reported in the BNF) were to be included. Where possible, if there were different licensed doses randomised to two or more groups, the results were combined. In analysis, all doses, regimens and routes of delivery were combined initially.

Different drugs, including those within the same class, were treated in separate reviews (for example, the bisphosphonates, alendronate and risedronate).

Concurrent interventions given to all participants in the trials were also recorded, particularly calcium and vitamin D intake.

#### **1.1.4 Types of comparisons**

The following comparisons were included:

1. Osteoporosis intervention (**A**) versus placebo
2. **A** versus no intervention
3. **A** plus calcium and/or vitamin D versus calcium and/or vitamin D alone
4. **A** versus calcium and/or vitamin D
5. **A** plus second intervention (**X**) versus **X** alone
6. Within a class of interventions, **A1** versus **A2**
7. Across classes of interventions **A** versus **Y**.

In analyses, the GDG decided that comparisons (1), (2) and (3) could be combined, but (4) was treated separately because calcium and vitamin D could be active interventions; (5) was treated separately because of possible drug interactions.

#### **1.1.5 Types of outcome measures**

For studies of interventions for the prevention of osteoporosis, the following **primary** outcomes were considered.

Incidence of fragility fracture, reported in trials either as primary or secondary outcomes:

- vertebral
- nonvertebral

- hip
- wrist
- proximal femur
- distal forearm
- all fractures
- all clinical fractures.

Fragility fractures are those associated with low trauma; that is, a fracture sustained as the result of a force equivalent to the force of a fall from a height equal to or less than that of an ordinary chair. They are also known as 'osteoporotic fractures'; they should not be confused with fractures in people with osteoporosis.

Studies solely reporting fractures associated with major trauma (for example, road accident) were excluded, but those reporting mixed trauma and nontrauma fractures were included. The latter were distinguished from those reporting nontrauma fractures alone.

The GDG also requested that a list was made of studies that did not report fracture outcomes, but did record the BMD. These studies were recorded in the text and classified as excluded studies (BMD only). The GDG also decided to exclude studies in people with other conditions (for example, liver transplantation patients), with the exception of those treated with glucocorticoids.

Other outcome measures recorded were:

- adverse effects associated with the intervention
- continuation and concordance (compliance).

The incidence of fracture outcome should have been measured in a dichotomous way, as the number of patients with a fracture. Studies recording only the number of fractures were included but not analysed unless there was sufficient information to calculate the number of patients with at least one fracture.

Clinical, or symptomatic, vertebral fractures are those fractures that cause sufficient discomfort for the patient to bring them to the attention of a health professional. They can be identified by X-ray. However, it is also possible for the patient to suffer vertebral fractures that do not cause sufficient discomfort to be reported by the patient, but that can also be identified by X-ray. Fractures that are identified radiographically are termed radiographic or morphometric, and include both symptomatic and asymptomatic fractures. For the most part, vertebral fracture data are likely to relate to morphometric fractures.

Data from one large study that reported both clinical and morphometric fractures suggested that the RR values for the two types of fracture are very similar. The GDG decided to treat clinical and morphometric fractures in the same meta-analysis, with preference given to morphometric fractures. Where appropriate this was examined in sensitivity analyses.

For vertebral fractures, various definitions of radiographic fractures have been developed. Definitions that require a 20% reduction in vertebral height are generally recognised as producing fewer false negatives and false positives than those that only require a 15% reduction. Therefore, data based on a 20% fracture definition were preferred because the reduction in specificity associated with the use of a 15% definition would reduce the perceived efficacy of the intervention in question. The use of a semi-quantitative method also results in greater specificity than the use of a 15% definition alone.

Consequently, the definition of vertebral fracture used was reported and considered in sensitivity analyses, and when both 15 and 20% definitions were given, the 20% definition was used to determine the number of people with a fracture.

Discontinuation of the study medication and concordance with treatment were considered separately. For concordance with treatment, the study should have stated, first, how it was measured, for example pill counts of oral medication and measurement of volume of injectable medication returned at each study visit, and second, recorded the number of participants who did not

comply with the medication regimen according to the authors' definition of concordance.

## **1.2 Search strategy**

The intervention reviews are updated versions of those produced in 2005 for the guideline and technology appraisals. Databases searched were MEDLINE, EMBASE, CINAHL and The Cochrane Library. The MEDLINE search strategy is given in appendix II. This was modified, when necessary, for use in other databases. The search was combined with standard search filters for RCTs and systematic reviews. The search was updated in June 2008.

## **1.3 Data extraction**

Data from included studies were extracted by one reviewer for each review and randomly checked by a second reviewer. They were then entered into a Microsoft Access relational database that had been specially designed for the guideline. The use of the database provided a more structured extraction; for example, only certain choices could be made for some items, although free text fields were also completed. The main advantage of using a database for this purpose is that a large measure of detail can be input, and then an overview obtained using database sorting procedures.

For intervention studies, the following data were extracted.

- Study details: study design; country where trial was conducted; study size; setting; funding.
- Participants:
  - characteristics: age (mean and range); ethnicity, comorbidities; weight/height/body mass index (BMI); smoking; mean time since the menopause; inclusion/exclusion criteria; calcium/vitamin D supplementation
  - type of osteoporosis (osteoporosis, osteopenia, normal BMD); T-score mean and range (and if not given, BMD mean and T-score calculated)
  - type of patient (e.g. postmenopausal women, men)

- type of risk group (not high risk, glucocorticoid-induced osteoporosis, other high risk group)
- presence of one or more prior fractures.
- Interventions:
  - class (for example, bisphosphonate)
  - drug (for example, alendronate)
  - intervention details, duration of intervention; dose; route of delivery (intravenous, oral); frequency of intervention
  - other interventions given concurrently to all patients.
- Comparator: usual care; placebo (details of what it is); other intervention.
- Outcomes: site of fracture measurement; time measured (including follow-up after end of intervention); method of measurement used (validity); means of determination of fracture (radiographic, clinical).

Other data extracted were:

- study quality (see below)
- results for each outcome.

If studies were published more than once, data were extracted from the most recent report in which there were differences; otherwise, all papers were used for data extraction. If studies were reported at different follow-up periods, each follow-up was reported, but the results at the end of the intervention period were used in the main analysis.

Masked assessment, when data extractors are blind to the details of the journal, authors and so on, was not undertaken.

#### **1.4      *Appraisal of methodological quality***

The methodological quality of each trial was assessed by one reviewer and randomly checked by a second.

For randomised trials, the following factors were considered in assessing the potential for bias.

- A priori sample size calculation:

- whether or not this was carried out.
- Method of generation of the randomisation sequence:
  - the means by which interventions are distributed amongst the participants
  - whether the method was reported or unclear (that is, no details given)
  - whether the reported method was adequate, inadequate or partial (table 1).
- Allocation concealment at randomisation:
  - the means of preventing the treatment assignment being known before the time of allocation
  - whether the method was reported or unclear (no details)
  - whether the reported method was adequate, inadequate or partial (table 1).
- Baseline comparability of treatment groups:
  - for relevant risk factors, especially age, T-score distribution, number of existing fractures, BMI, smoking, parental history of fracture.
- Patients stated to be blinded, especially for comparisons with placebo:
  - blinding involves hiding the nature of the intervention from participants, clinicians and treatment evaluators after allocation has taken place
  - blinding may be not be possible depending on the nature of the interventions
  - blinding may be more important for some outcomes than others:
- Outcome assessor stated to be blinded.
- No missing data for each outcome:
  - studies with at least 20% of data missing from any group were considered to be potentially biased, more so if there was differential dropout from any one group or if the missing data were known to be significantly different from the remaining data
  - studies with moderate amounts of missing data (20–50%) were considered in sensitivity analyses
  - those with 50% or more patients missing from any one group were regarded as flawed and not analysed further.
- Intention to treat analysis (ITT):

- trial participants should be analysed in the groups to which they were randomised, regardless of which (or how much) treatment they actually received, and regardless of other protocol irregularities
- all participants should be included regardless of whether their outcomes were actually collected.
- Note, however, that authors have devised different approaches to reporting missing data:
  - ◇ patients who had had a fracture were not followed further, but their positive fracture status is carried forward into the analysis; these are not considered to be missing data; the studies should report that this is what they have done
  - ◇ patients lost to follow-up are analysed as ITT, and the authors assume that those lost to follow-up do not have a fracture event; this is a conservative approach, but may be more critical if the loss to follow-up differs across groups or if the amount of missing data is large; such an approach should be considered in sensitivity analyses
  - ◇ data are analysed for every participant for whom the outcome was obtained ('available case analysis')
  - ◇ data are analysed for every participant for whom the outcome was obtained; the participants were analysed according to the group to which they were assigned ('per protocol analysis').

The studies were analysed using each of these methodological features, and then a summary was made of the studies that might have potential for bias. Studies that had poorly reported the methodological aspects of the study (unclear/unstated categories) were given the benefit of the doubt and were grouped with the adequate or partially adequate studies.

As described below, sensitivity analyses were performed to investigate aspects of methodological quality. If the only studies in an analysis had potential for bias, this was reported in the summary of the evidence.

## **1.5 Data synthesis**

### **1.5.1 Analytical approach**

Meta-analysis of similar trials was carried out using The Cochrane Collaboration's analysis software, Review Manager (Version 4.2). Trials were pooled using a fixed effects model and plotted on forest plots. When there was significant heterogeneity, a random effects model was used as a sensitivity analysis.

The most appropriate analytical approach would have been to treat fractures as time-to-event data. However, this information was generally not available, so fracture incidence data (and RRs) were used instead. RRs and 95% confidence intervals (CIs), obtained from Cox regression analysis based on time-to-event data, were used in preference to the raw data.

For studies that reported only the number of patients with a fracture in each group (but did not report the denominator), or only the percentage of patients with a fracture in each group, and the loss to follow-up was not reported, we used the number of patients randomised as the denominator. Such studies were also used in sensitivity analyses.

When it was possible to combine studies, outcomes were summarised for dichotomous data using RRs. The numbers needed to treat (NNTs) (the number of patients who would have to be treated for one to have an improved outcome), 95% CI and the control group rate (range of rates) to which they apply, were calculated from the risk difference when appropriate.

When there were differences between studies in the way the results were reported, for example, summary statistics only or raw data, the summary statistic (for example, RR) and its standard error were calculated from the 95% CI. The studies were then combined using the generic inverse variance method in Review Manager.

Heterogeneity between trials was assessed by visual inspection of forest plots (poor overlap of horizontal lines was noted), and by using statistical measures: the  $\chi^2$  test for heterogeneity and the level of inconsistency,  $I^2$ :

$$I^2 = [(\chi^2 - df) / \chi^2] \times 100\%$$

where df represents the degrees of freedom. There was considered to be heterogeneity between trials if the heterogeneity p-value was less than 0.1 and/or  $I^2$  was greater than 50%. Any heterogeneity was explored further (see subgroup analyses below) and unexplained heterogeneous results were not used as the basis for recommendations.

### **Table 1 Guide for assessing the adequacy of reporting of sequence generation and allocation concealment**

---

#### **Adequate sequence generation**

---

Coin toss, throwing a dice, shuffling, drawing lots (from a container).

**Partial:** drawing a card from a pack.

---

Computer or calculator-generated sequence (including minimisation and biased-coin/urn design).

**Partial:** 'random permuted blocks'.

---

Random number table or statistical tables.

**Partial:** random numbers, randomisation table.

---

Randomised Latin square design.

---

#### **Inadequate sequence generation**

---

For example, allocation by alteration, birthdate, day of week.

---

#### **Adequate allocation concealment**

---

Central randomisation: with contacting details and/or statement that central office retained schedule; must apply to all patients.

**Partial:** vague statement of central randomisation.

---

Independent third party: allocates interventions **and** retains schedule; or statement that allocator has no knowledge of patients.

**Partial:** third party, but unclear treatment allocation.

---

Third party cluster randomisation: third party has no knowledge of clusters.

**Partial:** unclear what third party knew.

---

Different parties (including one of the authors): should have no knowledge of the patients **and** retain schedule.

---

Secure computer-assisted method, eg locked file.

**Partial:** as adequate, but unclear access.

---

Sequentially numbered, opaque, sealed envelopes – all required, otherwise **partial**.

---

Serially numbered, identical containers, allocated sequentially – all required, otherwise **partial**.

---

#### **Inadequate allocation concealment**

---

For example, schedule known in advance, birth date, case record number.

---

### **1.5.2 Stratifications**

Stratification was planned only according to the type of comparison or, sometimes, the type of intervention (see individual reviews).

Studies were combined across the type of osteoporosis, presence of prior fractures and population group (postmenopausal women, men, glucocorticoid-induced osteoporosis).

### **1.5.3 Subgroup analyses**

Subgroup analyses were carried out to investigate heterogeneity or to investigate prespecified features. Post-hoc within-trial subgroup analyses were considered to be data driven and were only included under exceptional circumstances.

Generally, the studies were subdivided according to the following criteria:

- population group (postmenopausal women, men, glucocorticoid-induced osteoporosis)
- type of comparison (osteoporosis intervention versus placebo/no treatment; osteoporosis intervention plus vitamin D/calcium supplementation versus vitamin D/calcium supplementation).

The following general prespecified factors were proposed for subgroup analyses:

- age (below 75, over 75 years)
- dose of intervention
- duration of intervention
- type of fracture (fragility, mixed trauma/fragility).

Subgroup analyses specific to each review were also carried out.

### **1.5.4 Sensitivity analyses**

Sensitivity analyses, which were carried out to investigate assumptions within the analyses, included the following:

- methodological quality
- fixed effects model
- other features specific to each review.

To appraise methodological quality, particular attention was paid to allocation concealment and missing data. Studies were not included in the analyses if they had more than 50% missing data. Sensitivity analyses were performed on studies that had between 20 and 50% withdrawals or protocol deviations in any group (that were eliminated from the study's analyses).

### **1.5.5 Interpretation**

Some meta-analyses gave pooled summary statistics close to the null value. If the CI was narrow, this was considered to be 'evidence for no significant difference' between interventions and the approach became similar to that of an equivalence trial (Alderson 2004). If the CI was wide, there was considered to be insufficient information to determine if there was a difference between interventions. Generally, for dichotomous outcomes the RR or odds ratio (OR) was recorded on a 0.1–10 scale and a judgement made about the width of CI.

### **1.6 Grading evidence**

The GRADE (Grading of Recommendations Assessment, Development and Evaluation) scheme (GRADE working group 2004) was used to assess the quality of the evidence for each outcome using the approach described below. Evidence summaries were produced across all outcomes.

According to the GRADE scheme, evidence is classified as high, moderate, low or very low:

- high – further research is very unlikely to change our confidence in the estimate of effect
- moderate – further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- low – further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- very low – any estimate of effect is very uncertain.

The following procedure was adopted when using the GRADE scheme.

1. A quality rating was assigned, based on the study design, for example, RCTs started as high and observational studies as low.
2. This rating was up or downgraded according to specified criteria: study quality, consistency, directness, preciseness and reporting bias. These criteria are detailed below. Criteria were given a downgrade mark of -1 or -2, depending on the severity of the limitations. Sometimes a downgrade mark of -1/2 was applied.
3. The downgrade/upgrade marks were then summed and the quality rating revised. For example, a decrease of -2 points for an RCT would result in a rating of 'low'.
4. Wherever possible, reasoning was explained for the downgrade marks.

### **1.6.1 Study quality**

Study quality was assessed against standard criteria, depending on the study design. For randomised trials, the following factors were taken into account: the adequacy of allocation concealment; blinding of participants for comparisons and outcomes susceptible to bias; attrition (missing data); and baseline comparability. A downgrade mark of -1 was given for inadequate allocation concealment and for a loss to follow-up of more than 20% in any one group or overall. A loss to follow-up of 50% or more was given a downgrade of -2. If the evidence was a meta-analysis of several studies, the proportion and weighting of poor quality studies was considered. In some instances, sensitivity analyses were performed disregarding these studies and a separate rating was given for the new meta-analysis.

### **1.6.2 Consistency**

When several RCTs had widely differing estimates of treatment effect (heterogeneity or variability in results) the results were regarded as inconsistent. This was defined as a p-value for heterogeneity less than 0.1 and/or an  $I^2$  value greater than 50%. In such cases, a downgrade mark of -1 was given. If the p-value was less than 0.1 and the  $I^2$  value was greater than 80%, a downgrade mark of -2 was given. Predefined subgroup analyses were

performed when possible to investigate heterogeneity and these results were reported separately. Generally, single trials (especially smaller ones) were not considered as having inconsistency unless there were a priori defined subgroups showing widely different effects.

### **1.6.3 Directness**

Directness refers to the extent to which the population, interventions, comparisons and outcome measures are similar to those defined in the inclusion criteria for the reviews. Indirectness is only relevant if there is a compelling reason to expect important differences in the size of the effect. For example, many interventions have more or less the same relative effects across patient groups, so extrapolation is possible and reasonable. There are various types of indirectness found in studies:

- when the guideline-defined drugs differ from those in the studies, but are within the same class
- when there are no direct comparisons of interventions, investigators must make comparisons across studies; for example, investigators want to know the difference in effectiveness between interventions A and B, but there is only information on A versus placebo and B versus placebo.

### **1.6.4 Preciseness**

This is a rather subjective, but nevertheless important category. Evidence is considered to be imprecise in the following cases.

- The sample size is small. This is a subjective measure and is more important in a single study. If there is a power calculation for that outcome and comparison, it is used to decide if a study is 'small'; otherwise, the rule of thumb was used that, if the study has less than 25 patients in any one arm, this is too small. The rationale for this is that, below 25 patients, assumptions about normal distributions become much less valid. However, if these small studies are combined in a meta-analysis, this would be more acceptable.
- There are sparse data (only a few events and they are uninformative).

- If CIs are sufficiently wide that the effect estimate is consistent with both important harms and important benefits, and would lead to conflicting recommendations. If CIs are very wide, a downgrade mark of –2 is given. Generally, the width of CIs for RRs was estimated by examining the width on a 0.1–10 scale.

### **1.6.5 Reporting bias**

Reporting bias occurs in two main ways.

- Publication bias – papers are more likely to be published if their results are statistically significant. The existence of publication bias in the studies in a meta-analysis can be investigated in a limited way using funnel plots, in which the standard error is plotted against the log odds ratio, the log RR or the mean difference. Asymmetry is indicative of reporting bias. This method is usually only useful when there are at least five studies. Industry-sponsored studies are also regarded as potentially biased. This bias is more severe if the outcomes are subjective, and fracture outcomes were not considered to be in this category. Therefore, a downgrade mark of –1/2 was given for industry-funded studies.
- Outcome bias – authors do not report some outcomes (probably because they have nonsignificant results), even though they say in the methods section that they have measured them.

The GRADE approach, although rigorous, still requires judgements to be made. For example, what is a ‘wide’ CI; what is a ‘small’ study; how important is blinding of patients for a particular outcome? These difficulties have been described in the bullet points above, and judgements were made as appropriate.

## **2 Pharmacological interventions to prevent osteoporotic fractures**

Sixteen reviews of individual pharmacological interventions are presented in this section, grouped by class of drug. In each case the selection criteria described in the methods section were used, although the interventions were

specific to each review. The permitted doses for each drug (as current in the BNF, March 2008 version) are given in appendix I.

Update searches were carried out for all reviews and search strategies are given in appendix II. Included studies are reported in appendix III, excluded studies in appendix IV and methodological quality in appendix V.

Summary findings are given in table 2 for all drugs evaluated. This is a summary of appendix VI, in which the quality of evidence, by outcome, is graded. The evidence is given in the form of RRs of various fractures, colour coded by quality.

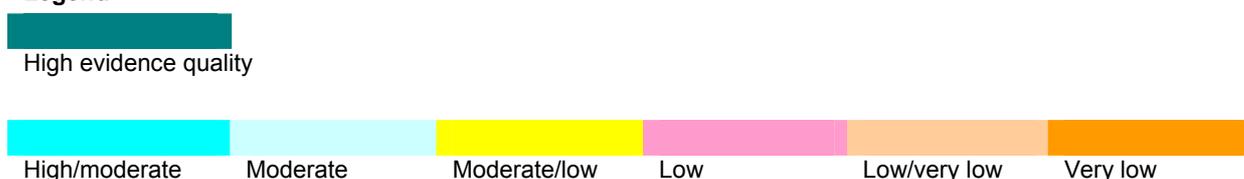
**Table 2. Summary of findings**

	Vertebral fracture	Nonvertebral fracture	Hip fracture	Arm fractures, wrist fractures, all fractures	
Alendronic acid vs placebo/no treatment	RR = 0.55 (95% CI 0.46 to 0.66) [high/moderate]	RR = 0.83 (95% CI 0.74 to 0.93) [high/moderate]	RR = 0.62 (95% CI 0.40 to 0.96) [high/moderate]	Wrist RR = 0.85 (95% CI 0.67 to 1.09) [low/very low]	
Etidronate vs placebo/no treatment	RR = 0.51 <sup>a</sup> (95%CI 0.31 to 0.83) [moderate]	RR = 0.72 (95%CI 0.29 to 1.80) [low]	RR = 1.02 (95%CI 0.21 to 4.94) [very low]	Wrist RR = 4.95 (95%CI 0.24 to 101.93) [very low]	All fractures RR = 0.78 (95%CI 0.42 to 1.44) [moderate]
Risedronate vs placebo	RR = 0.61 <sup>a</sup> (95%CI 0.50 to 0.74) [moderate/low]	RR = 0.81 (95% CI 0.72 to 0.90) [moderate/low]	RR = 0.73 (95% CI 0.58 to 0.92) [moderate/low]	Wrist RR = 0.68 (95% CI 0.43 to 1.07) [moderate/low]	Humerus RR = 0.46 (95% CI 0.23 to 0.93) [moderate/low]
Ibandronic acid vs placebo	RR = 0.51 (95%CI 0.34 to 0.74) [low/very low]	RR = 1.11 (95%CI 0.83 to 1.48) [low/very low]			
Zoledronic acid vs placebo	morphometric RR = 0.30 <sup>b</sup> (95% CI 0.24 to 0.38) [high/moderate]	RR = 0.75 (95% CI 0.66 to 0.85) [high/moderate]	RR = 0.62 (95% CI 0.47 to 0.83) [high/moderate]		

	Vertebral fracture	Nonvertebral fracture	Hip fracture	Arm fractures, wrist fractures, all fractures	
Strontium ranelate vs placebo	RR = 0.62 (95% CI 0.55 to 0.71) [moderate/low]	RR = 0.86 (95% CI 0.74 to 0.99) [moderate/low]	RR = 0.85 (95% CI 0.61 to 1.19) [moderate/low]	Wrist RR = 1.00 (95% CI 0.74 to 1.36) [moderate/low]	Humerus RR = 0.53 (95% CI 0.29 to 0.94) [moderate/low]
Teriparatide vs placebo	RR = 0.36 (95% CI 0.23 to 0.57) [moderate]	RR = 0.49 (95% CI 0.27 to 0.87) [moderate]	RR = 0.25 (95% CI 0.03 to 2.24) [very low]	Wrist RR = 0.29 (95% CI 0.06 to 1.38) [very low]	Humerus RR = 1.01 (95% CI 0.14 to 7.11) [very low]
PTH (1-84) vs placebo	RR = 0.39 (95% CI 0.22 to 0.69) [moderate/low]	RR = 0.97 (95% CI 0.71 to 1.33) [moderate/low]	Hip fracture leading to discontinuation RR = 0.65 (95% CI 0.11 to 3.86) [very low]		
Calcitonin vs placebo	RR = 0.41 <sup>b</sup> (95% CI 0.14 to 1.17) [low]	RR = 0.22 <sup>b</sup> (95% CI 0.02 to 1.96) [very low]	RR = 0.54 (95% CI 0.18 to 1.59) [very low]	Arm (all) RR = 0.79 (95% CI 0.38 to 1.61) [very low]	
HRT vs placebo	RR = 0.67 <sup>a</sup> (95% CI 0.48 to 0.93) [high/moderate]	RR = 0.73 (95% CI 0.65 to 0.81) [high]	RR = 0.63 (95% CI 0.42 to 0.93) [high]	All fractures RR = 0.70 (95% CI 0.63 to 0.78) [high]	
Raloxifene vs placebo	RR = 0.64 (95% CI 0.54 to 0.78) [moderate/low]	RR = 0.91 (95% CI 0.78 to 1.05) [low]	RR = 1.12 (95% CI 0.64 to 1.94) [low]	Wrist RR = 0.88 (95% CI 0.68 to 1.14) [very low]	
Nandrolone vs placebo	RR = 0.89 (95% CI 0.61 to 1.31) [low]	No evidence for this outcome			
Hydroxylated vitamin D vs placebo / no intervention	RR = 4.00 (95% CI 0.45 to 35.28) [very low]	RR = 0.46 (95% CI 0.18 to 1.18) [very low]			
Native vitamin D vs placebo/no intervention	RR = 0.66 (95% CI 0.40 to 1.08) [moderate]	RR = 1.02 <sup>b</sup> (95% CI 0.91 to 1.14) [high/moderate]	RR = 1.14 (95% CI 0.98 to 1.32) [moderate/low]		

	Vertebral fracture	Nonvertebral fracture	Hip fracture	Arm fractures, wrist fractures, all fractures	
vitamin D plus calcium vs placebo / no intervention	RR = 0.34 <sup>b</sup> (95% CI 0.01 to 8.34) [low/very low]	RR = 0.83 <sup>b</sup> (95% CI 0.73 to 0.94) [moderate]	RR = 0.79 <sup>b</sup> (95% CI 0.65 to 0.97) [moderate]	wrist and lower arm RR = 1.20 <sup>b</sup> (95% CI 0.73 to 1.98) [high/moderate]	all clinical fractures range of RRs 0.94 to 0.96 [moderate]
Calcium vs placebo	RR = 0.84 <sup>a</sup> (95% CI 0.66 to 1.08) [moderate]	RR = 0.92 <sup>a</sup> (95% CI 0.79 to 1.05) [high/moderate]	RR = 1.29 <sup>b</sup> (95% CI 0.89 to 1.88) [high/moderate]	Wrist RR = 1.05 (95% CI 0.57 to 1.92) [moderate]	All fractures 0.90 (95% CI 0.79 to 1.03) [moderate]
				Upper limb RR = 1.06 (95%CI 0.75 to 1.50) [high/moderate]	

#### Legend



<sup>a</sup> Sensitivity analysis available

<sup>b</sup> Sensitivity analysis

## 2.1 Bisphosphonates

### 2.1.1 Alendronic acid

#### Description of studies

A total of 61 papers were evaluated for inclusion. Sixteen of these papers reported only on BMD as an outcome and were excluded from the review (Adami 1995; Bone 2000; Chesnut 1995; Guo-ping 2005; Hiskins 1998; Ho 2005; Kasayama 2005; McClung 1998; McClung 2004; McClung 2005; Rossini 1994; Sambrook 2004; Shilbayeh 2004; Smith 2004; Uchida 2005; Wasnich 2004). Five studies were excluded because the comparator was

alfacalcidol, which is not licensed for osteoporosis in the UK (de Nijs 2006; Kushida 2002; Kushida 2004; Ringe 2001; Ringe 2004). A further 17 studies were excluded because the populations were higher-risk groups, such as liver transplant patients, and the GDG decided these would be unrepresentative.

Twenty-three reports of studies were included (Adachi 2001; Bone 1997; Bone 2004; Black 1996 [Fracture Intervention Trial – FIT]; Black 2000 [FIT]; Black 2003 [Parathyroid Hormone and Alendronate – PaTH]; Black 2005 [PaTH]; Black 2006 [Fracture Intervention Trial Long-term Extension – FLEX]; Carfora 1998; Cummings 1998 [FIT]; Dursun 2001; Greenspan 2002; Greenspan 2003; Hochberg 2005 [FIT]; Hosking 1998 [Early Postmenopausal Intervention Cohort study – EPIC]; Liberman 1995; Lindsay 1999; Miller 2004; Orwoll 2000; Pols 1999 [Fosamax International Trial – FOSIT]; Ravn 1999 [EPIC]; Saag 1998; Seeman 1999) that evaluated 15 RCTs.

One study (Liberman 1995) had three reports (Bone 2004; Liberman 1995; Seeman 1999); Seeman (1999) included the same 994 patients as Liberman (1995). Bone (2004) was an extension study of Liberman (1995) and reported results for years 6–10 in a sample of 247/994 (25%) patients who continued their randomised treatments.

Hosking (1998) and Ravn (1999) were two reports of the EPIC study, with Hosking (1998) forming an interim report.

The FIT was reported in four papers (Black 1996; 2000; Cummings 1998; Hochberg 2005). In this study, patients were stratified into those with a fracture and those without and then randomised to interventions; the former group was reported in Black (1996) (n = 2027) and the latter in Cummings (1998) (n = 4432). Black (2000) analysed a subset of women without vertebral fracture **and** a femoral neck T-score of –2.5 or less (n = 1631) and those with at least one vertebral fracture, resulting in a sample size of 3658 (see also Hochberg 2005).

The study by Adachi (2001) was a 1-year extension of a sample of patients (208/560 [37%]) from Saag (1998). The patients in Adachi (2001) were not representative of the original randomised group, in that the extension patients

had significantly fewer prior fractures. In addition, the patients in Adachi (2001) were receiving a higher median dose of prednisone (12 months after the start of the Saag [1998] study) compared with the nonextension participants. The study by Saag (1998), however, was only 48 weeks in duration and should have been excluded on the basis of too a short duration. The GDG decided that, in the absence of other studies in patients with glucocorticoid-induced osteoporosis, it was preferable to use the study by Saag (1998) as being more representative.

The FLEX trial (Black 2006) comprised patients who had been treated for 5 years with alendronic acid in the FIT study. These patients were then randomised to placebo or treatment groups (n = 1099 in total), and followed for a further 5 years. This study reported the number of patients with new fractures in the extension period only, rather than the number of patients over the whole 10 years, but is still a measure of the relative advantage of giving alendronate for 10 years rather than just for 5.

The PaTH trial (Black 2003; 2005) was designed to be of 2 years' duration. Patients were randomised to 1 year of the following interventions: (1) parathyroid hormone (1-84) (PTH 1-84) plus placebo; (2) PTH 1-84 plus alendronic acid (combination group); and (3) alendronic acid plus placebo. Then for the second year of the trial, patients in group (1) were randomised to alendronic acid (group 1a) or placebo (group 1b) and the patients in groups (2) and (3) were all given alendronic acid. A total of 94% of the patients completed the 2 years. The first year of this study was reported in Black (2003) and the second year in Black (2005). Comparisons relevant to the alendronic acid review are comparisons (2) versus (1) in Black (2003), and (1a) versus (1b) in Black (2005).

### ***Study design***

None of the studies were conducted exclusively in the UK. Of the 14 trials, five were multinational (Liberman 1995; Orwoll 2000; Pols 1999; Ravn 1999; Saag 1998), eight were conducted in the USA (Black 2005; FIT; FLEX; Bone 1997; Greenspan 2002; Greenspan 2003; Lindsay 1999; Miller 2004), one was in

Turkey (Dursun 2001), and the other in Italy (Carfora 1998). Of the multinational studies, at least one included UK centres (Ravn 1999).

Eleven studies received some funding from Merck, the manufacturers of alendronic acid (Black 2005; FIT; FLEX; Greenspan 2002; Liberman 1995; Lindsay 1999; Miller 2004; Orwoll 2000; Pols 1999; Ravn 1999; Saag 1998). One study did not explicitly state that Merck funded the study, but included coinvestigators who were employees of Merck (Bone 1997). One study was not funded by industry (Greenspan 2003). Two studies did not state a funding source (Carfora 1998; Dursun 2001).

Four studies had more than 100, but less than 200 patients (Carfora 1998, n = 136; Dursun 2001, n = 151; Miller 2004, n = 167; Orwoll 2000, n = 241). Six studies had more than 200, but less than 500 patients (Black 2005, n = 238; Bone 1997, n = 359; Greenspan 2002, n = 327; Greenspan 2003, n = 373; Lindsay 1999, n = 428; Saag 1998, n = 477). Five studies were large (Liberman 1995, n = 994; Pols 1999, n = 1908; Ravn 1999, n = 1609; FIT, n = 6459; FLEX, n = 1099).

One study took place in primary care (Greenspan 2003). Six studies took place in secondary care settings described as clinics (Black 2005; FIT; FLEX; Bone 1997; Carfora 1998; Dursun 2001) and one study (Greenspan 2002) took place in long-term care facilities. In one study, glucocorticoid therapy was managed by the patients' physicians, while osteoporosis treatment appears to have taken place in a clinic setting (Saag 1998). Six studies did not explicitly report the study setting (Liberman 1995; Lindsay 1999; Miller 2004; Orwoll, 2000; Pols 1999; Ravn 1999).

### ***Population***

Twelve studies aimed to evaluate the effectiveness of alendronic acid in postmenopausal women (Black 2005; FIT; FLEX; Bone 1997; Carfora 1998; Dursun 2001; Greenspan 2002; Greenspan 2003; Liberman 1995; Lindsay 1999; Pols 1999; Ravn 1999), although one of these (Lindsay 1999) included women who had a surgical menopause (29% had had bilateral oophorectomy). Two studies aimed to evaluate osteoporosis in men (Miller

2004; Orwoll 2000). One study consisted of patients receiving glucocorticoids for a variety of conditions, who were men and pre- and postmenopausal women (Saag 1998).

Overall, 13 studies included at least some, or all postmenopausal women (Black 2005; FIT; FLEX; Bone 1997; Carfora 1998; Dursun 2001; Greenspan 2002; Greenspan 2003; Liberman 1995; Lindsay 1999; Pols 1999; Ravn 1999; Saag 1998). The mean time since menopause, when reported, ranged from 8 to 25 years across the studies. One study (Saag 1998) included both men and women: 30% were men. Overall, three studies included men (Miller 2004; Orwoll 2000; Saag 1998).

Mean baseline femoral neck T-scores were reported (or calculated) for 11 studies and lumbar spine measures were reported in three (Bone 1997; Ravn 1999; Saag 1998); in the other study (Carfora 1998) the authors' classification of osteoporosis status was taken. On this basis, three studies were assessed as being in patients with osteoporosis only (T-score of  $-2.5$  SD or less; Black 1996 [FIT]; Carfora 1998; Liberman 1995), although Liberman (1995) did not give standard deviations for the T-scores; nine studies were assessed to include patients with osteoporosis or osteopenia (Black 2005; Black 2006 [FLEX]; Bone 1997; Cummings 1998 [FIT]; Greenspan 2002; Lindsay 1999; Miller 2004; Orwoll 2000; Pols 1999); and four studies were considered to cover all baseline BMDs (Dursun 2001; Greenspan 2003; Ravn 1999; Saag 1998).

Four studies did not report whether any of the patients had vertebral fractures at baseline (Carfora 1998; Dursun 2001; Pols 1999; Ravn 1999). Of the 11 remaining studies, the percentage of patients with at least one vertebral fracture at baseline varied from 6 to 70% across studies; two studies reported percentages of 30% or less (Liberman 1995; Saag 1998), five studies reported percentages greater than 30% and less than 50% (Black 2005; Bone 1997; FLEX; Greenspan 2003; Orwoll 2000) and four studies reported percentages greater than 50% (Black 1996 [FIT]; Greenspan 2002; Lindsay 1999; Miller 2004). Cummings (1998) analysed women with low BMD, but without vertebral fractures as part of the FIT study.

The age range across the studies was 21–91 years, with the mean age, when given, ranging from 43 to 73 years. The study that included the youngest patients was Saag (1998), in which the patients received glucocorticoids; in this study, the age ranged from 21 to 79 years. The study that included the oldest patients had an age range of 25–90 years (Miller 2004).

When reported, the mean height of patients in these studies ranged from 146 to 174 cm, the mean weight ranged from 49 to 79 kg, and the mean BMI ranged from 23 to 27 kg/m<sup>2</sup> across the studies; the lowest mean was in Lindsay (1999). Of the 13 studies that included women (see above), the mean height, when given, ranged from 145 to 169 cm, the mean weight ranged from 49 to 74 kg, and the mean BMI ranged from 23 to 29 kg/m<sup>2</sup>. Eight studies variously reported that 90–99% of the patients were white (Black 2005; Bone 1997; FIT; FLEX; Greenspan 2002; Lindsay 1999; Miller 2004; Orwoll 2000).

Of the 15 studies, only six reported whether any of the patients smoked (FIT; FLEX; Greenspan 2003; Lindsay 1999; Miller 2004; Orwoll 2000). Only two studies (Greenspan 2003; Lindsay 1999) reported whether any of the participants consumed alcohol.

### ***Interventions***

The following interventions were used:

- 5 mg/day alendronic acid (Bone 1997; FLEX; Ravn 1999)
- 10 mg/day alendronic acid (Black 2003; Carfora 1998; Dursun 2001; FLEX; Greenspan 2002; Greenspan 2003; Orwoll 2000; Pols 1999)
- 10 mg/day alendronic acid with hormone replacement therapy (HRT) (Lindsay 1999)
- 10 mg/day alendronic acid with PTH (1-84) (Black 2003)
- PTH (1-84) treatment for 1 year followed by 10 mg/day alendronic acid for 1 year (Black 2005)
- 70 mg/week alendronic acid (Miller 2004)
- combined doses:
  - 5 or 10 mg/day alendronic acid (Saag 1998)

- 5 mg/day for 2 years followed by 10 mg/day for 1–2 years alendronic acid (FIT)
- 5 or 10 mg/day, or 20 mg/day alendronic acid for 2 years followed by 5 mg/day alendronic acid for 1 year (Lieberman 1995).

Eight of the studies included more than two study arms (Black 2003; 2005; Bone 1997; FIT; Carfora 1998; Dursun 2001; Lieberman 1995; Ravn 1999). When possible, results for different licensed doses were combined. The following additional doses of alendronic acid were also examined in these studies, but have not been evaluated in this report:

- 1 mg/day alendronic acid (Bone 1997)
- 2.5 mg/day alendronic acid (Bone 1997; Ravn 1999)
- 20 mg/day alendronic acid (Carfora 1998, Lieberman 1995).

### **Comparisons**

The following comparisons were carried out.

- Alendronic acid versus placebo: 11 studies:
  - 1-year intervention period (Miller 2004; Pols 1999; Saag 1998: 48 weeks)
  - 2 years (Bone 1997; Greenspan 2002; Orwoll 2000)
  - 2.5 years (Carfora 1998)
  - 3 years (FIT; Greenspan 2003; Lieberman 1995)
  - 4 years (Ravn 1999).
- Alendronic acid versus no treatment for 1 year (Dursun 2001).
- Alendronic acid plus HRT versus placebo plus HRT for 1 year (Lindsay 1999).
- Alendronic acid plus PTH (1-84) versus placebo plus PTH (1-84) for 1 year (Black 2003).
- Alendronic acid versus placebo following alendronic acid in both arms:
  - 5 years' duration following 5 years of alendronic acid (FLEX).
- Alendronic acid versus placebo following PTH 1-84 in both arms:
  - 1-year duration following 1 year of PTH 1-84 (Black 2005).

In the majority of studies, all patients received either calcium supplements (ranging from 500 to 1500 mg/day across studies) (Bone 1997; Carfora 1998; Durson 2001; Liberman 1995; Pols 1999; Ravn 1999) or calcium and vitamin D supplements (ranging from 500 to 1000 mg/day calcium and 250–500 IU vitamin D across studies) (Saag 1998; Black 2005; FLEX; Miller 2004; Orwoll 2000). In one study, only those patients with inadequate intakes of calcium and vitamin D were given supplements (500 mg calcium and 250–400 IU vitamin D) (FIT). Three studies (Greenspan 2002; Greenspan 2003; Lindsay 1999) gave all patients vitamin D, but gave calcium supplements only to those with inadequate intakes.

Five studies reported concurrent medications; in one, patients were taking glucocorticoids (7.5 mg/day prednisone or its equivalent; Saag 1998). In Greenspan (2002) and Greenspan (2003), 64% and 33% respectively of the patients were taking aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs). All of the patients in Lindsay (1999) were taking HRT. Some (1.5–2.7%) of the patients in the FLEX study (Black 2006) were taking HRT or raloxifene, and in the study by Orwoll (2000), 4% of the men received doses of testosterone.

A washout period was not reported in the FLEX study (Black 2006), thus it is likely that the patients would still have carryover effects from the alendronic acid treatment received during the FIT. Two other studies reported a run-in period: 2 weeks on placebo in Lindsay (1999) and 2–4 weeks in Pols (1999) and Greenspan (2003) reported a 3 month run-in on HRT, alendronic acid placebo, calcium and multivitamins, as necessary.

### ***Methodological quality***

The method of sequence generation was adequate in four studies in which a computer-generated sequence was employed (Cummings 1998 [FIT]; Greenspan 2002; Greenspan 2003; Miller 2004), and was partially adequate in the FLEX study (the authors used a permuted block design). None of the other studies clearly reported a method of sequence generation.

Only two studies reported an adequate method of allocation concealment: the FIT study involved centralised randomisation by an independent third party

and Greenspan (2003) involved the research pharmacy which retained the code. Another study reported that the allocation schedule was blinded (Miller 2004). Allocation concealment was unclear in all other studies.

Twelve studies were described as double blind (Black 2005; Bone 1997; FIT; FLEX; Greenspan 2002; Greenspan 2003; Liberman 1995; Miller 2004; Orwoll 2000; Pols 1999; Ravn 1999; Saag 1998). Blinding was unclear in three studies (Carfora 1998; Durson 2001; Lindsay 1999).

Four studies reported an a priori sample size calculation, but only one was powered to detect fracture: Cummings (1998) (one arm of the FIT: women without vertebral fracture) required a sample size of 4000, assuming a 4% annual incidence of clinical fracture in the placebo group, to detect a 25% decrease in risk with 90% power at a 5% significance level. Bone (1997) estimated that a sample size of 60 per group was expected to provide 80% power to detect a 1.8% between-group difference in the mean percent change from baseline in lumbar BMD. Miller (2004) estimated that 80 participants in the treatment group and 40 in the placebo group were required to detect a 3% change in spine BMD, with a power of 97% at a 5% significance level. Greenspan (2003) stated that 92 participants per group were needed for 80% power to detect a 3% (or 90% power to detect a 4%) difference in BMD between each group and placebo. All of these studies had sample sizes greater than those estimated.

Power calculations were also reported for four other studies, only two of these were based on fracture risk: Black (1996) (the other arm of the FIT: women with existing vertebral fracture) reported 99% power to detect a 40% reduction in risk of vertebral fractures and a power of 90% to detect a 32% reduction, assuming a 6.5% annual incidence. The FLEX study (Black 2006) estimated that the study had 80% power to detect a risk reduction of 33% to 13.5%, based on 20% incidence of fracture. The other power calculations were based on BMD: with 90% power and a 5% significance level, Black (2005) expected to detect a difference between any two treatment groups in BMD of 4% at the spine and 2.4% at total hip. Pols (1999) estimated that their study had 99%

power to detect a 3.5% difference between groups in mean percentage change from baseline in BMD of lumbar spine, assuming a SD of 4.5%.

Almost all of the studies included in the review demonstrated baseline comparability of the groups, with the exception of Lindsay (1999), which reported a significantly higher proportion of women with a family history of fracture for the alendronate group, and a significantly higher proportion of smokers in that group. Carfora (1998) did not present baseline characteristics of the sample.

Only one study had withdrawals of more than 20% (Ravn 1999). Eleven studies reported withdrawals of 20% or less (Black 2005; Bone 1997; FIT; FLEX; Greenspan 2003; Liberman 1995; Lindsay 1999; Miller 2004; Orwoll 2000; Pols 1999; Saag 1998), and three studies did not clearly report the numbers of missing participants (Carfora 1998; Dursun 2001; Greenspan 2002). In the study by Ravn (1999), 47% (107/226) of patients who received alendronic acid and 27% (134/502) who received placebo withdrew from the study, however the authors analysed all patients randomised (ITT).

Six studies analysed data from available patients (Black 2005, Dursun 2001; Liberman 1995; Miller 2004; Pols 1999; Saag 1998). Seven studies included all the patients in the analysis (Bone 1997; Carfora 1998; FIT; FLEX; Greenspan 2003; Lindsay 1999; Ravn 1999). The number of patients included in the final analyses was not clearly reported in Orwoll (2000) and Greenspan (2002).

Methods to assess concordance were only reported in four studies (Black 1996 [FIT]; Black 2005; FLEX; Bone 1997). It was assessed by interview, self-report, or pill count.

Most of the studies described how vertebral fracture was assessed, usually as a radiographic 20–25% reduction in vertebral height. Orwoll (2000) reported using semi-quantitative and quantitative morphometric techniques, but did not describe these in detail. Two studies did not state how they assessed vertebral fracture (Carfora 1998; Ravn 1999), and four studies did not

specifically analyse vertebral fracture (Black 2005; Greenspan 2002; Greenspan 2003; Pols 1999).

Four studies evaluated vertebral fracture as a primary outcome (FIT; Carfora 1998; Dursun 2001; Liberman 1995), three assessed vertebral fracture as a secondary outcome (Bone 1997; Orwoll 2000; Saag 1998) and three studies (Black 2005; Greenspan 2002; Greenspan 2003) only evaluated 'clinical fractures'. One study considered fracture incidence to be an exploratory outcome (FLEX), and five studies considered fractures to be adverse events (Greenspan 2003; Lindsay 1999; Miller 2004; Pols 1999; Ravn 1999).

Of the included studies, two stated that pathological fractures, traumatic accident fractures or skull fractures were excluded (Black 1996 [FIT]; FLEX). One study (Miller 2004) reported trauma and fragility fractures separately. One 4-year study (Ravn 1998) reported in its 2-year interim report (Hosking 1998) that all fractures were traumatic and Ravn (1998) reported that none of the fractures were drug related; therefore the results were not included in the analyses. Two studies included traumatic fractures in their analyses of nonvertebral fracture (Bone 1997; Liberman 1995); these latter results were considered in sensitivity analyses; the rest of the studies did not state whether or not osteoporotic versus traumatic fractures were included.

Overall, the study by Ravn (1999) was considered to be at higher risk of bias due to the percentage of withdrawals and the differential missing data. The study by Saag (1998) was also considered at risk of bias because the study duration was too short, and Lindsay (1999) had potential for bias because of differences in baseline risk. These higher-risk studies were treated with caution and examined in sensitivity analyses. The studies in which fractures were recorded as adverse events (Greenspan 2003; Lindsay 1999; Miller 2004; Pols 1999) or in which traumatic nonvertebral fractures were included with fragility fractures (Bone 1997; Liberman 1995) were also considered in sensitivity analyses and the data were not analysed for the study by Ravn (1998).

## **Results**

### **Alendronic acid versus placebo/no treatment**

#### *Vertebral fractures*

Nine studies evaluated vertebral fracture (Black 1996 [FIT]; Bone 1997; Carfora 1998; Cummings 1998 [FIT]; Dursun 2001; Liberman 1995; Miller 2004; Orwoll 2000; Saag 1998); the two arms of the FIT study were treated as separate studies. Three studies reported only percentages: in Carfora (1998) the number randomised was assumed to be the denominator; also, the 5 and 10 mg groups were combined. In Bone (1997) the group sizes for lumbar spine BMD measurements were assumed, and in Orwoll (2000) the denominator was calculated from numbers and percentages given for painful vertebral fractures. Meta-analysis of nine studies in 8074 patients was carried out in subgroups of women, men and patients with glucocorticoid-induced osteoporosis (figure 1).

Overall, there were significantly fewer patients with vertebral fractures in the alendronate group compared with placebo/no treatment; RR 0.55 (95% CI 0.46 to 0.66), which corresponds to a NNT of 33 (95% CI 25 to 50), for a control group rate range of 4–35%. There was no statistical heterogeneity between the studies ( $I^2 = 0\%$ ;  $p = 0.59$ ). A sensitivity analysis, omitting studies at risk of bias (for example, Saag 1998) or in which there were uncertainties (Bone 1997; Carfora 1998; Orwoll 2000) showed very similar results. A funnel plot was also produced to examine if there was publication bias, and it was concluded that there was not (figure 2).

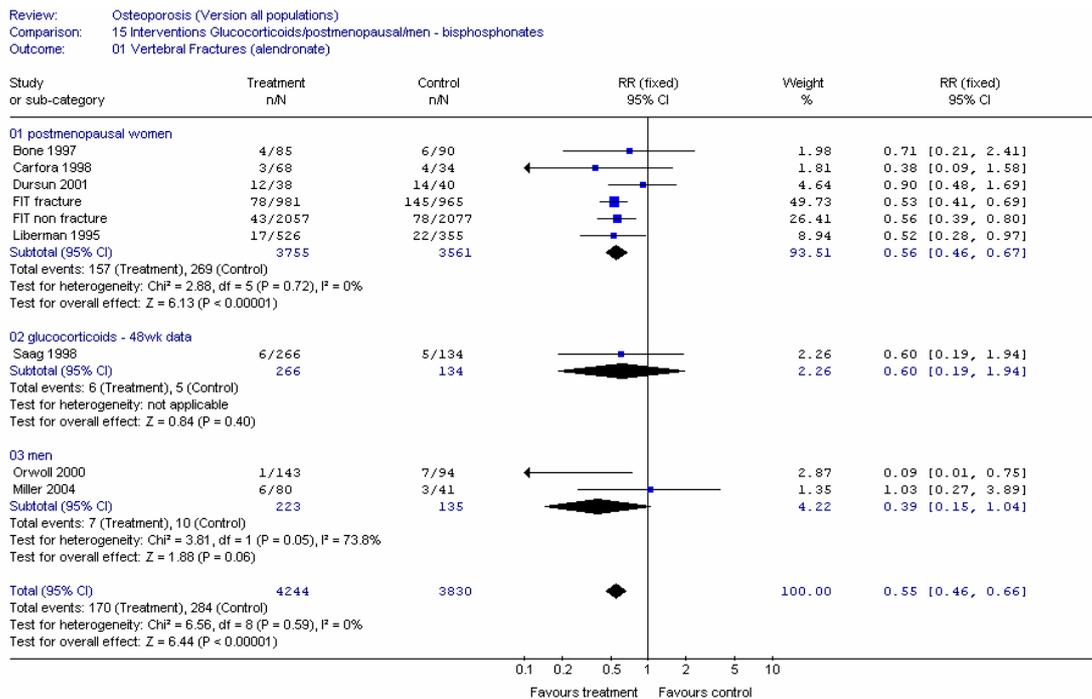


Figure 1. Number of patients with vertebral fractures

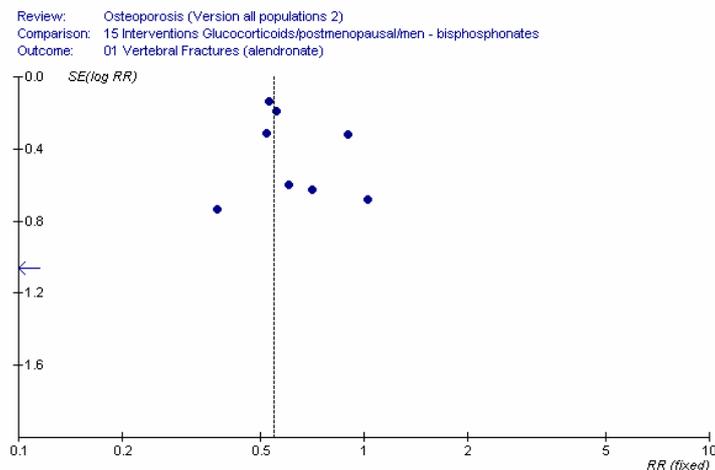


Figure 2. Number of patients with vertebral fractures – funnel plot

The population subgroups were then considered separately.

Six studies evaluated vertebral fracture in postmenopausal women (Black 1996 [FIT]; Bone 1997; Cummings 1998 [FIT]; Carfora 1998; Dursun 2001; Liberman 1995). There were significantly fewer patients with vertebral fractures for alendronate compared with placebo/no treatment, and no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.72$ )

Two studies evaluated men, and there was significant heterogeneity between studies ( $I^2 = 74\%$ ;  $p = 0.05$ ). One of these studies (Orwoll 2000) assessed vertebral fractures using both semi-quantitative methods and quantitative methods; they reported no significant difference in the former case ( $p = 0.12$ ) and a significant difference in the latter ( $p = 0.02$ ). The study by Miller (2004) reported fractures as adverse events.

Only one study evaluated vertebral fracture in patients with glucocorticoid-induced osteoporosis (Saag 1998). The trialists evaluated men and women treated with either 5 or 10 mg/day alendronic acid compared with placebo for 48 weeks. The CI was too wide to determine if there was a difference between interventions.

#### *Nonvertebral fractures*

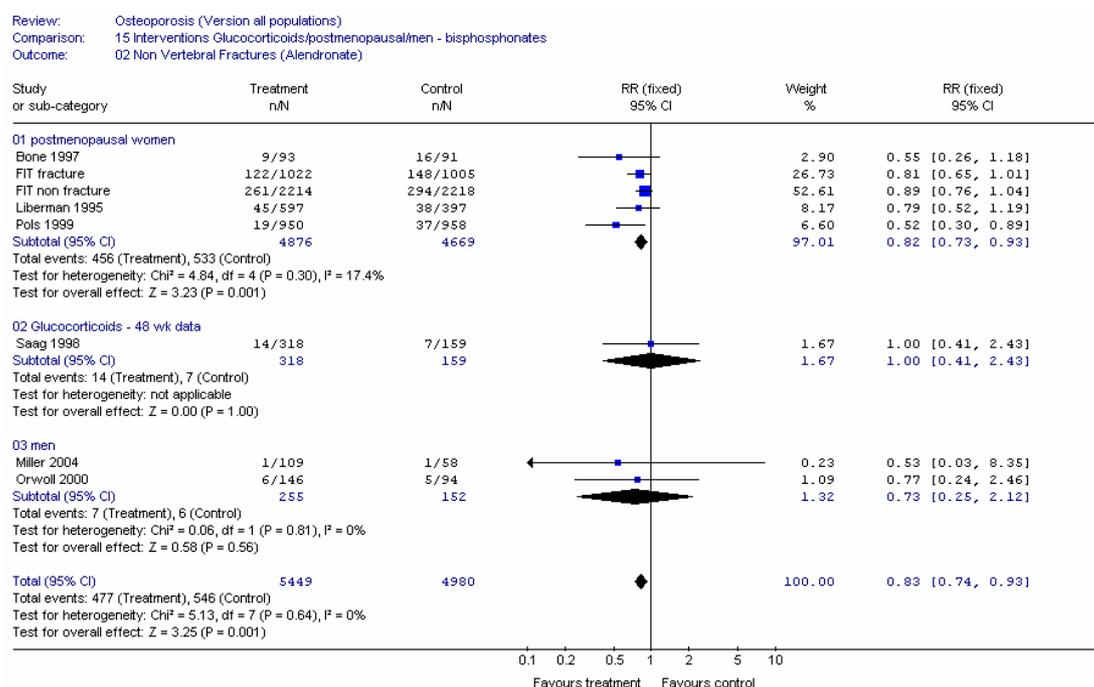
Nine studies recorded the incidence of nonvertebral fractures (Black 1996 [FIT]; Bone 1997; Cummings 1998 [FIT]; Greenspan 2002; Liberman 1995; Miller 2004; Orwoll 2000; Pols 1999; Saag 1998). Bone (1997) and Liberman (1995) included traumatic fractures in their dataset, and these studies were considered in sensitivity analyses.

Two studies (Miller 2004; Pols 1999) did not report the denominator for nonvertebral fractures (as adverse events). In each case, the number of patients randomised was assumed. One study (Saag 1998) reported only percentages; the number randomised was used as the denominator. One study (Greenspan 2002) did not give the number randomised in each group and this study was not included in the analyses.

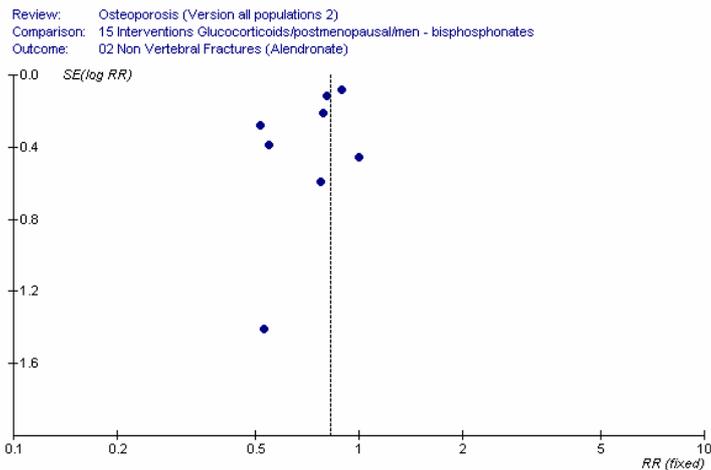
A meta-analysis was conducted on the eight studies in 10,429 patients with separate subgroups for women, men, and patients with glucocorticoid-induced osteoporosis (figure 3). Overall, there were significantly fewer patients with nonvertebral fractures in the alendronate group compared with placebo/no treatment; RR 0.83 (95% CI 0.74 to 0.93), which corresponds to a NNT of 50 (95% CI 33 to 100), for a control group rate range of 2–18%. There was no statistical heterogeneity between the studies ( $I^2 = 0\%$ ;  $p = 0.64$ ). A sensitivity analysis, omitting studies at risk of bias (for example, Bone 1997; Liberman

1995; Miller 2004; Saag 1998) or in which there were uncertainties (Orwoll 2000; Pols 1999), showed similar results. The funnel plot showed some asymmetry (figure 4).

On considering the population subgroups separately, five studies included postmenopausal women (Black 1996 [FIT]; Bone 1997; Cummings 1998 [FIT]; Liberman 1995; Pols 1999). There were significantly fewer patients with nonvertebral fractures for alendronate compared with placebo/no treatment, and no significant heterogeneity between studies ( $I^2 = 17\%$ ;  $p = 0.30$ ).



**Figure 3. Number of patients with nonvertebral fractures**



**Figure 4. Number of patients with nonvertebral fractures – funnel plot**

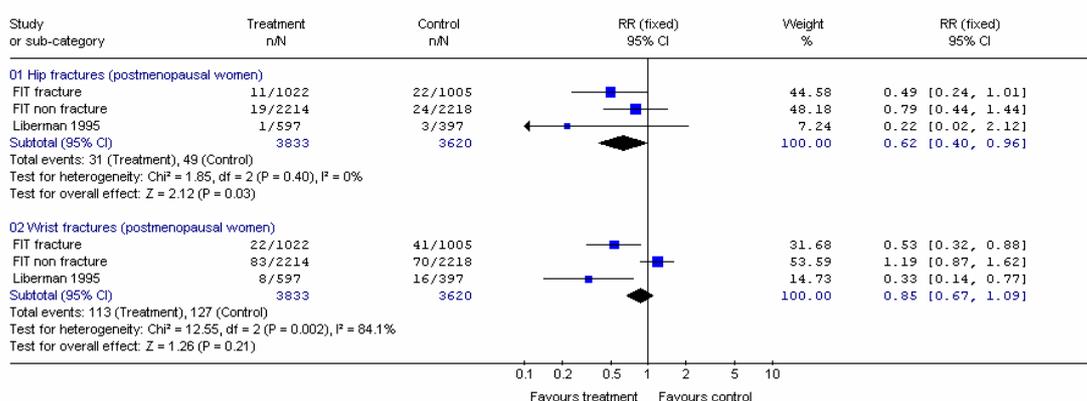
Two studies evaluated nonvertebral fractures in men (Orwoll 2000; Miller 2004). The CI was too wide to determine if there was a difference between interventions.

One study evaluated nonvertebral fracture in patients with glucocorticoid-induced osteoporosis (Saag 1998). There was no significant difference between interventions, although the CI was fairly wide.

#### *Hip fractures*

Three studies in 7453 patients evaluated hip fracture in postmenopausal women (Black 1996 [FIT]; Cummings 1998 [FIT]; Liberman 1995). Overall, there were significantly fewer hip fractures in the patients given alendronic acid: RR 0.62 (95% CI 0.40 to 0.96) (figure 5). This corresponds to an NNT of 100 (95% CI 100 to  $\infty$ ). There was no statistical heterogeneity between the studies ( $I^2 = 0\%$ ;  $p = 0.40$ ). An additional study (Greenspan 2002) recorded clinical hip fractures, but did not give the number of patients randomised to groups; there were twice as many fractures in the placebo group.

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausal/men - bisphosphonates  
 Outcome: 15 Hip and Wrist Fractures - alendronate, all populations



**Figure 5. Number of patients with hip fracture and with wrist fracture**

### Wrist fractures

Both arms of the large FIT study evaluated wrist fracture (Black 1996; Cummings 1998) and Liberman (1995) reported on wrist and forearm fractures (figure 5). There was significant heterogeneity between studies ( $I^2 = 84\%$ ;  $p = 0.002$ ); RR 0.85 (95% CI 0.67 to 1.09). Sensitivity analysis without the study by Liberman (1995) (which may have included traumatic fractures) did slightly affect the summary statistics, but did not remove the heterogeneity.

Heterogeneity was examined by considering the a priori specified subgroup analyses.

- Age (above and below 75 years): mean age was similar across the trials.
- Fractures at baseline: some patients in Liberman (1995) had fractures at baseline; all patients in Black (1996) and no patients in Cummings (1998) had baseline fractures – this may explain some of the heterogeneity.
- Black (1996) and Liberman (1995) were in people with confirmed osteoporosis, although Liberman (1995) did not give SDs for the T-scores, whereas Cummings (1998) included people with either osteoporosis or osteopenia; however, the mean T-score was similar for Liberman (1995) and Cummings (1998); this may explain some of the heterogeneity.
- Dose of alendronic acid: there was no difference in dose between the two FIT trial arms.

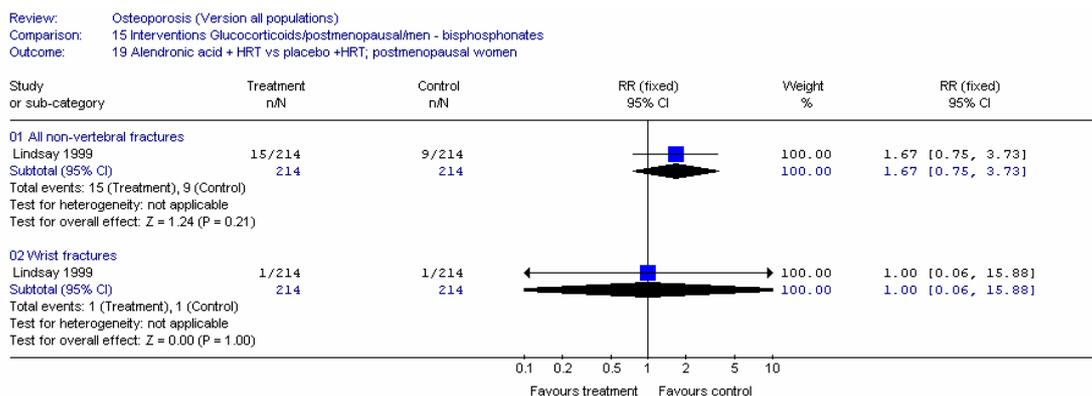
- Duration of intervention: there was no difference between the two FIT trial arms.

### Alendronic acid plus HRT versus placebo plus HRT

One study (Lindsay 1999) in 428 women examined the added effect of alendronic acid versus placebo in patients who all received HRT at doses sufficient to give an effect on osteoporotic fractures. The study only recorded fractures as adverse events, broken down by fracture site. No patient suffered a hip fracture or a vertebral fracture.

#### Nonvertebral fractures

There was no significant difference between interventions in the number of women with a nonvertebral fracture, although the CI was fairly wide (figure 6); RR 1.67 (95% CI 0.75 to 3.73).



**Figure 6. Alendronic acid plus HRT versus HRT; all nonvertebral fractures and all wrist fractures**

#### Wrist fractures

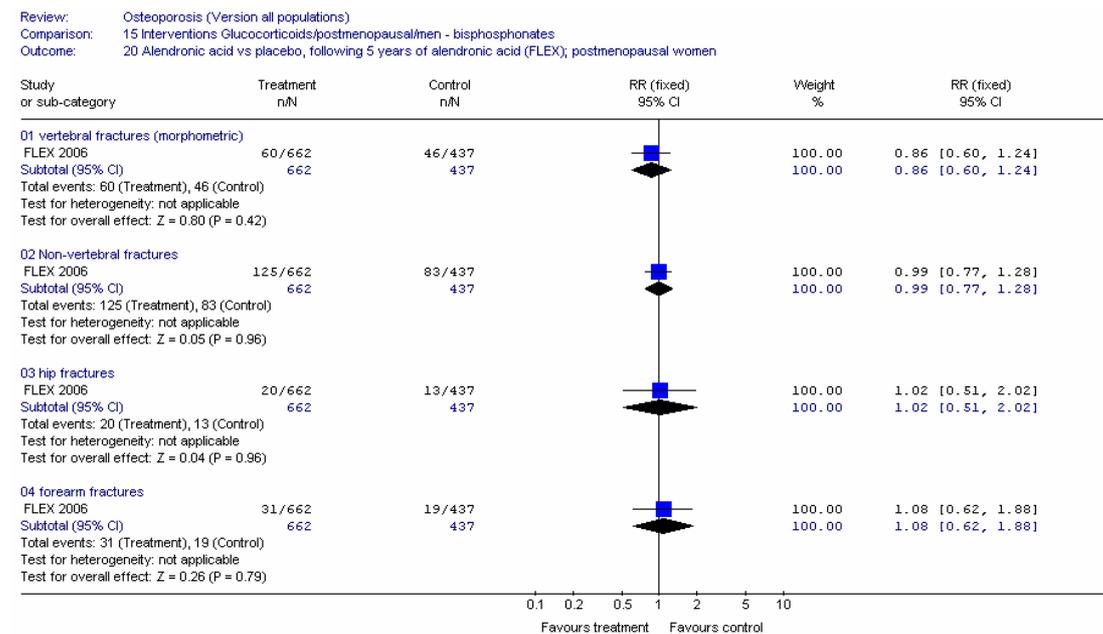
The CI was too wide to determine if there was a difference between interventions (figure 6).

### Alendronic acid plus PTH (1-84) versus placebo plus PTH (1-84)

One comparison of Black (2003) compared alendronic acid plus PTH (1-84) versus placebo plus PTH (1-84) in 119 patients. The study reported that the incidence of clinical fractures was similar in the two groups (approximately 3%), but no further data were given.

## Alendronic acid versus placebo following 5 years of alendronic acid: the FLEX study

Black (2006) randomised 1099 women who had been treated with alendronic acid in the FIT study for at least 3 years (mean 5 years) to either 5 or 10 mg/day alendronic acid or placebo for 5 years. This study therefore investigated how useful it would be to extend the period on alendronic acid to 10 years rather than stopping alendronic acid after 5 years. The authors of this study recorded the number of women with new fractures during the extension period only; the two alendronate groups were pooled. The authors recorded the number of women with a fracture in the period, but also gave RRs adjusted for clinic and stratum (with and without one or more morphometric vertebral deformities at the end of the FIT study or with a clinical fracture during FIT). There was little effect of adjustment and the raw data are reported here (figure 7).



**Figure 7. Number of women with fractures in the FLEX study**

### Vertebral fractures

There was no significant difference between placebo and the pooled alendronic acid treatment groups in the number of women with new morphometric vertebral fractures; RR 0.86 (95% CI 0.60 to 1.24) (figure 7).

### *Nonvertebral fractures, hip fractures and forearm fractures*

There was no significant difference between groups in the number of women with all nonspine fractures; RR 0.99 (95% CI 0.77 to 1.28) (figure 7).

### *Hip fractures*

There was no significant difference in the number of women with hip fractures; RR 1.02 (95% CI 0.51 to 2.02); (figure 7).

### *Forearm fractures*

There was no significant difference in the number of women with forearm fractures; RR 1.08 (95% CI 0.62 to 1.88) (figure 7).

### **Alendronic acid versus placebo following PTH (1-84)**

One comparison of Black (2005) examined the effectiveness of giving alendronic acid for 1 year following 1 year of PTH (1-84) in 119 women. Fracture incidence was determined over the full 2-year period. The study reported no significant difference in the proportion of women with a clinical fracture, but no data were given.

## **2.1.2 Etidronate**

### ***Description of studies***

A total of 33 papers were evaluated for inclusion. Fourteen studies (Fujita 2001; Ishida 2004; Iwamoto 2001; Jinnouchi 2000; Lyritis 1997; Nakayamada 2004; Pacifici 1988; Pitt 1998; Sato 2003; Shiomi 2002; Skingle 1997; Storm 1990; Wimalawansa 1998; Worth 1994) were excluded because the etidronate regimen was not listed in the BNF. Two studies (Hasegawa 2003; Yamaguchi 2003) reported only on BMD as an outcome and were excluded from the review, five other studies were excluded because they were in higher risk groups and one for other reasons given in Appendix IV.

Eleven studies were included in the review (Adachi 1997; Campbell 2004; Cortet 1999; Geusens 1998; Herd 1997; Jenkins 1999; Meunier 1997; Montessori 1997; Pouilles 1997; Roux 1998; Watts 1990).

### ***Study design***

Three studies were conducted in the UK (Campbell 2004; Herd 1997; Jenkins 1999); three in France (Cortet 1999; Meunier 1997; Pouilles 1997); and one each in the USA (Watts 1990); Canada (Adachi 1997); The Netherlands (Montessori 1997) and Belgium (Geusens 1998). One study (Roux 1998) was a multicentre trial held in the following countries: UK, Belgium, France, Germany, Italy and The Netherlands.

Four studies reported receiving a grant or support from Procter and Gamble Pharmaceuticals or one of its subsidiary companies (Adachi 1997; Meunier 1997; Montessori 1997; Geusens 1998). The pharmaceutical company provided the drugs for the Watts (1990) study. One study (Jenkins 1999) received funding from a public body and the rest did not say.

Two studies (Campbell 2004; Watts 1990) had more than two arms: Campbell (2004) also randomised patients to groups, to which etidronate alone and calcium alone were given, but these other interventions were disregarded because they were regimens not in the BNF. Watts (1990) had two other arms which were also given phosphate; these were not included either.

The largest study included 429 participants (Watts 1990) and the number of participants in the remaining trials were as follows: six studies had fewer than 100 patients (Cortet 1999, n = 83; Geusens 1998, n = 37; Jenkins 1999, n = 28; Meunier 1997, n = 54; Montessori 1997, n=80; Pouilles 1997, n=99); four had between 100 and 200 patients (Adachi 1997, n = 141; Campbell 2004, n = 349; Herd 1997, n =152; Roux 1998, n = 117). Two studies had 25 or fewer patients in the intervention arm (Geusens 1998, n = 18; Jenkins 1999, n = 15).

The majority of the studies were conducted in primary care, with two being carried out in secondary care (Cortet 1999; Watts 1990) and one not stating the setting (Geusens 1998).

## **Population**

Six studies included postmenopausal women only (Geusens 1998; Herd 1997; Meunier 1997; Montessori 1997; Pouilles 1997; Watts 1990). The mean time since the menopause ranged from 2 to 17.5 years. Five studies included both men and women (Adachi 1997; Campbell 2004; Cortet 1999; Jenkins 1999; Roux 1998). Of these studies, three included pre-menopausal women (Adachi 1997 12%; Cortet 1999 11%; Roux 1998 15%).

Six studies included people using glucocorticoids (Adachi 1997; Campbell 2004; Cortet 1999; Geusens 1998; Jenkins 1999; Roux 1998). Participants in these studies had comorbidities (rheumatological, dermatological, chronic active hepatitis, and respiratory illnesses).

Mean baseline femoral neck T-scores were reported (or calculated) for seven studies (Geusens 1998; Herd 1997; Jenkins 1999; Meunier 1997; Montessori 1997; Roux 1998; Watts 1990), one reported BMD at the lumbar spine (Adachi 1997), one at the proximal femur (Campbell 2007), one did not give enough information to specify the site (Cortet 1999) and BMD was not reported in the remaining study (Pouilles 1997). On this basis, no studies were assessed as being in patients with osteoporosis only; four were in patients with osteoporosis or osteopenia (Herd 1997; Meunier 1997; Montessori 1997; Watts 1990); two were in patients with osteopenia or normal BMD (Roux 1998; Jenkins 1999); four were in people with all categories of BMD (Adachi 1997; Campbell 2004; Cortet 1999; Geusens 1998) and one did not report BMD (Pouilles 1997).

One study (Watts 1990) reported that all patients had vertebral fractures at baseline. Herd (1997) reported no patients with fractures. Four studies reported that the patients had some vertebral fractures at baseline (Adachi 1997; Campbell 2004; Montessori 1997; Roux 1998). Two studies reported percentages of 30% or less (Campbell 2004; Roux 1998), and two reported percentages greater than 30% and less than 50% (Adachi 1997; Montessori 1997).

The age range across the studies, where given, was 19 to 87 years. The mean age of participants across the studies ranged from 53 to 69 years.

One study (Meunier 1997) reported smoking history in the patients (11% in each group); alcohol consumption was not reported in any of the studies.

All participants were Caucasian in four studies (Herd 1997; Meunier 1997; Montessori 1997; Pouilles 1997); patients were white or Asian in Watts (1990) and ethnicity was not reported in the remaining studies.

### ***Interventions***

The following interventions were given, all of which approximated closely to the BNF regimen of Etidronate 400mg/day for 14 days followed by calcium carbonate (with 500 mg/day elemental calcium) for 76 days:

- Etidronate (400 mg/day 14 days) followed by calcium (500 mg/day for 76 days)
  - calcium as carbonate (Adachi 1997 Roux 1998)
  - calcium as citrate (Montessori 1997)
  - calcium type unspecified (Cortet 1999; Herd 1997)
- Etidronate (400 mg/day 2 weeks) followed by calcium (500 mg/day for 11 weeks)
  - calcium as citrate (Pouilles 1997)
  - calcium type unspecified (Geusens 1998; Jenkins 1997; Meunier 1997)
- Etidronate (400 mg/day for 2 weeks every 3 months) followed by calcium carbonate (500 mg daily) (Campbell 2004)
- Placebo phosphate for 3 days then etidronate (400 mg/day 14 days) followed by calcium carbonate (500 mg/day for 74 days) (Watts 1990)

### ***Comparisons***

The following comparators were used:

- Placebo etidronate
  - placebo etidronate followed by calcium (500 mg/day for 76 days) (Adachi 1997; Cortet 1999; Herd 1997; Roux 1998)

- placebo etidronate followed by calcium (500 mg/day for 11 weeks) (Geusens 1998; Jenkins 1999; Meunier 1997)
- placebo phosphate for 3 days then placebo etidronate for 2 weeks then 500mg/day elemental calcium for 74 days (Watts 1990)
- No treatment (Campbell 2004)
- Calcium
  - calcium citrate 500mg day elemental calcium continuously (Campbell 2004; Montessori 1997)

The following comparisons were carried out:

- Etidronate regimen versus placebo etidronate (Adachi 1997; Cortet 1999; Geusens 1998; Herd 1997; Jenkins 1999; Meunier 1997; Pouilles 1997; Roux 1998; Watts 1990)
  - 1 year intervention period (Adachi 1997; Cortet 1999; Jenkins 1999; Roux 1998)
  - 2 years (Geusens 1998; Herd 1997; Meunier 1997; Pouilles 1997; Watts 1990)
- Etidronate regimen versus no treatment for 5 years (Campbell 2004)
- Etidronate regimen versus calcium
  - intervention period 3 years (Montessori 1997)
  - 5 years (Campbell 2004)

The majority of studies did not appear to give additional supplementation with either calcium or vitamin D. Two studies gave additional vitamin D: in Roux (1998) patients were allowed Vitamin D up to 1000UI/day and supplements were also given in Cortet (1999).

Compliance was assessed by pill counts (frequency not stated) in four studies (Herd 1997; Meunier 1997; Pouilles 1997; Roux 1998), patients were interviewed every 3 months on missed trial medication and if data were not available or considered incomplete, the patient's general practitioner was asked for additional information (Montessori 1997) and medication recorded in a daily diary (Jenkins 1999).

### ***Methodological quality***

The method of sequence generation was adequate (computer generated) in two studies (Montessori 1997; Watts 1990); partially adequate (permuted blocks) in one study (Campbell 2004) and unclear in the remaining studies.

Three studies (Adachi 1997; Campbell 2004; Herd 1997) reported stratification before randomisation. Patients were stratified by gender and menopausal status (Adachi 1997), years since menopause (Herd 1997) and according to the use of glucocorticoids (variation between method of intake and duration of medication) (Campbell 2004)

Allocation concealment was unclear in all studies.

The patients were blinded in the majority of studies. Exceptions were Montessori (1987), in which they were not blinded, and blinding was unclear in Campbell (2004), although it is likely, because of the different interventions, that the patients were unblinded. Seven studies (Adachi 1997; Campbell 2004; Geusens 1998; Meunier 1997; Montessori 1987; Pouilles 1997; Watts 1990) reported blinding of outcome assessors to the interventions and this was unclear in the remaining studies.

One study (Campbell 2004) carried out a power calculation with new clinical fractures (vertebral and non vertebral) as the primary outcome. Based on fracture rates of 8% versus 3%, in the etidronate and placebo groups respectively, it was calculated that 750 patients would detect a statistically significant difference at the 5% level, 80% power between the intervention and control group; the study only recruited 349 patients and so was underpowered. One study (Roux 1998) carried out a power calculation based on percent change in lumbar spine BMD. Based on a decrease in lumbar spine BMD of 2% versus 3%, for the etidronate and placebo groups, respectively, it was calculated that 42 patients were required per treatment group at 80% power.

Baseline comparability, assessed on age, gender, height and BMD was demonstrated in all the studies with exception of three (Adachi 1997; Jenkins 1999; Pouilles 1997). Adachi (1997) only reported comparability for age and

gender. In Jenkins (1999), there were significantly greater height and lumbar spine BMD values in the placebo group which also had a higher proportion of men. Pouilles (1997) reported a 'slightly lower' mean lumbar spine BMD in the etidronate group.

Missing data more than 20% were reported in three studies (Campbell 2004; Geusens 1998; Jenkins 1999). In Jenkins (1999), radiographs were not available for 60% of the etidronate group and 46% in the placebo group, i.e. a differential loss to follow up. Campbell (2004) reported deaths and loss to follow up of 33% in the etidronate group and 28% in the no treatment group. Geusens (1998) reported that 28% and 32% withdrew from the etidronate and placebo groups respectively. The remaining studies had missing data of less than 20%; in three studies there was differential missing data between groups: Roux (1998) reported withdrawals of 12% in the etidronate group, but only 5% in the placebo group. Herd (1997) had withdrawals of 15% and 8% in the etidronate and placebo groups respectively. In Cortet (1999), 2% and 8% (placebo) did not complete the study, but this was not considered an important difference.

Four studies (Campbell 2004; Geusens 1998; Herd 1997; Pouilles 1997) carried out an intention-to-treat analysis. The other studies did not say or used an available case analysis (Adachi 1997; Campbell 2004; Jenkins 1999; Montessori 1997; Watts 1990).

Two studies evaluated vertebral fracture as a primary outcome (Campbell 1990; Watts 1990), five assessed vertebral fracture as a secondary outcome (Adachi 1997; Geusens 1998; Herd 1997; Jenkins 1999; Montessori 1997); three studies reported fracture as an adverse event (Meunier 1997; Pouilles 1997; Roux 1998) and it was unclear in Cortet (1999).

Only one study (Geusens 1998) stated clearly that there were no traumatic fractures reported. Two studies (Pouilles 1997; Watts 1990) reported trauma and fragility fractures separately. One study (Meunier 1997) reported that all fractures were traumatic, therefore the fracture results from this study were not considered further.

The definition and method of confirming vertebral fractures varied between the studies.

In four (Adachi 1997; Campbell 2004; Montessori 1987; Watts 1990), vertebral fractures were defined as more than 20% reduction in vertebral height: Adachi (1997) examined the increase in vertebral-deformity score from baseline and used the following criteria to determine the vertebral-deformity score: grade 0: normal; grade 1, a 20 to 25% reduction in the height of the anterior, middle, or posterior dimension of the vertebral body in comparison with the adjacent vertebrae; grade 2, a 26 to 40% reduction; and grade 3, a reduction greater than 40%. Montessori (1987) used a semiquantitative vertebral-deformity score applied to individual vertebrae on the following criteria: normal, unfractured vertebra was assigned grade 0; mild fracture with 20 to 25% reduction in anterior, middle or posterior height (or all three) accompanied by a reduction in area of approximately 10 to 20% was assigned grade 1; moderate fracture with 25 to 40% reduction in any height accompanied by a reduction in area of approximately 20 to 40% was assigned grade 2; severe fracture with a reduction of more than 40% in any height and accompanying area was assigned grade 3. Intermediate grades (increments of 0.5) were assigned as appropriate. The other studies did not report how vertebral fractures were assessed, although in all but two (Meunier 1997; Pouilles 1997), a radiographic method was used.

Overall, six studies (Campbell 2004; Geusens 1998; Herd 1997; Jenkins 1999; Montessori 1987; Roux 1998) were regarded as having a higher potential for bias: three because more than 20% data was missing (Campbell 2004; Geusens 1998; Jenkins 1999); two because of differences in missing data (but less than 20% was missing) (Herd 1997; Roux 1998); two because the patients were unblinded (Campbell 2004; Montessori 1987) and one because of differences in baseline characteristics (Jenkins 1999). It was decided that the Jenkins (1999) study had too large an amount of missing data and this study was not included in the analysis. The other studies were considered in sensitivity analyses.

## **Results**

### **Etidronate versus placebo/no treatment**

All studies but two compared an etidronate/calcium regimen with a placebo/calcium regimen. The exceptions were Campbell (2004) which had no treatment and continuous calcium as comparators, and Montessori (1997), for which the comparator was continuous calcium. It was decided that the comparisons with continuous calcium could be combined as subgroups with the comparison with placebo/calcium, because the latter comparator gave calcium for most of the treatment period. In addition, subgroups are used for postmenopausal women and people taking glucocorticoids.

#### *Vertebral fractures*

Eight studies (9 comparisons) reported the number of patients with vertebral fractures (Adachi 1997; Campbell 2004 [2 comparisons]; Cortet 1999; Geusens 1998; Herd 1997; Montessori 1997; Roux 1998; Watts 1990).

Meta-analysis in 1039 patients showed there were significantly fewer people with fractures in the etidronate group (figure 8); RR 0.51 (95% CI 0.31 to 0.83). This corresponds to a number needed to treat of 25 (95% CI 13 to 100) for a control group range of 3 to 16%. There was no heterogeneity ( $I^2 = 0\%$ ,  $p = 0.98$ ) and a funnel plot was fairly symmetrical (figure 9).

Review: Osteoporosis (Version all reviews)  
 Comparison: 29 Etidronate vs placebo/no treatment  
 Outcome: 01 Vertebral Fractures

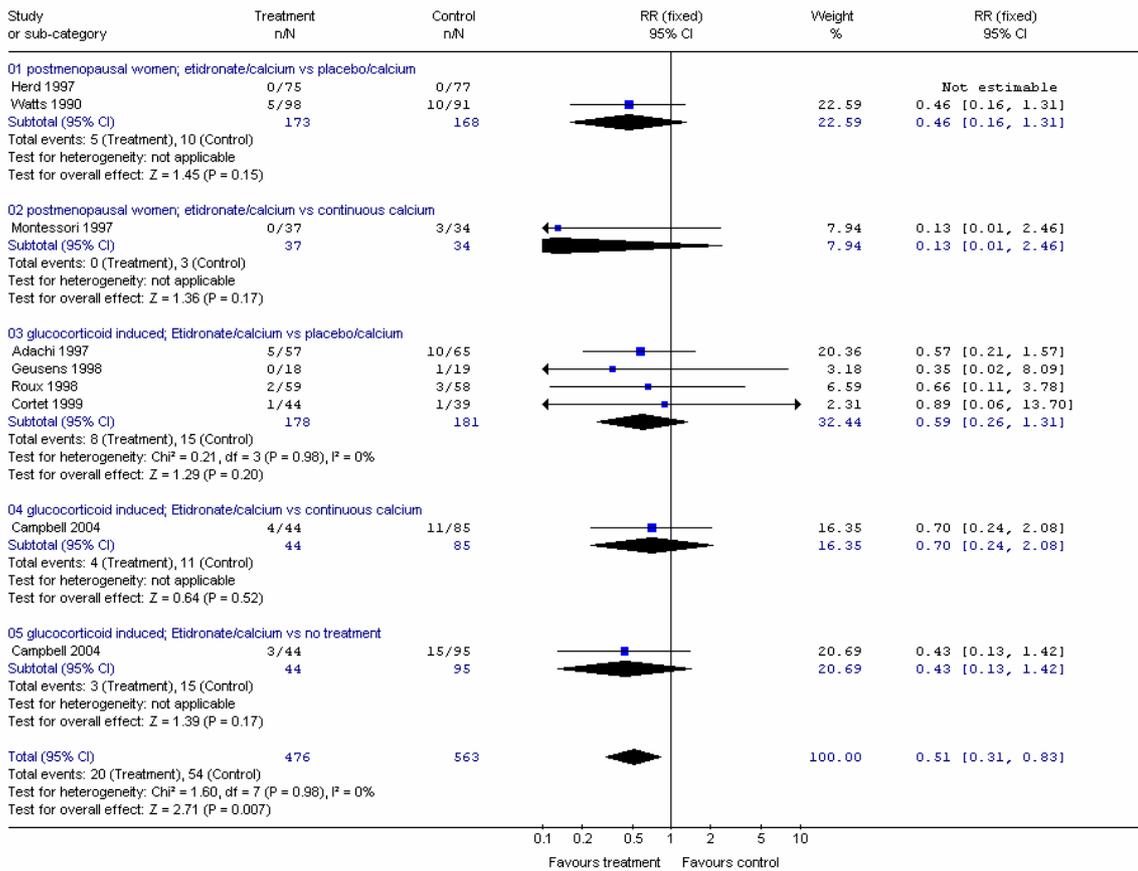


Figure 8. Number of patients with vertebral fractures

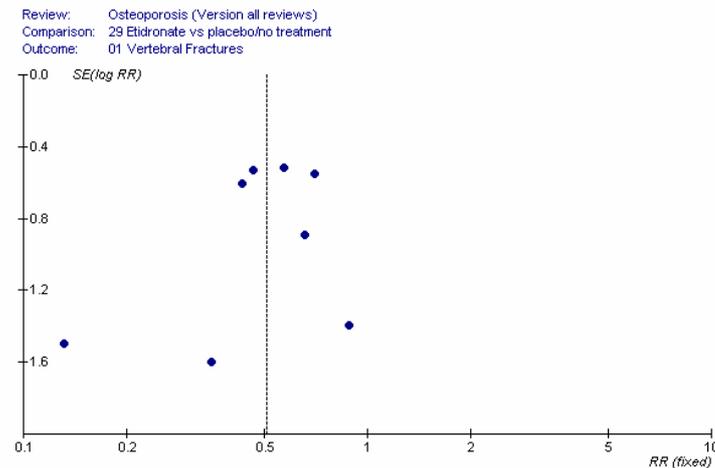
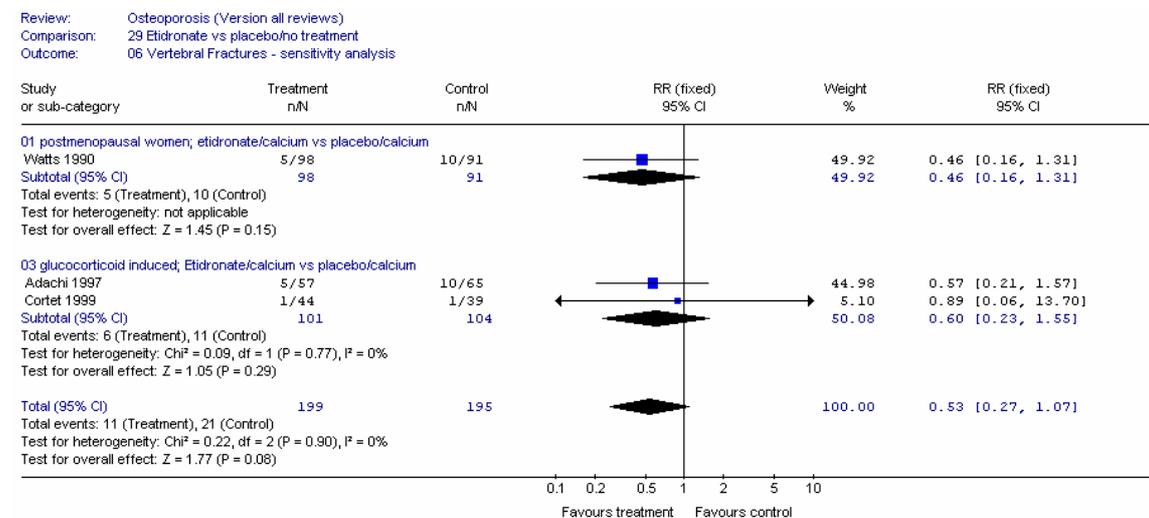


Figure 9. Funnel plot for number of patients with vertebral fractures

A sensitivity analysis omitting studies that were at higher risk of bias (Campbell 2004; Geusens 1998; Herd 1997; Montessori 1997; Roux 1998) was conducted and the overall estimate in the remaining studies was similar,

although no longer significant (figure 10); RR = 0.53 (95% CI 0.27 to 1.07). There was no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.90$ ).



**Figure 10. Sensitivity analysis: number of patients with vertebral fractures**

### *Nonvertebral fracture*

Four studies reported the number of people with nonvertebral fractures (Geusens 1998; Pouilles 1997; Roux 1998; Watts 1990). Some of the Roux (1998) data were via personal communication to the HTA authors (Kanis 2007). For Watts (1990), the study did not appear to report the number of patients with a fracture, just the numbers of each type of nonvertebral fracture; it was assumed that there was one of each type of non-traumatic fracture per person. This was thought to be a better approximation than using the number of people with all nonvertebral fractures because the majority of these were not osteoporotic.

Meta-analysis of four studies (Geusens 1998; Pouilles 1997; Roux 1998; Watts 1990) in 472 patients showed there was no significant difference between interventions (figure 11), but the confidence interval was fairly wide. There was no significant heterogeneity between studies ( $I^2 = 18\%$ ;  $p = 0.84$ ). A sensitivity analysis was carried out in the absence of the two studies with higher potential for bias; this left the two studies in postmenopausal women (figure 11). The confidence interval was too wide to determine if there was a difference between interventions, and there was some heterogeneity between studies ( $I^2 = 36\%$ ;  $p = 0.21$ ).

Review: Osteoporosis (Version all reviews)  
 Comparison: 29 Etidronate vs placebo/no treatment  
 Outcome: 02 Non Vertebral Fractures

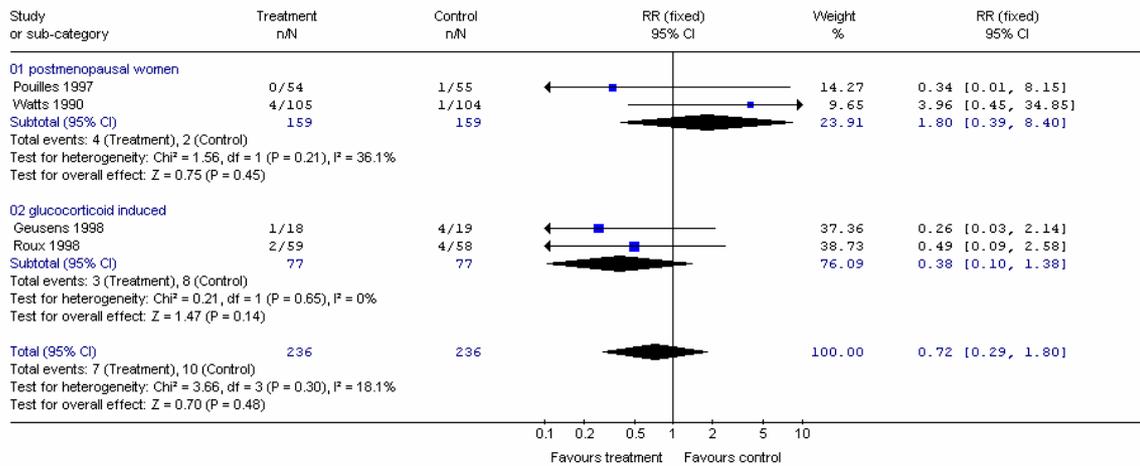


Figure 11. Number of patients with nonvertebral fractures

### Hip fracture

Two studies reported the number of patients with hip fractures (Geusens 1998; Watts 1990). Meta-analysis of the two studies in 246 patients gave a confidence interval that was too wide to determine if there was a difference between interventions (figure 12).

Review: Osteoporosis (Version all reviews)  
 Comparison: 29 Etidronate vs placebo/no treatment  
 Outcome: 03 Hip Fractures

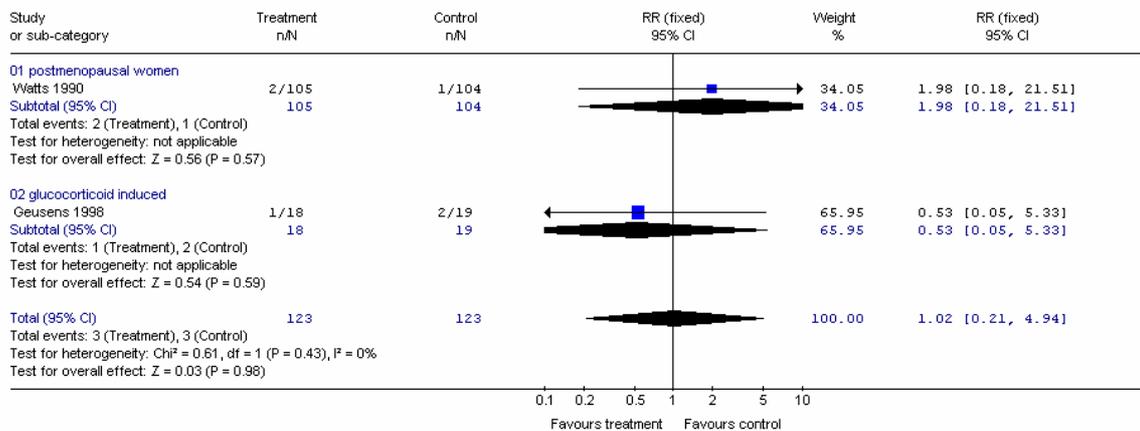
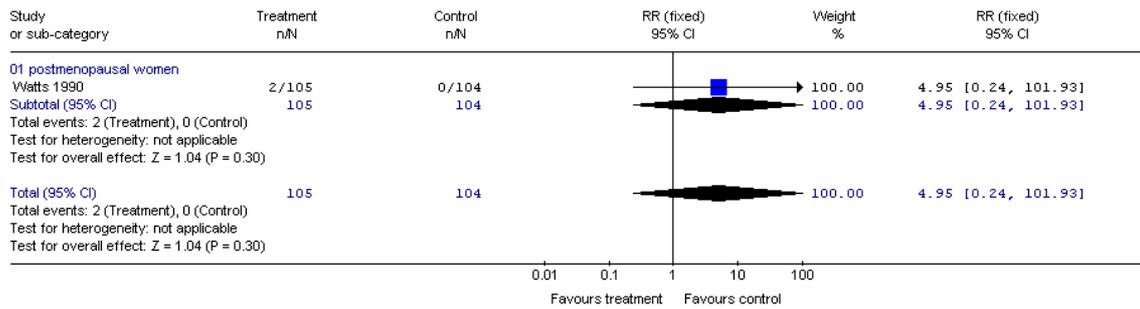


Figure 12. Number of patients with hip fractures

### Wrist fracture

One study (Watts 1990) in 209 patients reported the number of patients with a wrist fracture (figure 13). The confidence interval was much too wide to determine if there was a difference between interventions.

Review: Osteoporosis (Version all reviews)  
 Comparison: 29 Etidronate vs placebo/no treatment  
 Outcome: 05 Wrist Fractures



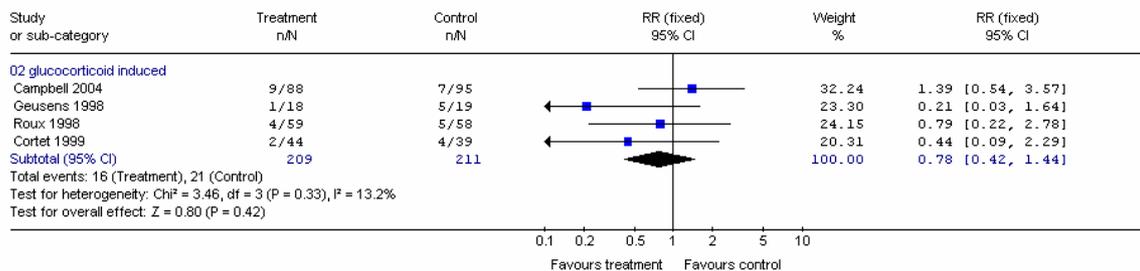
**Figure 13. Number of patients with wrist fractures**

NB forest plot scale 0.1 to 100

### All symptomatic fractures

Four studies in patients receiving glucocorticoids reported the number of people with all symptomatic fractures. Meta-analysis in 420 patients showed no significant difference between interventions (figure 14); RR 0.78 (95%CI 0.42 to 1.44). There was no significant heterogeneity between studies ( $I^2 = 13\%$ ;  $p = 0.33$ ).

Review: Osteoporosis (Version all reviews)  
 Comparison: 29 Etidronate vs placebo/no treatment  
 Outcome: 04 All symptomatic fractures



**Figure 14. Number of patients with all symptomatic fractures**

## 2.1.3 Risedronate sodium

### Description of studies

A total of 21 papers were evaluated for inclusion. Only three of these reported BMD as an outcome and were excluded from the review (Durchschlag 2006; Leung 2005; Majima 2006). Six studies were excluded because they were in patients with other conditions that were not representative. One study was excluded because the dose was not in the BNF (Clemmensen 1997). Eleven reports were included (Cohen 1999; Fogelman 2000; Harris 1999 [Vertebral Efficacy with Risedronate Therapy – VERT-NA]; Hooper 2005; McClung 2001 [Hip Intervention Program – HIP]; Mellström 2004 [VERT]; Mortensen 1998;

Reid 2000; Reginster 2000 [VERT]; Sørensen 2003 [VERT]; Ste-Marie 2004 [VERT-NA]), describing eight trials.

Three of these reports are extension studies: Ste-Marie (2004) was a 2-year extension of Harris (1999), and reported results for years 3–5 in a sample of 86/2458 (3%) patients, who continued their randomised treatments. Sørensen (2003) was a 2-year extension of Reginster (2000), and reported results for years 3–5 in a sample of 265/814 (33%) patients who continued their randomised treatments. Mellström (2004) was a 2-year extension of Sørensen (2003), in 164/265 (62%) of the patients, that is, 164/814 (20%) of the original trial, who were continued on risedronate regardless of their initial allocation. In addition, Mortensen (1998) also reported an extension trial in their original paper: patients were initially randomised to 1 year of intervention after they were offered the option of continuing with the treatment for 1 year and then receiving no treatment for a third year; 68/111 (61%) took up this option.

### ***Study design***

Two studies (Fogelman 2000; Reid 2000) were conducted in Europe and both were multicentre. Harris (1999) and Cohen (1999) were conducted in North America; Hooper (2005) was conducted as a multicentre trial in Australia; McClung (2001) was also multicentre and conducted in the USA, Canada, Europe, Australia and New Zealand. Mortensen (1998) was undertaken in the USA and Denmark; Reginster (2000) used 80 centres in Europe and Australia.

The trialists in all eight studies received industry support from Procter & Gamble, Hoechst Marion Roussel and Aventis Pharma.

McClung (2001) was the largest study with 9331 participants. Harris (1999) randomised 2458, and Reginster (2000) 1226 people. The remaining trials were smaller (Cohen 1999, n = 228; Hooper 2005, n = 383; Fogelman 2000, n = 359; Reid 2000, n = 290; and Mortensen 1998, n = 111).

## **Population**

Reid (2000) and Cohen (1999) included both men and women who were at risk of developing osteoporosis from use of corticosteroids; the proportions of men were 36% and 25–27%, respectively, and both studies also included premenopausal women (7–9% and 14–15%, respectively). Participants in these studies also had comorbidities (rheumatoid arthritis, asthma, polymyalgia rheumatica, systemic lupus erythematosus, temporal arteritis, vasculitis, chronic obstructive pulmonary disease, polymyositis, chronic interstitial lung disease and skin conditions). The other trials randomised postmenopausal women only; the mean time since the menopause ranged from around 1 to 24 years.

Three studies included people with osteoporosis or osteopenia. Participants in Harris (1999) had a mean T-score of  $-2.6$  (SD 1.1); in Fogelman (2000) the mean T-score was  $-2.8$  (SD 0.7); in Reginster (2000) the mean T-score was  $-2.8$  (SD 1.4).

Hooper (2005) and Cohen (1999) included people with normal BMD or osteopenia, with T-scores of  $-0.4$  (SD 1.1) and  $-0.4$  to  $-0.7$  (SD 1.75), respectively. Two studies (Mortensen 1998; Reid 2000) had T-scores indicative of mixed normal BMD/osteopenia/osteoporosis: Reid (2000) reported a mean T-score of  $-1.7$  (SD 1.6) and Mortensen (1998) was  $-1.8$  (SD 0.9) for risedronate and  $-1.7$  (SD 0.8) for placebo).

McClung (2001) enrolled women into two groups. One group ( $n = 5445$ ) consisted of women aged 70–79 years who had osteoporosis, either a femoral neck T-score below  $-4$  SD or below  $-3$  SD, with at least one risk factor for hip fracture: these included fall-related risk factors. The other group ( $n = 3886$ ) consisted of women aged 80 years and older, having one or more nonskeletal risk factors (as above), a femoral neck T-score below  $-4$  SD or a T-score below  $-3$  SD plus a hip-axis length of at least 11.1 cm. The two groups were randomised separately. In the younger group the mean T-score was  $-3.7$  (SD 1.6), but a mean was not given for the older group, which probably included women with osteopenia and normal BMD.

All participants included in Harris (1999) and Reginster (2000) had baseline fractures. In four studies (Cohen 1999; Fogelman 2000; McClung 2001; Reid 2000) some patients had a prevalent fracture at baseline assessment. Hooper (2005) and Mortensen (1998) did not state whether participants had fractures.

The mean age of participants across studies ranged from 51 to 71 years, with individual ages ranging from 18 to 85 years. McClung (2001) included people with an age range from 70 to more than 80 years. There were limited descriptions of the type of setting for the trials.

Use of cigarettes was reported by three studies (Fogelman 2000; Harris 1999; Hooper 2005), which all reported that some participants had a history of smoking. Ethnicity data were limited. All participants in Mortensen (1998) were white.

### ***Interventions***

All studies compared risedronate with placebo. McClung (2001) randomised two doses of risedronate 2.5 and 5 mg/day, but only reported the combined results in comparison with placebo; we have included this study even though 2.5 mg is not a licensed dose for risedronate. The other studies used 5 mg/day of risedronate. In six studies (Cohen 1999; McClung 2001; Harris 1999; Hooper 2005; Reid 2000; Reginster 2000) participants were also given calcium and vitamin D concomitantly. Fogelman (2000) gave participants calcium only. In Mortensen (1998), participants were stratified according to their calcium levels; patients were excluded if they took vitamin D at levels above 400 IU/day or if they took high-dose calcium at more than 1500 mg/day).

### ***Comparisons***

The following comparisons were carried out.

- Risedronate versus placebo: eight studies:
  - 1-year intervention period (Cohen 1999; Mortensen 1998; Reid 2000)
  - 2 years (Fogelman 2000; Hooper 2005)
  - 3 years (Harris 1999; McClung 2001; Reginster 2000).

- Risedronate versus placebo: follow-up studies:
  - Mortensen (1998), extension of Mortensen (1998): 61% chose option of having a further year of their randomised treatment, followed by 1 year without treatment
  - Ste-Marie (2004), 2-year extension of Harris (1999): 3% continued their randomised treatments
  - Sørensen (2003), 2-year extension of Reginster (2000): 33% continued their randomised treatments.
- Mellström (2004), 2-year extension of Sørensen (2003): 62% (20% of Reginster 2000) received risedronate regardless of their initial allocation.

### ***Methodological quality***

All eight studies were randomised. In Fogelman (2000), McClung (2001), Mortensen (1998), Reginster (2000) and Reid (2000) no further details of the techniques used for randomisation were reported. Harris (1999) and Hooper (2005) used computer-generated randomisation. Allocation concealment was only reported in one study (Harris 1999), assessed as partially adequate: the randomisation schedule was generated by a third party.

Hooper (2005) did not report if the trial was double blind. The other trials were all stated to be double blind. Reginster (2000) reported that patients and clinic personnel maintained blinding throughout the study, and Harris (1999) reported that investigators and other staff remained blind to treatment assignments.

Details recorded at baseline in Fogelman (2000) found participants were comparable on age, weight, height, smoking history and T-scores. In McClung (2001) baseline comparability was found on age, weight, height, years since menopause and T-scores. Mortensen (1998) reported limited baseline data that were shown to be comparable on age and BMD. Participants in Reginster (2000) were comparable on age, height, BMD and time since menopause. In Harris (1999) participants were comparable on age, weight, height, years since menopause, smoking history and T-scores. Hooper (2005) reported baseline comparability on age, years since menopause and BMD. Reid (2000)

reported comparability on age, the ratios of male to female participants between treatment groups, pre- and-postmenopausal women, corticosteroid use, and T-scores.

Sample size calculations were carried out in five studies (Fogelman 2000; Harris 1999; Hooper 2005; Reginster 2000; Reid 2000). Fogelman (2000) calculated that 180 patients per group would be needed to detect a 6% difference in change in BMD at 24 months. Reginster (2000) reported that the sample was based on an expected annual 17% vertebral fracture incidence in the placebo group; assuming a 50% dropout over 3 years, the study had a 90% power to detect a 40% reduction in fracture risk. The Harris (1999) sample was calculated to have a 90% power to detect a 40% reduction in vertebral fracture risk. Hooper (2005) calculated a sample of 96 per group in order to provide a 90% power to detect a difference of 3% in mean percentage change in lumbar BMD at 24 months. Reid (2000) calculated a 90% power to detect a group difference of at least a 3% change from lumbar baseline BMD at 12 months.

Fogelman (2000) reported missing data in each group, with 40 participants (22%) dropping out from risedronate after 2 years and 37 (21%) leaving from the placebo group. McClung (2001) reported 2197 (35%) participants leaving the study early from the risedronate groups, and 1127 (36%) from placebo. In the older McClung (2001) subgroup, there were 42% missing data. In Mortensen (1998) five people (14%) dropped out of the risedronate group, and four (11%) from the placebo group by the end of the intervention period (1 year). In Reginster (2000), 156 (38%) and 186 (46%) participants dropped out of the risedronate and placebo groups, respectively, by 3 years. In Harris (1999), 324/821 (39%) and 365/820 (45%) participants dropped out from the risedronate and placebo groups after 3 years. Hooper (2005) reported 26 participants (20%) who had missing data from the risedronate groups and 33 (26%) from the placebo group after 2 years. Cohen (1999) reported missing data for 13/76 (17%) and 19/77 (25%) in the risedronate and placebo groups, respectively. Reid (2000) did not clearly report study attrition.

Most studies used ITT analyses. McClung (2001) used available case analyses while Hooper (2005) did not clearly report using ITT, although all participants were included in the analyses.

All studies reported assessing fractures radiographically. Fogelman (2000), McClung (2001) and Mortensen (1998) did not report cut-off points for reduction in height. Cohen (1999), Harris (1999), Hooper (2005), Reid (2000) and Reginster (2000) used a 15% cut-off point in height reduction.

Fractures were reported as primary outcomes in all studies except Hooper (2005), in which fractures were recorded as safety data. Nonvertebral fractures were defined as being osteoporotic in McClung (2001) and Reginster (2000). Three studies (Fogelman 2000; Hooper 2005; Reid 2000) did not define the types of nonvertebral fracture. Harris (1999) reported that nonvertebral fractures included trauma fractures, and in Mortensen (1998) nonvertebral fracture data were given, but all were related to traumatic events (and so were not included in our analysis).

Ste-Marie 2004 was a 2-year extension of Harris (1999) in only 3% of patients, who continued their randomised treatments. We considered this sample to be unrepresentative of the original randomised participants so that the randomisation was questionable: this study was not considered further. Sørensen (2003) was a 2-year extension of Reginster (2000) in 33% patients, who continued their randomised treatments. Although this was more representative, we decided not to consider this study further either. The study by Mortensen (1998) also reported an extension trial, within their original paper, in 61% of patients who chose whether or not to continue. We considered this to be a biased sample and the extension part of the study is not considered further.

Overall, of the original studies, six (Cohen 1999; Harris 1999; Hooper 2005; McClung 2001; Reid 2000; Reginster 2000) were regarded to have potential for bias, four because they had more than 20% missing data (Cohen 1999; Harris 1999; McClung 2001; Reginster 2000) and the first three also had differences between groups in missing data. One study included patients who

had an unlicensed dose of risedronate; 50% of those receiving risedronate had 2.5 mg/day (McClung 2001). Five studies (Cohen 1999; Harris 1999; Hooper 2005; Reid 2000; Reginster 2000) were also at risk of bias for the vertebral fracture outcome, because the measurement method was based on a change of 15% height. Harris (1999) also included traumatic fractures. Therefore, these five studies were considered in sensitivity analyses. We did not consider any of the extension studies to be reliable, and these data were not analysed.

## **Results**

### **Risedronate versus placebo**

#### *Vertebral fractures*

Two of the studies (Harris 1999; Reginster 2000) reported estimates of fracture incidence using analysis of Kaplan–Meier plots; they also reported the RR calculated using Cox regression analysis. The unadjusted figures are reported here and the effect of adjustment was checked in a sensitivity analysis.

A meta-analysis of seven studies consisting of 2845 participants comparing risedronate with placebo (figure 15) found significantly fewer people with vertebral fractures in the risedronate group after 3 years of intervention; RR 0.61 (95% CI 0.50 to 0.74); this corresponds to a NNT of 17 (95% CI 12 to 25), for a control group rate range of 0–26%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.64$ ). A funnel plot (figure 16) was mainly symmetrical, with an outlier.

Sensitivity analysis with the adjusted values for Harris (1999) and Reginster (2000) gave a more efficacious RR of 0.56 (95% CI 0.45 to 0.69) (figure 17). Sensitivity analyses in the absence of the studies with potential for bias left only two small studies (Fogelman 2000; Mortensen 1999) and the full analysis was preferred.

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausalmen - bisphosphonates  
 Outcome: 09 Vertebral Fractures (risedronate)

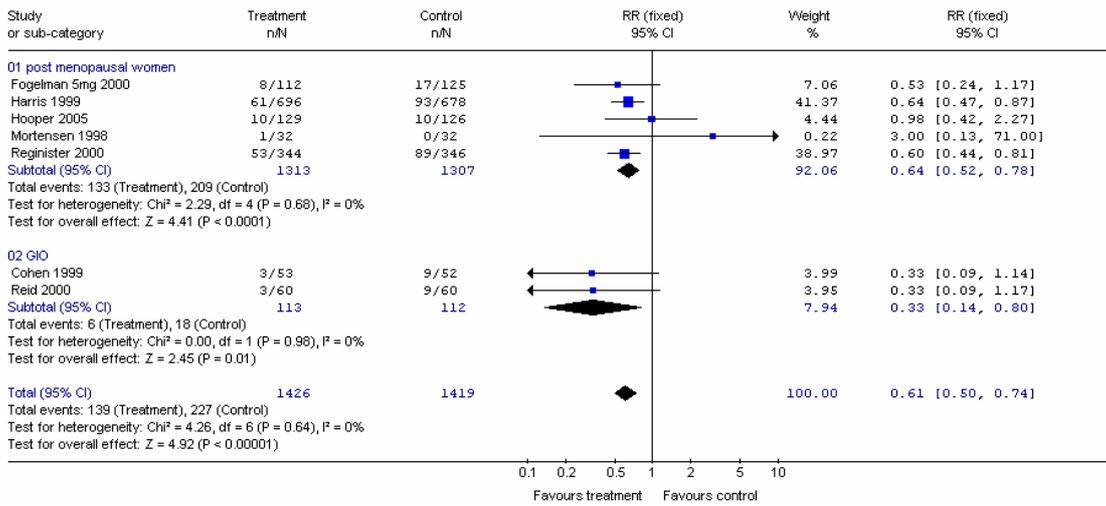


Figure 15. Number of patients with vertebral fractures

Review: Osteoporosis (Version 130608)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausalmen - bisphosphonates  
 Outcome: 09 Vertebral Fractures (risedronate)

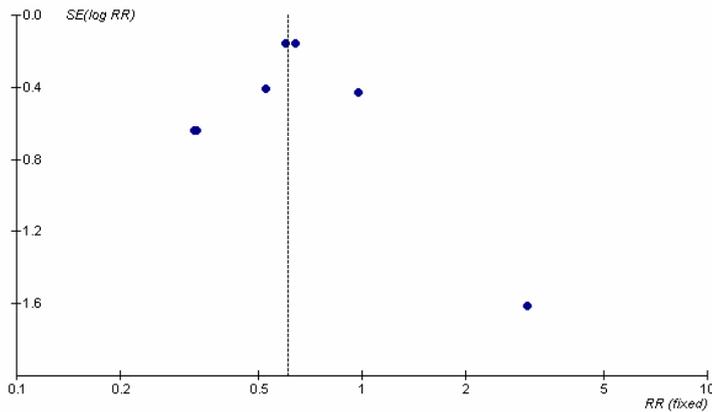
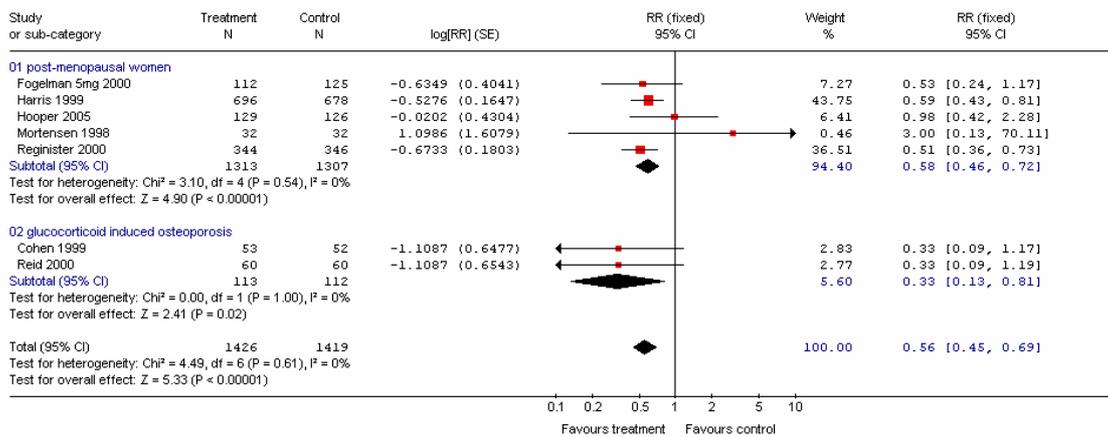


Figure 16. Number of patients with vertebral fractures – funnel plot

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausal/men - bisphosphonates  
 Outcome: 21 Vertebral fractures (regression RRs) risedronate



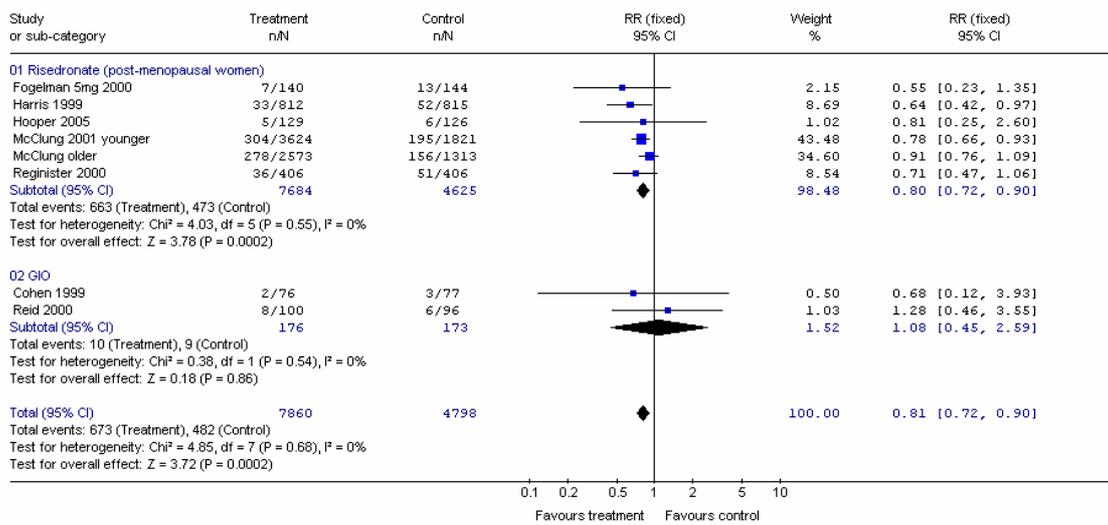
**Figure 17. Sensitivity analysis: number of patients with vertebral fractures (using adjusted RR for Harris 1999 and Reginster 2000)**

### *Nonvertebral fractures*

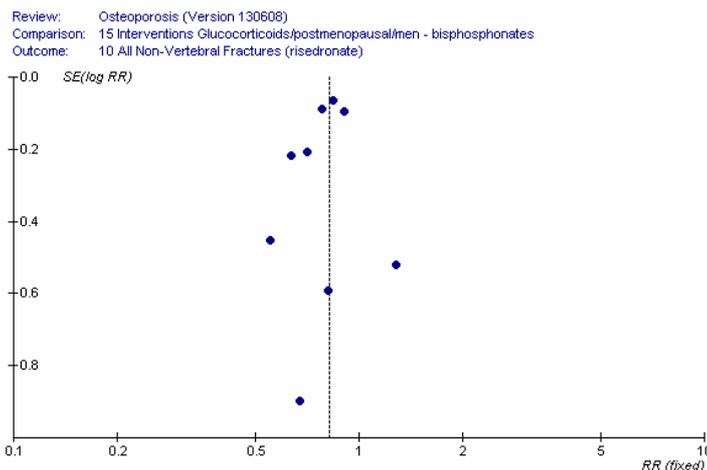
Two studies (Fogelman 2000; Reid 2000) reported the data as the number of people with fractures and percentages; these were used to calculate the denominators. The study by McClung (2001) reported percentages of nonvertebral fractures only; we assumed the denominators were the same as those given for hip fractures (that is, the number randomised) and calculated the number of events. We analysed the two strata of the the study by McClung (2001) separately. Three of the studies (Harris 1999; McClung 2001; Reginster 2000) reported estimates of the fracture incidence using analysis of Kaplan–Meier plots; they also reported the RR calculated using Cox regression analysis. The unadjusted figures are reported here and the effect of adjustment was checked in a sensitivity analysis.

Meta-analyses of seven studies in 12,658 participants comparing risedronate with placebo (figure 18) found significantly fewer people with nonvertebral fractures in the risedronate group; RR 0.81 (95% CI 0.72 to 0.90); this corresponds to a NNT of 50 (95% CI 33 to 100), for a control group rate range of 4–13%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.68$ ). Sensitivity analysis using adjusted values had no effect on the summary statistics. The funnel plot (figure 19) showed no asymmetry.

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausalmen - bisphosphonates  
 Outcome: 10 All Non-Vertebral Fractures (risedronate)



**Figure 18. Number of patients with nonvertebral fractures**

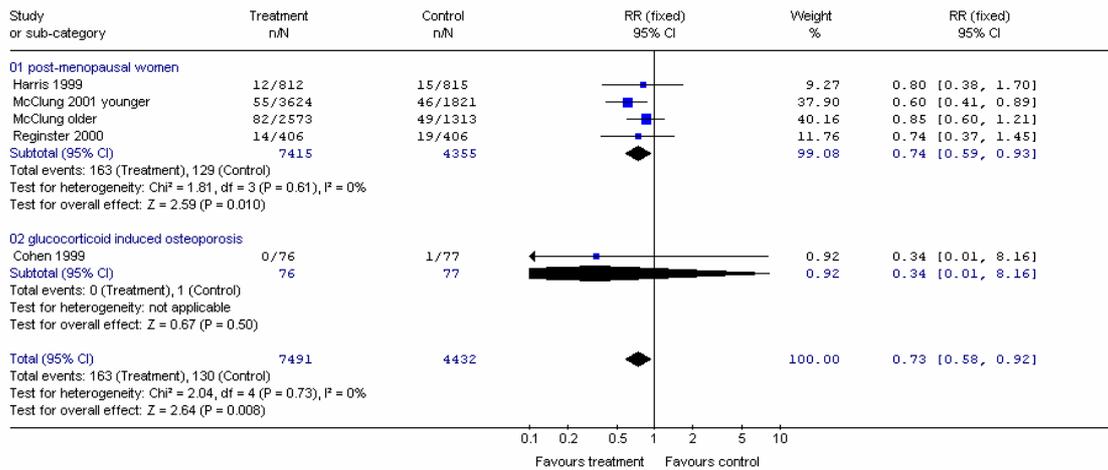


**Figure 19. Number of patients with nonvertebral fractures – funnel plot**

### Hip fractures

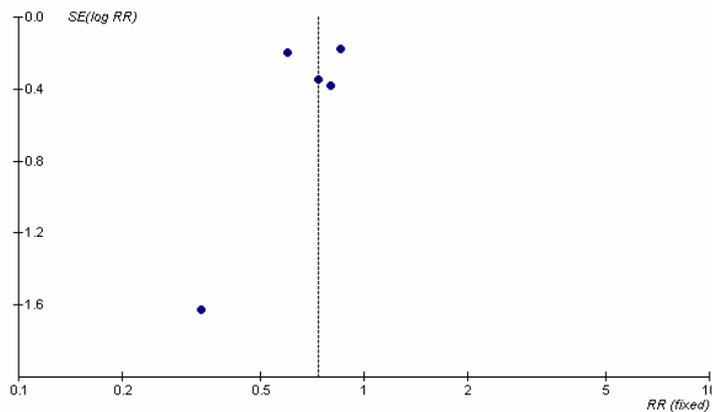
A meta-analysis of four studies in 11,923 participants comparing risedronate with placebo (figure 20) found significantly fewer people with hip fractures in the risedronate group at the end of the trial; RR 0.73 (95% CI 0.58 to 0.92); this corresponds to a NNT of 100 (95% CI 100 to ∞), for a control group rate range of 2–8%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.73$ ). All of the studies had potential for bias, so sensitivity analyses were not carried out. The funnel plot (figure 21) showed some asymmetry.

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausal/men - bisphosphonates  
 Outcome: 17 Hip Fractures - risedronate



**Figure 20. Number of patients with hip fractures**

Review: Osteoporosis (Version 130608)  
 Comparison: 15 Interventions Glucocorticoids/postmenopausal/men - bisphosphonates  
 Outcome: 17 Hip Fractures - risedronate

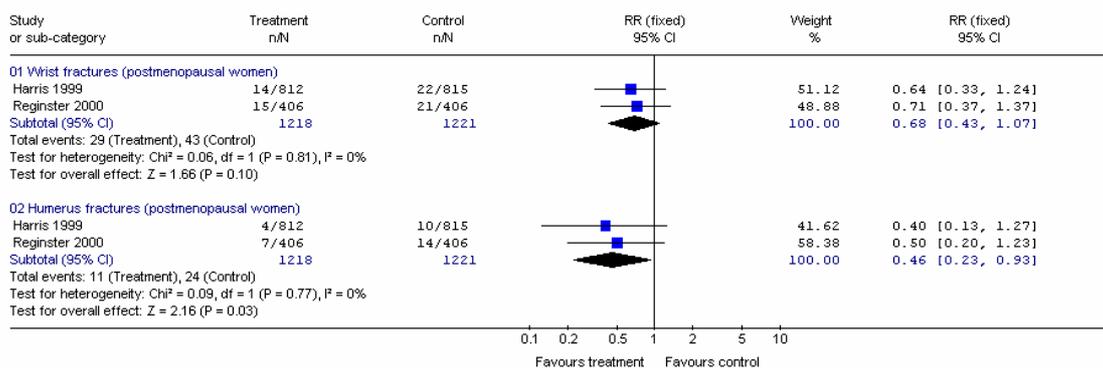


**Figure 21. Number of patients with hip fractures – funnel plot**

### Wrist fractures

A meta-analysis of two studies in 2439 participants (figure 22) found no significant difference in wrist fractures after 3 years of intervention between risedronate and placebo; RR 0.68 (95% CI 0.43 to 1.07).

Review: Osteoporosis (Version all populations)  
 Comparison: 15 Interventions: Glucocorticoids/postmenopausal/men - bisphosphonates  
 Outcome: 24 Wrist and humerus Fractures - risedronate, all populations



**Figure 22. Number of patients with wrist fractures and number with humerus fractures**

### *Humerus fractures*

A meta-analysis of two studies in 2439 participants (figure 22) showed significantly fewer people had humerus fractures in the risedronate group; RR 0.46 (95% CI 0.23 to 0.93), although the CI was fairly wide. This corresponds to an NNT of 100 (95% CI 50 to ∞), for a control group rate range of 1–3%.

## **2.1.4 Ibandronic acid**

### ***Description of studies***

Fifteen papers were evaluated for inclusion. Four of these only evaluated bone mineral density (BMD) as an outcome (Adami 2004; McClung 2004; Stakkestad 2003; Thiébaud 1997) and were excluded from the review. Two studies were excluded because the population was higher risk groups (Body 2004; Lester 2007). Another study was excluded because it evaluated two doses of ibandronic acid (0.5 mg or 1 mg iv injections every 3 months) (Recker 2004), neither of which are licensed in the UK. Eight reports of studies were included, describing three randomised trials (Chesnut 2004 [oral iBandronate Osteoporosis vertebral fracture trial in North America and Europe; BONE]; Chesnut 2005 [BONE]; Delmas 2004 [BONE]; Delmas 2006 (Dosing IntraVenous Administration; DIVA); Eisman 2008 [DIVA]; Felsenberg 2005 [BONE]; Miller 2005 [Monthly Oral iBandronate In LadiEs; MOBILE]; Reginster 2006 [MOBILE]).

The BONE study had four reports (Chesnut 2004; Chesnut 2005; Delmas 2004; Felsenberg 2005). Chesnut (2005) reported geographical subgroups of the main trial (Chesnut 2004); Felsenberg (2005) reported a post-hoc analysis restricting the outcome to vertebral fractures of greater severity and Delmas (2004) reported biochemical marker outcomes.

The MOBILE study (Reginster 2006) had two reports (Miller 2005; Reginster 2006); Miller (2005) was a 1 year interim report of the 2 year study. The DIVA study (Eisman 2008) was designed to examine BMD and provided fracture data as adverse events. Delmas (2006) was an interim study at one year and Eisman (2008) reported data at two years.

### ***Study design***

All trials were multinational. All were supported by grants from Roche (Chesnut 2005) or F. Hoffman-La Roche Ltd. and GlaxoSmithKline (Eisman 2008; Reginster 2006); all studies had at least one author who was an employee of the pharmaceutical company. In Eisman (2008), the drug company also had a role in the study design and interpretation of the data.

All were large studies ranging from 1395 to 2946 patients in total; the number in the intervention arm ranged from 401 to 982. None of the studies stated whether the trials were conducted in primary or secondary care settings.

### ***Population***

Participants in all the trials were postmenopausal women; the time since the menopause was on average, 21 years in the Chesnut (2004), and 18 to 19 years in Eisman (2008) and Reginster (2006).

Women in Chesnut (2004) had osteoporosis or osteopenia, with an average femoral neck T-score of  $-2.0$  (SD 0.9). In Eisman (2008) and Reginster (2006), all women had osteoporosis; both studies reported a mean lumbar (L2 to L4) T-score of  $-3.3$  (no SD reported) and each had an inclusion criterion of a T-score of  $-2.5$  or lower. Neither reported a mean T-score at the femoral neck.

In Chesnut (2004), 93 to 94% of the women had at least one fracture at baseline. In Eisman (2008), 43 to 44% of women had a prevalent fracture, and 47 to 49% of women had a history of previous fractures in Reginster (2006).

All three trials included women aged between 55 and 80 years, with a mean age ranging from 66 to 69 years. The average body mass index (BMI) in all trials was 25 or 26 kg/m<sup>2</sup>. None of the studies stated whether or not any of the participants smoked or consumed alcohol.

Twenty-five to 28% of women participating in Chesnut (2004) had pre-existing gastro-intestinal disorders. The other two studies did not report any comorbidities.

### ***Interventions***

The interventions included were:

- 150 mg/month ibandronic acid per month oral tablet (Reginster 2006)
- 3 mg per 3 months ibandronic acid iv injections (Eisman 2008)
- 2.5 mg/day ibandronic acid oral tablet (Chesnut 2004; Reginster 2006; Eisman 2008)
  - NB 2.5 mg/day is not currently included in the BNF, but is of interest for indirect comparisons

All three trials included more than two study arms. The following additional doses were examined, but are not evaluated in this report:

- 20 mg oral ibandronic acid every other day for 12 doses every 3 months (Chesnut 2004)
- 50 mg on two consecutive days per month oral ibandronic acid (Reginster 2006)
- 100 mg monthly oral ibandronic acid (Reginster 2006)
- 2 mg ibandronic acid iv injections every 2 months (Eisman 2008)

In all three included trials, daily 500 mg calcium and 400 IU vitamin D were provided to all participants.

## **Comparisons**

The following comparisons were carried out:

- 2.5 mg ibandronic acid per day oral tablet versus placebo for 3 years (Chesnut 2004)
- 150 mg ibandronic acid per month oral tablet (plus daily placebo) versus 2.5 mg ibandronic acid per day oral tablet (plus monthly placebo) for 2 years (Reginster 2006)
- 3 mg ibandronic acid iv injections every 3 months (plus daily oral placebo) versus 2.5 mg ibandronic acid per day oral tablet (plus iv placebo, either 2 or 3 monthly because of combining two placebo groups) for 1 year (Eisman 2008)

There were no comparisons with placebo for either of the BNF listed doses, instead the comparator for these interventions was 2.5mg ibandronic acid per day. Therefore, the studies comparing 2.5mg with placebo were also included to allow indirect comparisons to be carried out.

No studies stated if there was a washout period, although patients in all studies were excluded if they had previously received bisphosphonates.

## **Methodological quality**

The method of sequence generation was adequate in one study (Reginster 2006), which used a minimisation approach. It was not clearly reported in the other studies. Allocation concealment was adequate in two studies that used a centralised call-in system (Eisman 2008; Reginster 2006), but concealment methods were not clear in Chesnut (2004). All studies were reported to be double-blind, however, combining the placebo groups in the analysis of the Eisman (2008) study could have led to some unblinding of the patients (the two 2.5mg daily groups received iv placebo either two- or three-monthly). The outcome assessors were blinded in Chesnut (2004) for confirmation of fractures, but it was unclear if the outcome assessors were blinded in Eisman (2008) and Reginster (2006).

All studies provided details of an a priori sample size calculation: in Chesnut (2004), it was estimated that more than 2,040 patients (in total) were required to assess the incidence of new morphometric vertebral fractures; in Eisman (2008) and Reginster (2006), a sample size of 318 per group was estimated to detect lumbar spine bone mineral density (BMD) changes from baseline in order to demonstrate non-inferiority. All of these studies had sample sizes greater than those estimated.

No differences in baseline characteristics between the study groups were reported for any of the three studies. All similarly reported on the following characteristics: age, weight, height, BMI, years since menopause, lumbar spine T-score, hip T-score, prevalent fractures and biochemical markers.

In the Chesnut (2004), more than 20% of the participants did not complete treatment after three years: 334/982 (34%) in the ibandronic acid group and 354/982 (36%) in the placebo group; in Eisman (2008), 21% in the 3 monthly group and 17% in the daily group failed to complete the trial, and fewer than 20% did not complete two years of treatment in Reginster (2006). In all three studies the analyses included patients who received at least one dose of study medication and who attended at least one follow-up visit (i.e. during the trial) Around 99% of participants randomised (to treatment groups of interest) were analysed in each of the studies. Methods to assess concordance were not stated in any of the trials.

In Chesnut (2004), vertebral fracture was measured by a relative height reduction in a vertebral body of at least 20%, together with an absolute decrease from the baseline radiograph of at least 4 mm in any vertebral body height. The other studies did not state how they assessed vertebral fractures, where these were recorded.

Chesnut (2004) evaluated morphometric vertebral fracture as a primary outcome and non-vertebral fractures as adverse events; the study stated that non-osteoporotic non-vertebral fractures were excluded (defined as fractures of the hands, feet, face and skull). In Eisman (2008) and Reginster (2006), clinical fractures (vertebral and non-vertebral) were identified symptomatically,

and confirmed radiographically as adverse events; these studies did not state if the fractures were osteoporotic.

Overall, Chesnut (2004) was considered to be at higher risk of bias because of the proportion of withdrawals. There was some unblinding of patients in Eisman (2008) and the proportion withdrawing from the study also gave some risk of bias. Eisman (2008) and Reginster (2006) were also considered with caution because fractures were recorded as adverse events and they did not distinguish osteoporotic fractures.

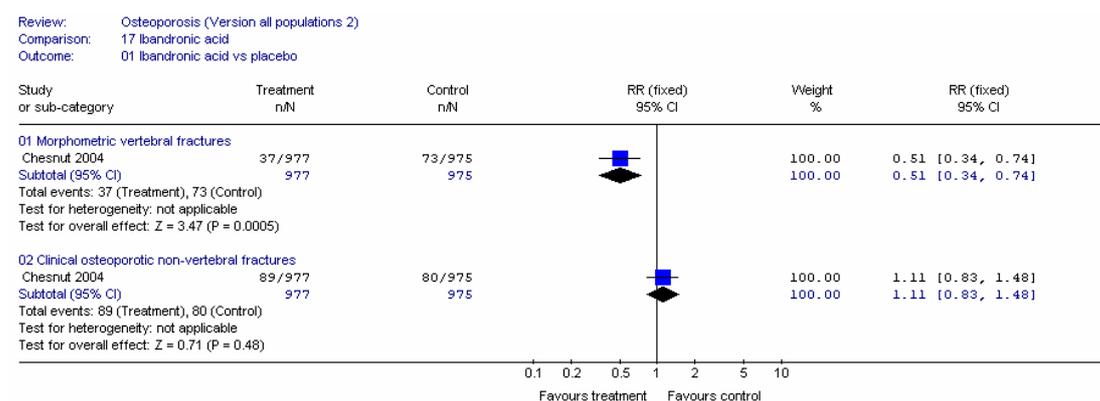
## Results

### Ibandronic acid versus placebo

One study (Chesnut 2004) compared ibandronic acid versus placebo. We note that the dose used (2.5 mg per day oral tablet) is not included in the BNF and results are given here in order to carry out indirect comparisons.

#### Vertebral fracture

One study (Chesnut 2004) in 1952 patients showed that significantly fewer patients had a vertebral fracture when taking 2.5 mg ibandronic acid in comparison to placebo (figure 23); RR: 0.51 (95%CI 0.34 to 0.74); this corresponds to a number needed to treat of 25 (95%CI 17 to 50) for a control group rate of 7%.



**Figure 23. Number of patients with vertebral fractures and number of patients with nonvertebral fractures**

The study reported that their analysis demonstrated an interaction based on the T-score above and below  $-2.0$  SD, and that a Cox regression analysis

was carried out to allow for this. The resulting relative risk reduction was: RRR 0.62 (95%CI 0.41 to 0.75).

#### *Nonvertebral fracture*

One study (Chesnut 2004) in 1952 patients showed there was no significant difference between interventions in the number of patients with non-vertebral clinical osteoporotic fractures (figure 23); RR: 1.11 (95% CI 0.83 to 1.48). Numbers of patients were calculated from percentages, although the denominators were clearly stated.

#### **Ibandronic acid dose 1 versus dose 2**

Two studies compared different doses of ibandronic acid and/or routes of administration versus 2.5mg oral ibandronic acid:

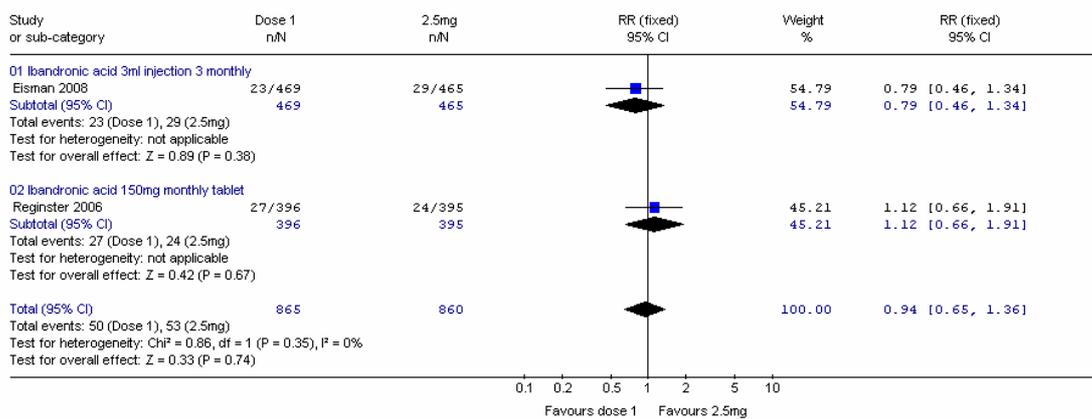
- 3 mg ibandronic acid iv injections every 3 months (plus daily oral placebo) versus 2.5 mg ibandronic acid per day oral tablet (plus iv placebo) over 2 years (Eisman 2008).
- 150 mg ibandronic acid per month oral tablet (plus daily placebo) versus 2.5 mg ibandronic acid per day oral tablet (plus monthly placebo) over 2 years (Reginster 2006).

#### *Clinical fractures*

In both studies, clinical fractures were evaluated as a measure of safety and were confirmed radiographically. Although different regimens were compared with the 2.5mg daily dose, meta-analysis was carried out. Meta-analysis of the two trials in 1,725 patients (figure 24) showed no significant difference in the number of patients with clinical fractures: RR 0.94 (95% CI: 0.65 to 1.36). There was no heterogeneity between studies ( $I^2 = 0\%$ ,  $p = 0.35$ ).

Eisman (2008) and Reginster (2006) were designed to be non-inferiority trials, but this was based on BMD outcomes; they were not powered to examine this for fracture outcomes and fractures were reported as adverse events.

Review: Osteoporosis (Version all populations 2)  
 Comparison: 17 Ibandronic acid  
 Outcome: 02 Ibandronic acid vs 2.5mg daily ibandronic acid; All clinical fractures



**Figure 24. Number of patients with all clinical fractures**

### **Indirect comparison analysis**

There were no studies comparing licensed doses of ibandronic acid versus placebo, so any comparison of the licensed doses with placebo can only be achieved indirectly. In this report, the outcome measures for the placebo and the head to head comparisons are not the same (morphometric vertebral fractures and all clinical fractures). In addition, the two studies comparing different doses are not non-inferiority trials for fracture outcomes. We decided, therefore, to use the data for the 2.5mg dose and to downgrade the quality of the evidence because of indirectness.

## **2.1.5 Zoledronic acid**

### **Description of studies**

A total of 12 papers were evaluated for inclusion. Ten studies were excluded because they were in patients with other conditions that were not representative.

Two studies were included in this review (Black 2007 [Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly – HORIZON]; Lyles 2007 [Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly Recurrent Fracture Trial – HORIZON Recurrent Fracture Trial]).

### **Study design**

Both studies were multinational; in Black (2007), approximately half of the participants were from Europe and the remainder were from North and South

America, and Asia. In Lyles (2007) patients were from Europe, North America and Latin America. The authors received research grants or consultancies from the pharmaceutical industry. Both studies were supported by Novartis Pharma.

In total, 7765 people were randomised to either zoledronic acid or placebo in the Black (2007) study and 2127 were randomised in the Lyles (2007) study..

### ***Population***

One study (Black 2007) included postmenopausal women only; the other study (Lyles 2007) included both men and women over 50 years; neither study reported the mean time since the menopause.

In the Lyles (2007) study, patients were included in the study only if they had undergone repair of a hip fracture and if they were unwilling or unable to take oral bisphosphonates. This may not have been a representative population.

Mean baseline femoral neck T-scores were reported in both studies. In Black (2007), the majority (71%) had a T-score measured at the femoral neck of less than  $-2.5$  SD; some women (28%), had a T-score between  $-2.5$  and  $-1.5$  SD, and were included only if there was radiological evidence of at least two mild vertebral fractures or one moderate vertebral fracture. The study by Lyles (2007) reported that 42% of the patients had femoral neck T-scores of  $-2.5$  SD or below, 34% had T-scores from above  $-2.5$  to  $-1.5$  and 11% had T-scores above  $-1.5$  SD. Thus the participants in each study had osteoporosis or osteopenia.

All patients in the Lyles (2007) study had a prior hip fracture (entry criterion was that they had had a surgical repair of a non-trauma hip fracture within 90 days). About 63% of the participants in the Black (2007) study had one or more vertebral fractures at baseline.

In the Black (2007) study, participants were stratified according to whether they had previously taken osteoporosis medication and were randomised within these strata. The majority (79%) were not taking osteoporosis medicines at baseline. Patient characteristics were not reported separately for

the two strata. In the Lyles (2007) study, prior use of osteoporosis medication was permitted, and in both studies patients were required to have had a washout period, the duration of which depended on the duration and type of previous medication.

In the Lyles (2007) study, patients were allowed concomitant medication at the discretion of the investigator. If the BMD value showed a decline of 8% or more in the first 12 months or 10% or more over 2 years, patients could choose to: (1) continue the study; (2) add calcitonin or HRT or tibolone or raloxifene and then continue in the study; (3) have oral bisphosphonates or teriparatide, with discontinuation of zoledronic acid, while continuing to be followed; (5) stop active participation in the study. The proportion of patients receiving concomitant medication was 9% and 12% in the zoledronic acid and placebo groups respectively; this was not a significant difference ( $p = 0.07$ ).

The mean age of participants across the studies ranged from 73 to 74.5 years. Patient ages in the Black (2007) study ranged from 65 to 89 years and the Lyles (2007) study reported that 18% of the patients were under 65 years and 14% were 85 years and older.

The BMI mean was 24.8 to 25.4 kg/m<sup>2</sup> across the studies. No data were available on history of smoking or alcohol use in either study. The setting was not reported.

### ***Interventions***

Zoledronic acid 5 mg was given as a 15-min intravenous infusion at baseline, and again at 12 and 24 months in both studies. The duration of the studies was 3 years in Black (2007) and a median of 1.9 years in Lyles (2007). In the latter, the study was terminated when efficacy had been established, based on the number of clinical fractures. In both studies, all participants received oral calcium 1000–1500 mg/day and vitamin D 400–1200 IU/day (Black 2007) and 800 IU/day (Lyles 2007). The comparison group received placebo.

### ***Methodological quality***

A partially adequate method of sequence generation was reported in Black (2007): the participants were randomised using permuted blocks. In Lyles (2007) the method was considered to be adequate because the study reported that an 'interactive voice response system created randomised permuted blocks', which suggests computer randomisation. Allocation concealment was reported only in Lyles (2007), for which a central location was described with an interactive voice-response system.

In both studies, both investigators and participants were blind to treatment allocation.

Participants in Black (2007) were comparable across the two groups on femoral neck baseline T-scores, BMD, age and BMI. Participants in Lyles (2007) were comparable on age, gender, ethnicity, BMI, BMD, and the number receiving concomitant treatment.

In the Black (2007) study, a sample size calculation was performed with a 90% power to detect a 50% reduction in morphometric vertebral fractures in the zoledronic acid group assuming an annual incidence of 1.9% in the placebo group, with 2252 participants in stratum 1 (no osteoporosis medications at randomisation). In the Lyles (2007) study, the power calculation was event driven and required 211 clinical fractures to have a power of 90% to detect 35% reduction in the rate of fracture.

In the Black (2007) study, 627 people (16%) in the zoledronic acid group did not complete the trial, and there were 592 (15%) in the placebo group. Efficacy analyses were performed on all participants except for 29 who were discontinued from one of the study sites. The Lyles (2007) study had more than 20% of patients who did not complete the trial (28% and 30% in the zoledronic acid and placebo groups respectively), and all participants were included in an ITT analysis.

Fracture data were primary outcomes in both studies, although Lyles (2007) stated that vertebral, nonvertebral and hip fractures were secondary outcomes, with any clinical fracture being the primary outcome. In both

studies, vertebral fractures were measured radiographically by an expert reader using a cut-off point of at least a 20% reduction in height. Nonvertebral fractures were said to be osteoporotic in Black (2007) and not stated in Lyles (2007), and fractures of the toe, facial bone and finger, or fractures from excessive trauma were excluded from both studies.

Both studies calculated the number of people with clinical fractures from Kaplan–Meier plots, analysed using a proportional hazards model, stratified by study stratum in Black (2007).

Overall the Lyles (2007) study was considered to have higher potential for bias because of missing data and the use of concomitant osteoporosis medications. This study was also considered to be partly unrepresentative because it was restricted to patients who were unable or unwilling to take oral bisphosphonates. The study was therefore considered in sensitivity analyses.

## **Results**

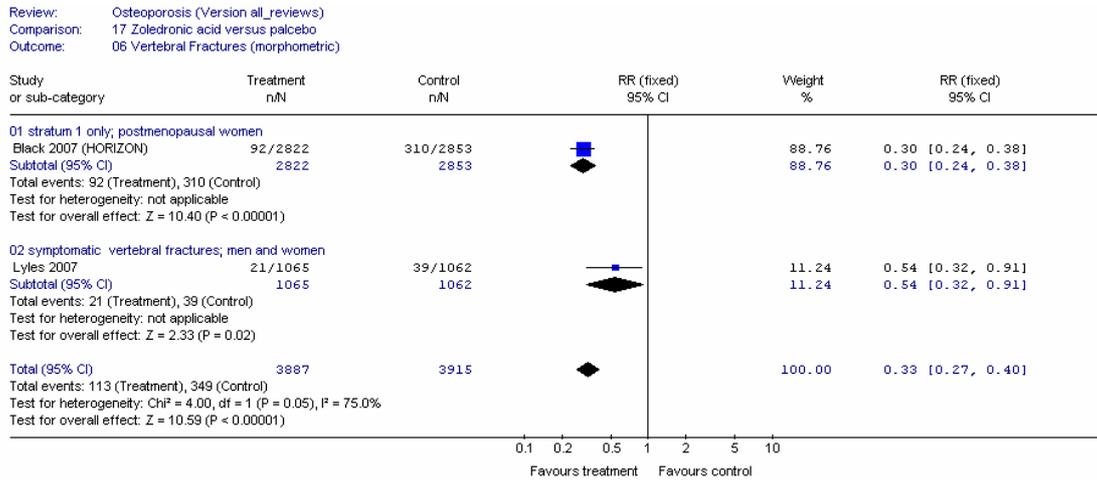
### **Zoledronic acid versus placebo**

#### *Morphometric vertebral fractures*

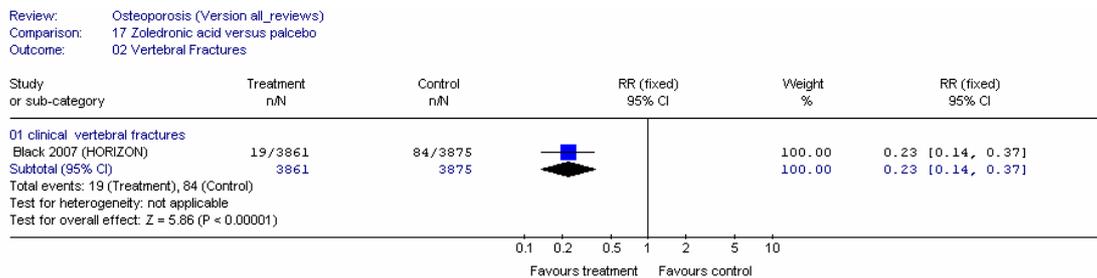
Both studies reported morphometric vertebral fractures; in Black (2007) this was carried out only for stratum 1; that is, the 5675 participants who were not taking osteoporosis medicines at the time of randomisation. In Lyles (2007), only symptomatic vertebral fractures were confirmed radiographically, which may have led to an underestimate. Meta-analysis was carried out of the two studies in 7802 patients. Each study showed there were significantly fewer people with vertebral fractures in the zoledronic acid group compared with placebo (figure 25a), but there was significant heterogeneity between studies ( $I^2 = 75\%$ ;  $p = 0.05$ ). A sensitivity analysis in the absence of the Lyles (2007) study, which had a higher potential for bias, and for which the use of concomitant medications may have diluted the vertebral fracture rate, showed significantly fewer people had a fracture in the zoledronic acid group; RR 0.30 (95% CI 0.24 to 0.38). This corresponds to a NNT of 13 (95% CI 12 to 17) for a placebo group rate of 11%.

### Clinical vertebral fractures

The Black (2007) study also reported the number of patients with clinically assessed vertebral fractures in both strata (7736 patients) RR 0.23 (95% CI 0.14 to 0.37). The NNT was 50 (95% CI 50 to 100) for a placebo group rate of 2%.



**Figure 25 a. Number of patients with morphometric vertebral fractures**

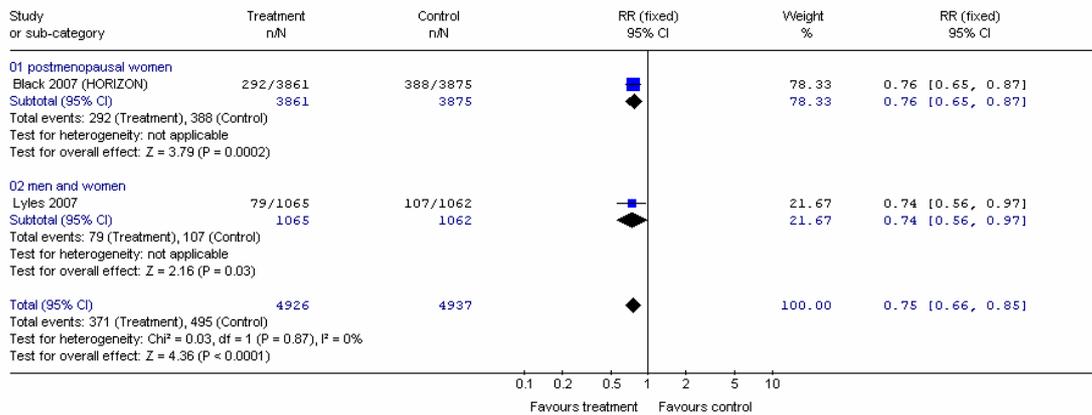


**Figure 25 b. Number of patients with clinical vertebral fractures**

### Nonvertebral fractures

Both studies reported the number of people with nonvertebral fractures, with both strata being included for the Black (2007) study. Meta-analysis of the two studies in 9863 patients showed there were significantly fewer people with nonvertebral fractures in the zoledronic acid group compared with placebo (figure 26); RR 0.75 (95% CI 0.66 to 0.85), NNT 50 (95% CI 25 to 100), for a control group rate of 10%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.87$ ). A sensitivity analysis without the Lyles (2007) study made little difference to the summary statistics.

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 17 Zoledronic acid versus placebo  
 Outcome: 03 Non Vertebral Fractures

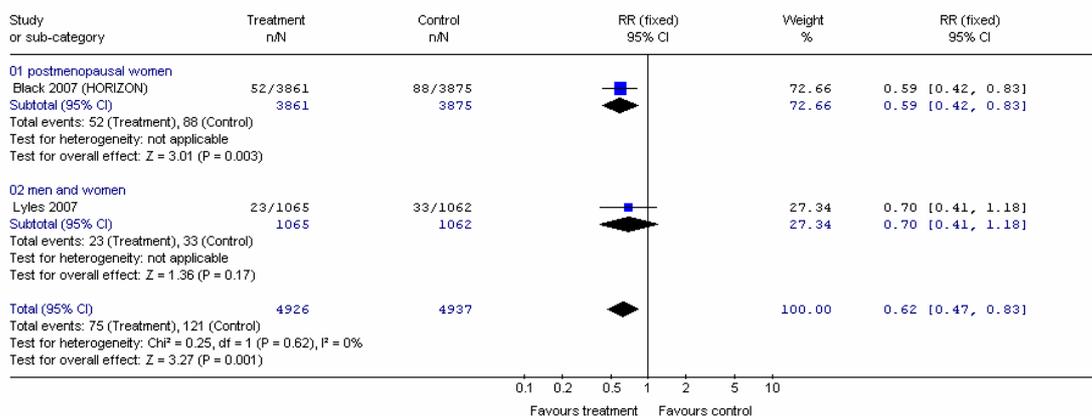


**Figure 26. Number of patients with nonvertebral fractures**

### Hip fractures

Both studies reported the number of people with hip fractures, with both strata being included in the Black (2007) study. Meta-analysis of the two studies in 9863 patients showed there were significantly fewer people with hip fractures in the zoledronic acid group compared with placebo (figure 27); RR 0.62 (95% CI 0.47 to 0.83), NNT 100 (95% CI 50 to ∞), for a control group rate of 2–3%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.62$ ). A sensitivity analysis without the Lyles (2007) study made little difference to the summary statistics.

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 17 Zoledronic acid versus placebo  
 Outcome: 05 Hip Fractures



**Figure 27. Number of patients with hip fractures**

## **2.1.6 Direct comparisons of bisphosphonates**

### ***Characteristics of studies***

A total of 12 papers were evaluated for inclusion. Two of these papers reported only on BMD as an outcome and were excluded from the review (Atmaca 2006; Yildirim 2005). Two studies (Iwamoto 2005; Kushida 2004) were excluded because the regimen was not in the BNF.

Eight studies were included, reporting five trials (Bonnick 2006 [Fosamax Actonel Comparison Trial; FACT]; Miller 2008 [Monthly Oral Therapy with Ibandronate for Osteoporosis iNtervention; MOTION]; Muscoso 2004; Reid 2006b [FACTS; multinational trial based on FACT]; Reid 2008 [FACTS]; Rosen 2005 [FACT]; Sarioglu 2006; Sebba 2004 [FACT]).

Of the eight included studies, one, Sebba (2004) reported patient characteristics and other outcome data from the FACT trial (Rosen 2005); it is considered as an adjunct paper to Rosen (2005).

Bonnick (2006) was a 12 month extension study of a sample of patients (833/1053 [79%]) from Rosen (2005). Seventy-two of the original 78 sites chose to participate in the extension study and only patients who had completed the Rosen (2005) study were eligible to continue. The demographics of the extension cohort were very similar to those of the original population. Patients continued their randomised and blinded interventions.

Reid (2008) was a 12 month extension study of a sample of patients (798/936 [85%]) from Reid (2006b); the sample constituted the proportion of completers who chose to enter the next phase. The demographics of the extension cohort were very similar to those of the original population. Patients continued their randomised and blinded interventions.

### ***Study design***

One trial was conducted in the USA (Rosen 2005); one in Turkey (Sarioglu 2006); one in Italy (Muscoso 2004); and two were multinational trials: Reid (2006b) was conducted in 27 countries throughout Europe, the Middle East,

the Americas and Asia Pacific and Miller (2008) was conducted in 65 centres in North America, Latin America, Europe and South Africa.

Three trials had funding from the pharmaceutical industry (Miller 2008; Reid 2006b; Rosen 2005) and the other two did not state their funding.

The largest study had 2000 participants (Muscoso 2004). One study had fewer than 100 patients (Sarioglu 2006), Reid (2006b) randomised 936 patients and the other trials included more than 1000 participants (Rosen 2005 n = 1053; Miller 2008 n = 1760). The Muscoso (2004) study had only 100 patients in the risedronate arm.

### ***Population***

All the studies were in postmenopausal women; the mean time since the menopause ranged from 12 to 18.5 years.

Mean baseline femoral neck T-scores were reported (or calculated) for three trials (Rosen 2005; Reid 2006b; Sarioglu 2006) and lumbar spine measures were reported in two (Miller 2008; Muscoso 2004). On this basis, two studies were assessed as being in patients with osteoporosis (Miller 2008; Muscoso 2004); two had patients with osteoporosis or osteopenia (Rosen 2005; Reid 2006b) and the other study included participants with all BMD values (Sarioglu 2006).

One study did not report whether any of the patients had vertebral fractures at baseline (Muscoso 2004). Of the four remaining studies, the percentage of patients with at least one vertebral fracture at baseline varied from 10% (Sarioglu 2006) to 39% (Miller 2008) across studies.

The mean age of the patients ranged from 57 to 71 years. Ethnicity was reported in three studies (Miller 2008; Reid 2006b; Rosen 2005): Rosen (2005) reported a distribution of 95.3% Caucasian, 0.9% Black, 1.4% Asian and 2.4% were identified as 'other' ethnicity; in Reid (2006b) 78.5% participants were identified as White, 8.8% Hispanic, 7.6% Asian and 5.1% identified as 'other' ethnicity; Miller (2008) reported 81-83% were Caucasian.

BMI was reported in two studies: Rosen (2005): 25.4 kg/m<sup>2</sup>; Sarioglu 2006: 27.4 kg/m<sup>2</sup>. Current tobacco use was reported in one study (Rosen 2005/Sebba 2004) and the proportion of smokers was 14%.

### ***Interventions***

The following interventions were used:

- Alendronic acid:
  - 10 mg/day alendronic acid (Muscoso 2004)
  - 70 mg/week alendronic acid (Miller 2008; Reid 2006b; Rosen 2005; Sarioglu 2006)
- Risedronate:
  - 5 mg/day (Muscoso 2004; Sarioglu 2006)
  - 35 mg/week (Reid 2006b; Rosen 2005)
- Ibandronic acid
  - 150 mg/month (Miller 2008)

### ***Comparisons***

The following comparisons were carried out:

- Alendronic acid versus risedronate
  - Intervention duration 1 year (Reid 2006b; Rosen 2005; Sarioglu 2006)
    - ◇ Extension of Rosen (2005) to 2 years in 79% patients (Bonnick 2006)
    - ◇ Extension of Reid (2006b) to 2 years in 85% patients (Reid 2008)
  - Intervention duration 2 years (Muscoso 2004)
- Ibandronic acid versus alendronic acid
  - Intervention duration 1 year (Miller 2008)

In all trials all patients received calcium and vitamin D, and two studies reported that some of the patients used an NSAID during the study: Reid (2006b) 28% and Rosen (2005) 41%..

In Rosen (2005), patients recorded medication use over the 12 months period of treatment and concordance was validated with tablet counts of returned medication at each study visit. The other studies did not record concordance.

### ***Methodological quality***

The method of sequence generation was adequate in two studies (Reid 2006b; Rosen 2005); both used computer generation. The other studies did not describe their methods. One study (Reid 2006b) reported a partially adequate method of allocation concealment, using numbered containers allocated sequentially; the schedule was generated by the study statistician. None of the other studies described allocation concealment.

Patients were blinded in four studies; double blinding was reported in three (Miller 2008; Reid 2006b; Rosen 2005) and one was described to be single blind (Sarioglu 2006); patient blinding was deduced in the latter because the frequency of drug dose varied. Muscoso (2004) was an open study, so the patients were not blinded. The outcome assessors were clearly blinded only in one study (Rosen 2005).

Two studies (Reid 2006b; Rosen 2005) reported power calculations for BMD outcome. In Rosen (2005) a sample size of 366 evaluable patients per group at 90% power was required to detect a treatment difference in the change from baseline in hip trochanter BMD of 1.2 percentage points between alendronic acid and risedronate groups. In Reid (2006b), the primary endpoint was change from baseline in hip trochanter BMD at 12 months and to detect a significant ( $p < 0.05$ ) difference of 1.2% change from baseline between the alendronate and risedronate groups, at 90% power, 860 were enrolled to ensure at least 732 patients would be evaluated at the end. Miller (2008) reported non-inferiority calculations on the basis of change from baseline of mean BMD. None of the trials were powered to detect fracture outcomes, and Reid (2006b) suggested that more than 50,000 participants would be needed to detect a difference in fracture rate of more than 10%.

All the studies demonstrated baseline comparability, but for Muscoso (2004) this was restricted to age only. Baseline comparability was demonstrated in relation to age in all of the studies, BMI (Rosen 2005; Sarioglu 2006), years since menopause (Reid 2006b; Rosen 2005; Sarioglu 2006); race, baseline T score and fracture history (Reid 2006b; Rosen 2005). In one report of Rosen (2005) (Sebba 2004), there was a significant difference in use of tobacco

[70/520: 284/533] with a higher number of patients in the risedronate group. However, the study also reported percentages of current smokers (13.5%: 14%), so there was a discrepancy between the percentage and the number of smokers reported in the risedronate group. We assumed that the number of patients was incorrect.

One study had missing data of more than 20% (Miller 2008): the study was not completed by 21.7% and 22.6% in the alendronic acid and ibandronic acid groups respectively. Two studies (Reid 2006b; Rosen 2005) reported missing data of less than 20%: in Reid (2006b) 8.1% and 9.4% in the alendronic acid and risedronate groups did not complete the study; in Rosen (2005), 15.8% in the alendronic acid group and 14.8% in the risedronate group discontinued the study. One study (Sarioglu 2006) stated there were missing data but did not provide numbers and it was not stated in Muscoso (2004)

None of the studies reported an intention-to-treat analysis for fracture outcomes. Miller (2008) reported an available case analysis; Reid (2006b) and Rosen (2005) reported per protocol analyses and the other studies were unclear.

The method of ascertaining fracture varied. One study based assessment of fracture on an unspecified radiographic method (Sarioglu 2006), it was not stated in one study (Muscoso 2004); and two studies (Reid 2006b and Rosen 2005) stated that fractures were reported as adverse events and confirmation of fractures by a radiographic method was not undertaken. Miller (2008) reported separately the number of people with vertebral and nonvertebral fractures, but these were recorded as adverse events and no details were given on methods of assessment.

Fracture was a secondary outcome in two studies (Muscoso 2004; Sarioglu 2006) and was reported as adverse events in the remaining studies. Only one study clearly reported osteoporotic fractures (Miller 2008) and two others ( Reid 2006b; Rosen 2005) did not distinguish osteoporotic and trauma fractures; the other studies did not state if the fractures were traumatic..

Overall, two studies (Miller 2008; Muscoso 2004) were considered to be at higher risk of bias: the patients were not blinded in Muscoso (2004) and there was slightly more than 20% missing data in Miller (2008). The studies in which fractures were recorded as adverse events (Reid 2006b; Rosen 2005; Miller 2008) or in which traumatic non-vertebral fractures were included with fragility fractures (all studies except Miller 2008) were also considered in sensitivity analyses, as appropriate. We note that the two extension studies (Bonnick 2006; Reid 2008) did not include all the randomised participants, so there is some potential for bias.

## **Results**

### **Alendronic acid versus risedronate**

Four studies compared alendronic acid versus risedronate (Rosen 2005; Reid 2006b; Muscoso 2004; Sarioglu 2006), but one (Muscoso 2004) only reported the number of fractures, so this study was not considered further. No fractures were detected throughout the study period for Sarioglu (2006).

#### *All fractures*

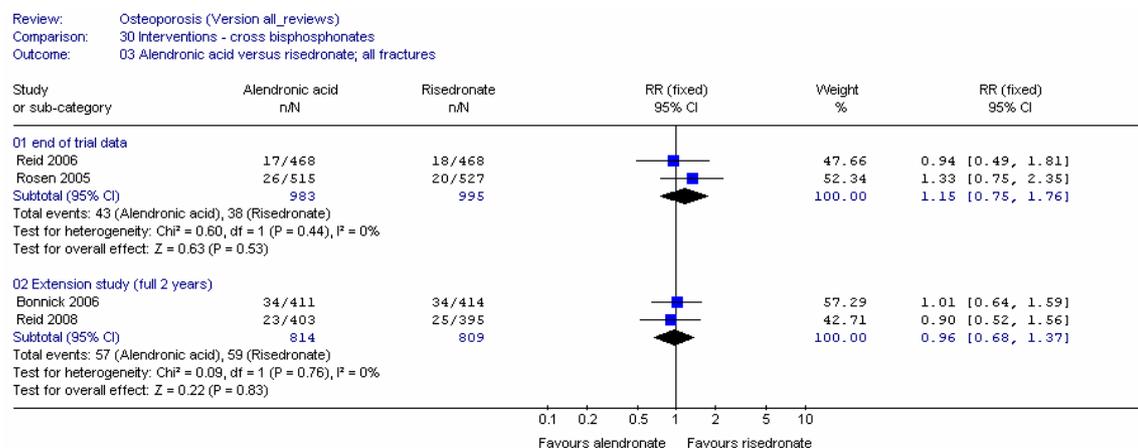
Two studies (Rosen 2005; Reid 2006b) reported all clinical fractures (including trauma) as adverse events; both studies reported results at the end of 12 months of treatment. Bonnick (2006) and Reid (2008) reported results for 12-month extensions of the Rosen (2004) and Reid (2008) studies respectively. Results were reported for the full two year period in both cases.

Reid (2006b) reported the percentage of people with fractures; it was unclear what to use for the denominator so the numbers randomised were taken, Reid (2008) clearly reported both the numerators and denominators.

Meta-analysis of the two studies (Rosen 2005; Reid 2006b) in 1978 patients showed no significant difference in the number of patients with a fracture after 1 year of treatment with either alendronate or risedronate (figure 28); RR 1.15 (95%CI 0.75 to 1.76); there was no heterogeneity between studies ( $I^2 = 0\%$ ,  $p = 0.44$ ).

Meta-analysis of the two extension studies (Bonnick 2006; Reid 2008) in 1623 patients reported the number of patients with a fracture after a one year

extension of the original trials (figure 28). There was no significant difference between interventions; RR 0.96 (95%CI 0.68 to 1.37) and no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.76$ ).



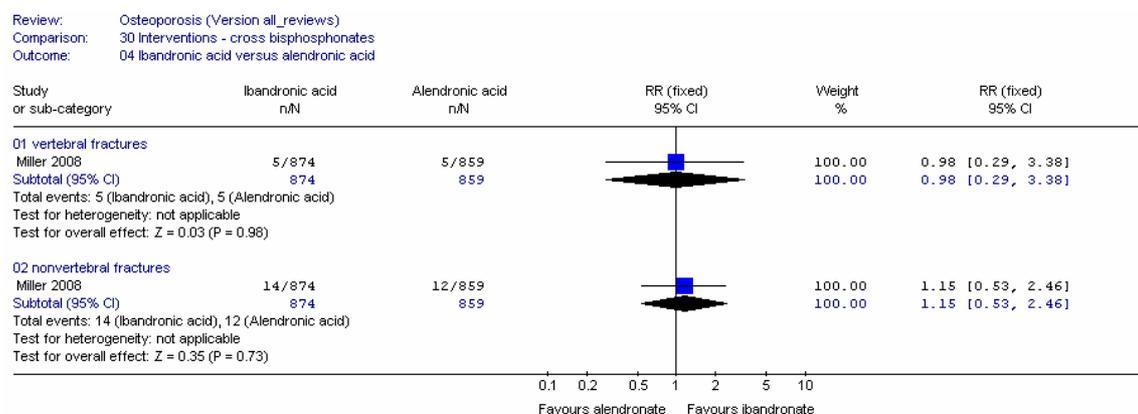
**Figure 28. Number of patients with all fractures  
 ibandronic acid versus alendronic acid**

### Vertebral fracture

One study in 1733 patients reported the number of people with vertebral osteoporotic fractures. The confidence interval was too wide to determine if there was a difference between interventions (figure 29).

### Nonvertebral fracture

One study in 1733 patients showed there was no significant difference in the number of people with nonvertebral osteoporotic fractures (figure 29); RR 1.15 (95% CI 0.53 to 2.46). The confidence interval was fairly wide.



**Figure 29. Number of patients with vertebral and number of patients with nonvertebral fractures**

## **2.2      *Strontium ranelate***

### ***Description of studies***

Five papers were evaluated for inclusion and all were included in the review. The five reports described three trials (the Treatment of Peripheral Osteoporosis–TROPOS study [Adami 2006; Reginster 2005; Reginster 2007]; the STRontium Administration. for Treatment of OSteoporosis–STRATOS study [Meunier 2002]; the Spinal Osteoporosis Therapeutic Intervention–SOTI study [Meunier 2004]).

### ***Study design***

TROPOS was conducted at 75 centres in 11 European countries and also in Australia. SOTI was also a multicentre study utilising 72 study centres in 11 European countries and also in Australia. STRATOS undertook the investigations at 31 centres in nine European countries. No details were given regarding the centres.

The trialists in all three studies received industry support from Servier. TROPOS gave interventions for 5 years, but the main results were reported after 3 years; 5-year results were included in a conference abstract (Reginster 2007). SOTI gave interventions for 3 years and STRATOS for 2 years.

TROPOS was the largest study with 5091 women randomised to either strontium ranelate or placebo. SOTI included 1649 women with osteoporosis. STRATOS was the smallest study and included 178 women. All studies were conducted in secondary care.

### ***Population***

All three studies were in postmenopausal women; the mean time since menopause ranged from 18 to 28 years across the studies.

TROPOS included women who had a mean femoral neck T-score of  $-3.13$  (SD 0.6); SOTI reported that the women had a mean T-score of  $-2.8$  (SD 0.8) measured at the femoral neck. In STRATOS the mean lumbar T-score for the strontium ranelate group was  $-3.86$  (SD 1.1) and  $-3.97$  (SD 0.95) for the

placebo group. On this basis, two studies (TROPOS and STRATOS) were classified as being in women with osteoporosis only, and the other study (SOTI) was considered to be in women with osteoporosis or osteopenia.

In TROPOS, 54–55% of the participants had existing fractures at baseline, while all the women in the other two studies (STRATOS and SOTI) had fractures at baseline. In the SOTI study some of the participants were smokers, but this was not reported in the other studies.

TROPOS included two population groups: women aged 74 or over, or aged 70–74 years with additional risk factors such as a history of osteoporotic fracture after menopause, resident in a retirement home, frequent falls, or a maternal history of osteoporotic fractures of the hip and spine; 39% were in the former group, but the patients were not stratified into these groups before randomisation. Participants in the SOTI study had a mean age of 69 years. STRATOS included women with an age range of 45–78 years.

Two studies (STRATOS and SOTI) reported BMI means of 25.4 and 26.2 kg/m<sup>2</sup>; one study (SOTI) reported that 11–12% of the patients smoked cigarettes. Descriptions of the participants' ethnicity were limited, although STRATOS enrolled white women.

### ***Interventions***

All studies compared strontium ranelate with placebo; the duration of the STRATOS trial was 2 years and the other studies lasted 3 years. Strontium ranelate was given at the licensed dose of 2 g per day in all studies; STRATOS also randomised patients to two lower doses of strontium ranelate (0.5 and 1 g per day), but the results are not reported here. All studies gave the women calcium and vitamin D concomitantly to ensure adequate levels (calcium, above 1000 mg/day).

### ***Methodological quality***

All three studies were said to be randomised. TROPOS and SOTI stated that participants were 'randomly assigned', but did not provide further details on how this was achieved. STRATOS stated that women were randomised in

blocks of four with consecutive therapeutic unit numbers allocated to patients in chronological order. None of the studies reported on allocation concealment.

All studies used double-blind methodology, although it was not made clear which of the investigators were blinded. However, SOTI stated that the evaluations of spinal radiographs were collected centrally by independent investigators and then transferred to Servier. Also, none of the studies tested the robustness of the blindness of those rating outcomes. Furthermore, none of the participants were tested to see if they knew which treatment regime they had received, but patient blinding was assumed because of the use of placebos.

In the TROPOS study, sample size calculations had been planned, but the original power calculation was rendered obsolete due to changes in statistical methods. SOTI stated that the study did not have sufficient power for adequate statistical comparison of the two groups. In the STRATOS study, sample sizes were calculated to include 65 patients per group, to ensure 90% power to detect lumbar BMD slope of 4% with a SD of 6% between placebo and strontium ranelate groups; this presupposed a treatment withdrawal rate of 20% in the first year.

TROPOS reported missing data in each group, with 867 participants (34%) dropping out from the strontium ranelate group after 3 years and 904 participants (35%) dropping out of the placebo group. SOTI reported that 200 participants (24%) in the strontium ranelate group left by 3 years and 189 people (32%) dropped out of the placebo group. Yearly vertebral X-rays were performed in 1817/2526 (72%) of the strontium ranelate group and 1823/2503 (73%) of the placebo group. STRATOS reported that 77.1% completed the study, although 12 participants withdrew due to adverse events from the strontium ranelate group and 11 from the placebo group. All three studies used ITT analyses.

TROPOS defined the ITT analysis as patients who had received at least one sachet of medication and provided at least one post-baseline assessment of

nonvertebral fracture occurrence. SOTI defined the ITT analysis as being performed on participants who had received at least one packet of treatment and for whom at least one spinal radiograph was obtained after baseline. Thus in these two studies, the last recording was carried forward. STRATOS did not elaborate any further on ITT. In two of the studies (TROPOS and SOTI), however, Kaplan–Meier graphs were drawn and the data analysed as time-to-event, with unadjusted Cox regression used to calculate the RRs. In the other study (STRATOS) the RR was calculated from the raw data.

The baseline characteristics of the women in the TROPOS study were comparable on age, time since menopause, number of prevalent fractures and BMD T-scores. SOTI reported that participants were comparable on age, cigarette smoking, years since menopause, number of previous vertebral fractures and BMI. Participants in STRATOS were comparable on age, BMI, years since menopause, number of vertebral fractures and lumbar T-scores.

Fractures were primary outcomes in all the studies. SOTI and STRATOS reported assessing fractures radiographically. TROPOS reported that radiographic vertebral fractures were not mandatory but were obtained when possible. In all the studies, the method of vertebral fracture was defined by radiographic assessment using a decrease of at least 20–25% in height; in the TROPOS study vertebral fractures were secondary outcomes.

Only two studies reported on nonvertebral fractures: SOTI implied that nonvertebral fractures measured were fragility fractures ('skull, face, finger, toe and coccygeal fractures were not considered to be osteoporotic fractures'). In TROPOS nonvertebral osteoporotic fractures were based on radiographic reports of the hospital/emergency department. Major nonvertebral osteoporotic fractures were defined as hip, wrist, pelvis and sacrum, ribs-sternum, clavicle or humerus. Fractures of the coccyx, skull, jaw, face, phalanx (fingers and toes) and ankle were not considered as being related to osteoporosis.

Overall, all of the studies had potential for bias because all had missing data greater than 20%, however this was more significant for the STRATOS study:

missing data were partly taken into account in the time-to-event analyses for TROPOS and SOTI.

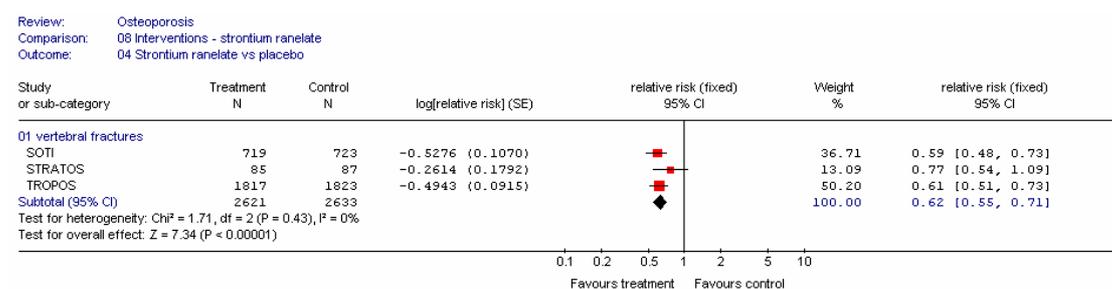
## Results

### Strontium ranelate versus placebo

#### Vertebral fractures

There was difficulty in interpreting some of the papers; for example, the papers by Adami (2006) and Reginster (2005) reported the results in various ways, some of which seemed contradictory, with disagreements between the RRs cited by the authors and the RRs we calculated using the percentage data given in the papers. This is probably because the RRs in the papers were calculated from Cox regression of Kaplan–Meier time-to-event data. Therefore we decided to use the RRs given by the authors as being more precise estimates. A meta-analysis of three studies of 5254 patients showed there were significantly fewer people with vertebral fractures in the strontium ranelate group compared with placebo at the end of the trial (figure 30); RR 0.62 (95% CI 0.55 to 0.71). There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.43$ ).

The 5-year results of the TROPOS study gave a RR of 0.76 (95% CI 0.65 to 0.87), which is less efficacious than that reported at 3 years.

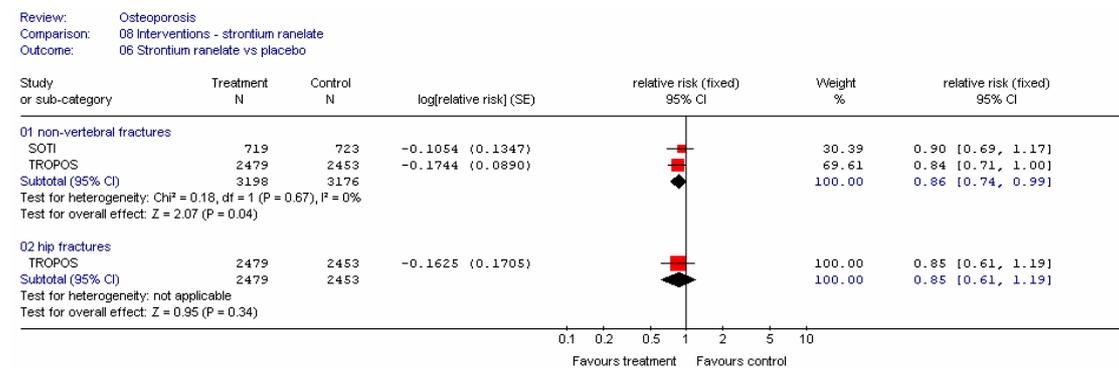


**Figure 30. Number of patients with vertebral fractures**

#### Nonvertebral fractures

A meta-analysis of two studies consisting of 6374 women showed that there were significantly fewer women with nonvertebral fractures for strontium ranelate compared with placebo after 3 years of intervention (figure 31); RR 0.86 (95% CI 0.74 to 0.99). There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.67$ ).

The 5-year results of the TROPOS study gave a RR of 0.85 (95% CI 0.77 to 0.99), which is very similar to that reported at 3 years.



**Figure 31. Number of patients with nonvertebral and hip fractures**

### *Hip fractures*

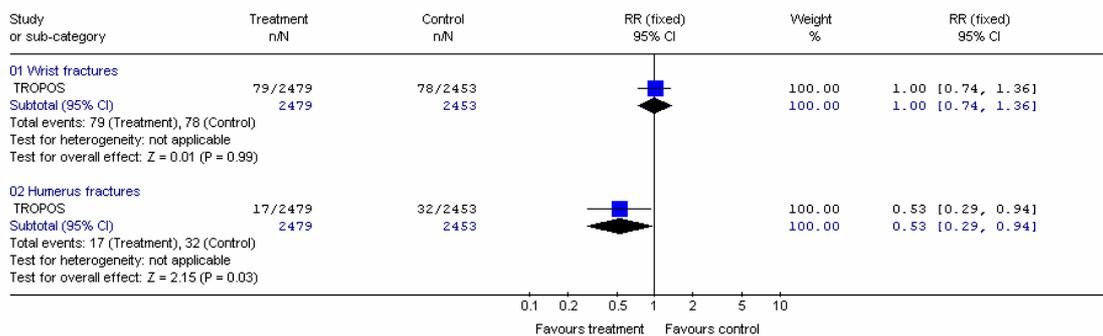
One study (TROPOS in Reginster 2005) recorded the incidence of hip fracture in all 4932 ITT patients (figure 31). There was no significant difference between interventions after 3 years of intervention; RR 0.85 (95% CI 0.61 to 1.19).

The authors of the TROPOS study have reported a subgroup analysis for the women aged 74 years and older; however, although two groups of women were specifically selected, the patients in these subgroups were not randomised separately to treatments and the results of this within-trial subgroup are not considered here.

### *Wrist fractures*

One study (TROPOS in Reginster 2005) recorded the percentage incidence of wrist fracture; the ITT population of 4932 patients was assumed (figure 32). There was no significant difference between interventions after 3 years of intervention; RR 1.00 (95% CI 0.74 to 1.36).

Review: Osteoporosis  
 Comparison: 08 Interventions - strontium ranelate  
 Outcome: 05 Strontium ranelate vs placebo



**Figure 32. Number of patients with wrist and humerus fractures**

### *Humerus fractures*

One study (TROPOS in Reginster 2005) recorded the percentage incidence of humerus fracture; the ITT population of 4932 patients was assumed (figure 32). There were significantly fewer women with a fracture in the strontium ranelate group; RR 0.53 (95% CI 0.29 to 0.94), although the CI is fairly wide.

### *Concordance*

Two studies (TROPOS and STRATOS) reported on concordance. The TROPOS study recorded a mean global compliance of 82%, similar in the two groups, and STRATOS assessed concordance by the number of tablets returned to be 85.4 and 91.2% in the strontium ranelate and placebo groups, respectively.

## **2.3 Calcitonin and parathyroid hormone**

### **2.3.1 Teriparatide**

#### ***Description of studies***

A total of 38 papers were evaluated for inclusion. One of these (Ste-Marie 2006) only reported BMD as an outcome and was excluded from the review. A further 16 studies were excluded for reasons given in appendix IV. Twenty-one reports were included (Boonen 2006 [Fracture Prevention Trial – FPT]; Cosman 2001; 2005; Crans 2004 [FPT]; Dempster 2001 [FPT]; Gallagher 2005 [FPT]; 2006 [FPT]; Genant 2005 [FPT]; Jiang 2003 [FPT]; Kaufman 2005; Kurland 2000; Lane 1998; 2000a; 2000b; Lindsay 1997; 2004 [FPT];

Marcus 2003 [FPT]; Miller 2007; Neer 2001 [FPT]; Orwoll 2003; Prince 2005 [FPT]), describing six trials.

Of the 21 included studies, 12 (Boonen 2001; Crans 2004; Dempster 2001; Gallagher 2005; 2006; Genant 2005; Miller 2007; Jiang 2003; Lindsay 2004; Marcus 2003; Neer 2001; Prince 2005) were reports of the FPT, a multinational and multicentre study looking into the effect of teriparatide on fractures; nine of these studies reported on population subgroups and are not considered further (Boonen 2001; Crans 2004; Dempster 2001; Gallagher 2005; 2006; Genant 2005; Miller 2007; Jiang 2003; Marcus 2003). Lindsay (1997) was an early report on some of the patients in Cosman (2001).

Six studies were included represented by the following papers (Cosman 2001; 2005; Kurland 2005; Lane 1998; Neer 2001 [FPT]; Orwoll 2003) and four follow-up reports (Kaufman 2005; Lane 2000; Lindsay 2004 [FPT]; Prince 2005 [FPT]).

Lindsay (2004) and Prince (2005) were follow-up studies of Neer (2001) (FPT), with follow-up periods of 18 and 30 months, respectively; 77% of patients volunteered for the follow-up studies. Patients were allowed osteoporosis drugs during the follow-up period, and 44 and 52% did so in the teriparatide and placebo groups: this is a statistically significant difference, and leads to confounding of results.

Orwoll (2003) was a study in men, and the median duration of treatment was 11 months (from less than 2 to 15 months) because the trial was stopped due to a report on adverse effects in rats. Kaufman (2005) was an 18-month follow-up study of Orwoll (2003) in volunteers; 81% (355/437) of the total number of patients from the main study were followed. Patients were allowed osteoporosis drugs during the follow-up period and a significantly higher proportion did so in the placebo group compared with the combined teriparatide groups (36 versus 25%). This also gives rise to confounding of results.

Lane (2000) was a 12-month follow-up study of Lane (1998); 100% of patients were followed. Cosman (2001) also reported a 12-month follow-up period in

their original paper, with all patients followed. It was unclear if patients were allowed osteoporosis drugs during the follow-up period in either study.

### ***Study design***

The main FPT (Neer 2001) was conducted in 99 centres of 17 countries, and Orwoll (2003) was carried out in 37 centres of 11 countries. Both trials were terminated early by the sponsors because of evidence from a long-term toxicology study, in which Fischer rats developed osteosarcomas after receiving teriparatide. However, there have been no cases of human osteosarcomas in any of the trials. The remaining four trials were conducted in the USA.

Two studies (Neer 2001; Orwoll 2003) received industry support from Eli Lilly (manufacturers); one study (Cosman 2005) had authors who were employees of pharmaceutical companies including Eli Lilly, Merck, Roche-Glaxo Smith Kline, Novartis, Pfizer, NPS, Procter and Gamble, Aventis, and Wyeth. The remaining studies were funded by public bodies.

The total number of patients in each study ranged from 23 (Kurland 2000) to 1637 (Neer 2001). Four studies (Cosman 2001; 2005; Lane 1998; Kurland 2000) had fewer than 100 patients in intervention or control arms as follows: Kurland (2000) had fewer than 20 patients in both intervention (n = 10) and control (n = 13) arms; Lane (1998) had 28 and 23 patients; Cosman (2001) had 27 and 25 patients; and Cosman (2005) had 43 and 40 patients. Orwoll (2003) had more than 100 patients in each arm and the FPT had more than 500 patients in each arm.

All the studies were randomised trials. Three took place in primary care (Lane 1998; Neer 2001; Orwoll 2003); one (Kurland 2000) included patients from secondary care; and the others included patients from both primary and secondary care (Cosman 2001; 2005).

### ***Population***

Two studies (Kurland 2000; Orwoll 2003) included only men. The other studies randomised postmenopausal women only; the mean time since the

menopause ranged from 15 to 22 years. Lane (1998) included postmenopausal women who were at risk of developing osteoporosis from use of corticosteroids; participants in this study also had a range of comorbidities (chronic noninfectious inflammatory conditions: rheumatoid arthritis, systemic lupus erythematosus, vasculitis, polymyalgia rheumatica, asthma, kidney transplant); they were treated with prednisone for at least 12 months before randomisation at an average dose of 5.0–20 mg/day.

Each study included people with osteoporosis or osteopenia. Cosman (2005) included women with osteoporosis or osteopenia who either had a T-score of  $-2.5$  SD or less at the lumbar spine, femoral neck or total hip; or had a T-score of  $-2$  SD or less plus a history of fracture in adulthood (defined as an age of at least 40 years). The FPT study included women with osteoporosis or osteopenia who had a T-score of  $-1$  SD or less (Neer 2001).

All patients in the FPT study had to have at least one moderate or two mild atraumatic vertebral fractures on radiographs of the thoracic and lumbar spine in order to be enrolled in the study (Neer 2001). The other six studies had some patients with a history of fracture at baseline. Lane (1998) reported this for 29 and 26% of the patients in the teriparatide and control groups, respectively, and Cosman (2001) reported that 58 and 46% of the patients in the teriparatide and control groups had X-ray documented prevalent vertebral fractures. Cosman (2005) had proportions of patients with prior nonspinal fracture of 60, 58 and 30%, respectively, for daily teriparatide, cyclic teriparatide and control; and prevalent vertebral fractures of 51, 54 and 49% for the same intervention groups. Kurland (2000) reported that 78% of the patients had sustained fractures. The proportion of vertebral fractures was not reported in Orwoll (2003).

The age range of patients across studies was 30–86 years, with the mean age, when given, ranging from 50 to 58.6 years. The youngest patients (30 years) were included in Kurland (2000). The youngest patients in the studies in postmenopausal women (42 years) were those included in the FPT study (Neer 2001)

Only two studies reported on ethnicity: 98–99% of patients in Neer (2001) and 99–100% in Orwoll (2003) were reported to be white.

Four studies (Lane 1998; Neer 2001; Orwoll 2003 [men]; Kurland 2000 [men]) reported BMI; in the women this ranged from 25.6 to 27 kg/m<sup>2</sup> and in men was 24.3–25.9 kg/m<sup>2</sup>

Four studies (Lane 1998; Kurland 2000; Neer 2001; Orwoll 2003) reported on smoking status. Lane (1998) reported that none of the patients were current smokers. In the FPT, the reported proportion of smokers was 18.5 and 15.8% for the treatment and control groups, respectively (Neer 2001); Orwoll (2003) reported 30–32% current smokers. Kurland (2000) reported that mean cigarette use for teriparatide and control groups was 1400 and 992 packs/year, respectively. This study also reported that 76% of the patients in the teriparatide and 69% of the patients in the control group were current consumers of alcohol; the other studies did not report on this.

The patients who volunteered to take part in the FPT follow-up studies were similar clinically and demographically, but younger, taller and with higher BMI than the nonvolunteers (Lindsay 2004; Prince 2005). Patients in these studies were also similar to the original studies in the number of baseline vertebral fractures. Kaufmann (2005) reported that patient characteristics were not different between those enrolled in the follow-up study compared with the original population, but osteoporosis drug use was significantly different between groups in the follow-up period. All patients were followed in Lane (1998) and Cosman (2001).

### ***Interventions***

The following interventions were used.

- Teriparatide subcutaneous 20 microgram self-injection once per day (Neer 2001; Orwoll 2003).
- Teriparatide subcutaneous 25 microgram self-injection once per day (Kurland 2000).

- Teriparatide subcutaneous 25 microgram self-injection once per day plus HRT (Cosman 2001; Lane 1998)
  - HRT was as prescribed in Cosman (2001) and was oestrogen in Lane (1998).
- Continuous teriparatide subcutaneous 25 microgram self-injection once per day plus 70 mg dose alendronic acid once per week (Cosman 2005).
- Cyclic teriparatide subcutaneous 25 microgram self-injection once per day plus 70 mg dose alendronic acid once per week. Each treatment cycle lasted 3 months followed by 3 months without PTJH (Cosman 2005).

Although 25 micrograms was not the dose given in the BNF, we considered this to be close enough to the licensed dose (20 micrograms) and included the studies.

Three studies had more than two study arms (Cosman 2005; Neer 2001; Orwoll 2003). The following additional doses of teriparatide were also examined in these studies, but have not been evaluated in this report: 40 micrograms once per day teriparatide (Neer 2001; Orwoll 2003).

Two studies had a washout period (Kurland 2000; Neer 2001): the FPT study gave a washout with placebo for 2 weeks before the start of the study, but also excluded patients who had received drugs that affect bone metabolism in the previous 2–24 months. Kurland (2000) gave calcium in a dose of 1500 mg/day plus 400 IU of vitamin D for a period of 12–15 months before the start of the trial. In Cosman (2001) patients were on HRT for at least 2 years before randomisation to ensure that BMD was stabilised.

In five studies (Cosman 2001; Kurland 2000; Lane 1998; Neer 2001; Orwoll 2000) all the participants were also given calcium and vitamin D concomitantly. In these studies the calcium dose was 1.5 g/day (Cosman 2001; Kurland 2000; Lane 1998) or 1 g/day (Neer 2001; Orwoll 2003). Vitamin D intake ranged from 400 to 1200 IU/day. In the other study (Cosman 2005), some patients received calcium and vitamin D. Calcium intake was maintained between 1200 and 1500 mg/day and assessed by a food

frequency questionnaire. Vitamin D supplements were provided to achieve levels of 20 nanograms/ml maximum.

During the follow-up periods of the FPT and Orwoll (2003) (Kaufman 2005; Lindsay 2004; Prince 2005), calcium and vitamin D supplements were given as needed.

In Orwoll (2003), 3.3% (5/151) of the patients in the teriparatide and 6% (9/147) of the patients in the control groups were on androgens. Concurrent medications were not stated in the other studies.

### ***Comparisons***

The following comparisons were carried out.

- Teriparatide versus placebo once a day:
  - for 11 months median in which 71% received teriparatide or placebo for at least 9 months (Orwoll 2003):
    - ◇ follow-up for 18 months of 81% of patients (Kaufman 2005)
  - for 18 months in two studies (Kurland 2000; Neer 2001 [FPT]):
    - ◇ follow-up of 77% of FPT patients for 18 months (Lindsay 2004)
    - ◇ follow-up of 77% of FPT patients for 30 months (Prince 2005).
- Teriparatide versus no treatment, with HRT in both arms:
  - for 12 months with oestrogen at a dose of 0.625 mg/day (Lane 1998)
  - for 36 months with HRT type, amount and frequency not stated (Cosman 2001).
- Teriparatide versus no treatment, with alendronic acid in both arms for 15 months (Cosman 2005):
  - Continuous teriparatide once a day in one group
  - Cyclic teriparatide once a day for 3 months, then no treatment for 3 months in one group.

### ***Methodological quality***

The method of sequence generation was adequate in five studies. Four stated that they used computer generation (Cosman 2001; 2005; Kurland 2000; Lane 1998) and Orwoll (2003) used a random number table, stratified on initial

morning testosterone measurement (normal versus low for the patient's age). The method of sequence generation was not reported in the FPT study (Neer 2001).

Allocation concealment was partially adequate in two studies: in Cosman (2001), assignment was done by one of the authors who did not have contact with patients; in Orwoll (2003), centralised randomisation was carried out, with no further details. Allocation concealment was not reported in the other studies.

All but one study (Lane 1998) reported that the outcome assessors were blinded to the interventions. Patient blinding was reported in two studies (Kurland 2000; Orwoll 2003) and assumed in the FPT study because a placebo was used (Neer 2001). In three studies the patients were not blinded (Cosman 2001; 2005; Lane 1998). All studies stated that patients self-administered the teriparatide injections.

All the studies included in the review demonstrated baseline comparability of the groups. Lane (1998) found the groups comparable on age, weight, height, years since menopause, duration of oestrogen therapy and corticosteroids, BMI, BMD and number of baseline vertebral fractures. Neer (2001) reported comparability for age, prevalent fractures, BMI, BMD, current smoking, years since menopause, previous osteoporosis therapy, calcium intake, number of vertebral fractures and lumbar spine BMD. Orwoll (2003) found the groups comparable on age, race, BMI, calcium intake, smoking and alcohol status, previous osteoporosis therapy, proportion of patients with low serum free testosterone and vertebral BMD.

Two studies were not comparable at baseline on one patient characteristic: Cosman (2001) reported that the two groups were statistically significantly different in age (the teriparatide group was younger). In Cosman (2005), there was a significant difference among intervention groups in the number of nonspinal clinical fractures during adulthood: 60% (26/43) in the daily teriparatide versus 58% (23/40) in the cyclic teriparatide versus 30% (13/43) in the placebo groups. The study stated that once fractures of toes, fingers,

feet and ankles were excluded there was no significant difference among groups. Kurland (2000) reported that the patients differed significantly in calcium intake, having 1400 mg/day in the teriparatide group versus 992 mg/day in the control group.

Sample size calculations were carried out in two studies: Kurland (2000) calculated that 50 participants in each intervention group were required to detect a 20–25% reduction in vertebral height with 80% power at a 5% significance level. Cosman (2005) calculated 33 women in each intervention group, to detect the absolute difference of 3% in spinal BMD increment between the two teriparatide (daily and cyclic) groups (Cosman 2005) at 90% statistical power. Cosman (2001) carried out a retrospective power calculation and found 15 patients per group were needed to detect changes in BMD.

Three studies (Cosman 2001; Lane 1998; Orwoll 2003) had more than 20% missing data in at least one of the groups: Cosman (2001) reported 22% (6/27) and 0% loss to follow-up in the teriparatide plus HRT and HRT groups, respectively (that is, a large difference between groups). Lane (1998) had 23.5% loss to follow-up in the teriparatide group and 12% in the control group (that is, a large difference between groups). In Kurland (2000), although there was little loss to follow-up, there were missing data for the vertebral fracture outcome because some of the radiographs could not be properly evaluated: this applied to 40% (4/10) and 8% (1/13) of the teriparatide and placebo groups, respectively (that is, a large difference between groups for this outcome). The other studies had less than 20% missing data.

Three of the follow-up studies had less than 20% loss to follow-up: Lindsay (2004) and Prince (2005) had 15% missing data; Cosman (2001) had 16% missing from the placebo group compared with 0% from the teriparatide group (that is, a large difference between groups); Kaufmann 2005 stated that 21% of the patients did not have usable radiographs and Lane (2000) had no missing data in the follow-up year.

Three studies stated they used an ITT analysis, despite missing data: Orwoll (2003) achieved this by carrying forward the fracture incidence. Neer (2001)

included all the patients for nonvertebral fracture outcomes, but did not give further details; vertebral fractures were based on the number of patients with adequate radiographs. Cosman (2001) stated that analyses were ITT, but this was unconfirmed in the results. The remaining three studies used an available case analysis.

All studies reported assessing fractures radiographically, using a decrease of 20–25% in vertebral height. Orwoll (2003) (as described in Kurland 2000) stated that the method of assessment was semi-quantitative and referred to the method of Genant.

Fractures were reported as primary outcomes in all studies except Lane (1998), in which they were secondary outcomes, and Orwoll (2005), in which fractures were recorded as safety data. Nonvertebral fractures were defined as being osteoporotic in Neer (2001); the other studies did not describe the type of fractures.

Overall, five studies had potential for bias (Cosman 2001; 2005; Kurland 2000; Lane 1998; Orwoll 2003). Cosman (2001) had a large difference in missing data between groups (22% versus 0% in the control group); there was also a significant baseline difference in age and the patients were not blinded. Furthermore, the study showed vertebral fracture data discrepancies compared with the results presented in an early report (Lindsay 1997) – one vertebral fracture was reported in the intervention group, but there were none in the later report.

Lane (1998) also had a large difference in missing data between groups (23.5% versus 12% in the control group) and the patients were not blinded. Cosman (2005) did not blind the patients. Kurland (2000) reported baseline differences in calcium intake and there was potential for bias for the vertebral fracture outcome only, because of missing data for 40% and 8% (1/13) of the patients randomised to teriparatide and control, respectively.

Orwoll (2003) had potential for bias because the study duration was not sufficiently long (11 months median): this study was included only because it

was in men (and would otherwise have been excluded from the review). Studies with potential for bias were investigated in sensitivity analyses.

Three of the follow-up studies (Lindsay 2004; Kaufmann 2005; Prince 2005) were treated with caution because there was confounding from the use of other osteoporosis drugs during the follow-up period, and volunteers were enrolled for the follow-up studies. The fracture outcomes for these studies were not reported, but the use of osteoporosis drugs was given as an outcome measure. The other two studies, Lane (2000) and Cosman (2001), were considered further.

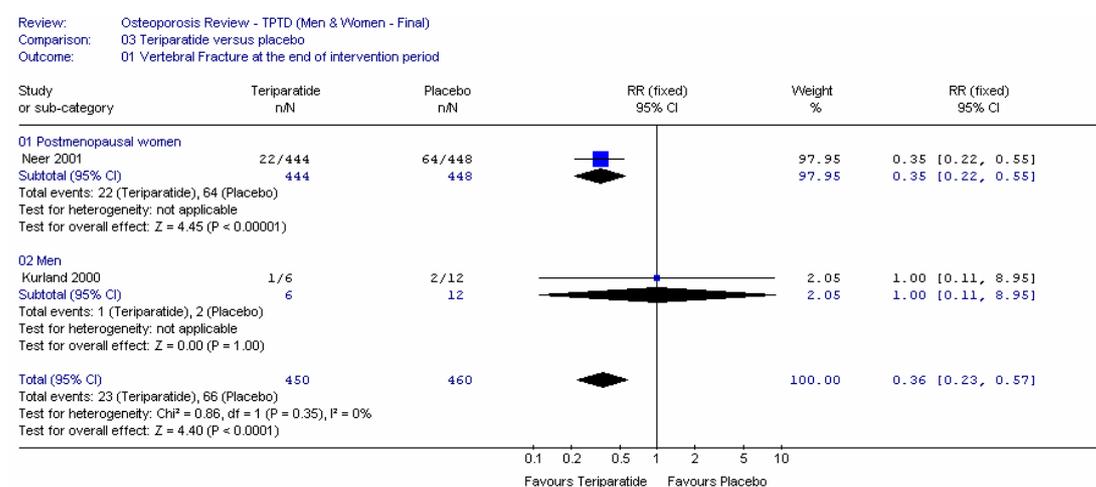
## Results

### Teriparatide versus placebo

#### Vertebral fractures

Two studies (Neer 2001; Kurland 2000) in 910 patients reported the number of patients with vertebral fractures. A meta-analysis showed there were significantly fewer patients with a vertebral fracture in the teriparatide group compared with placebo (figure 33); RR 0.36 (95% CI 0.23 to 0.57). This corresponds to a NNT of 11 (95% CI 8 to 20), for a control group rate of 14–17%. There was no heterogeneity between interventions ( $I^2 = 0\%$ ;  $p = 0.35$ ).

The study by Kurland (2000) had a significant amount of missing data in the teriparatide group for this outcome and the RR of Neer (2001) in 892 patients is preferred; RR 0.35 (95% CI 0.22 to 0.55); NNT 12 (95% CI 8 to 17).

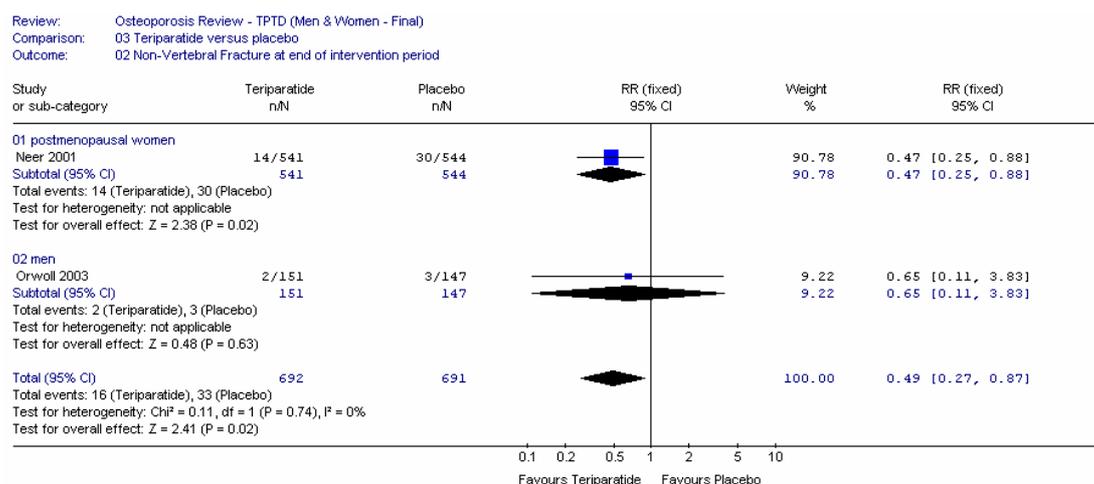


**Figure 33. Number of patients with vertebral fractures**

### Nonvertebral fractures

Two studies (Neer 2001; Orwoll 2003) in 1383 patients reported the number of patients with nonvertebral fractures. In Orwoll (2003) the denominator for the groups was not given and assumed to be the number randomised. A meta-analysis showed there were significantly fewer patients with a nonvertebral fracture in the teriparatide group compared with placebo (figure 34); RR 0.49 (95% CI 0.27 to 0.87). This corresponds to a NNT of 50 (95% CI 25 to 100), for a control group rate of 2–4%. There was no heterogeneity between interventions ( $I^2 = 0\%$ ;  $p = 0.74$ )

The study by Orwoll (2003) had a duration less than 12 months and the RR of the postmenopausal women study (Neer 2001) in 1085 patients is preferred; RR 0.47 (95% CI 0.25 to 0.88); NNT 33 (95% CI 20 to 100). The results in this study refer only to fragility fractures.



**Figure 34. Number of patients with nonvertebral fractures**

### Hip fractures

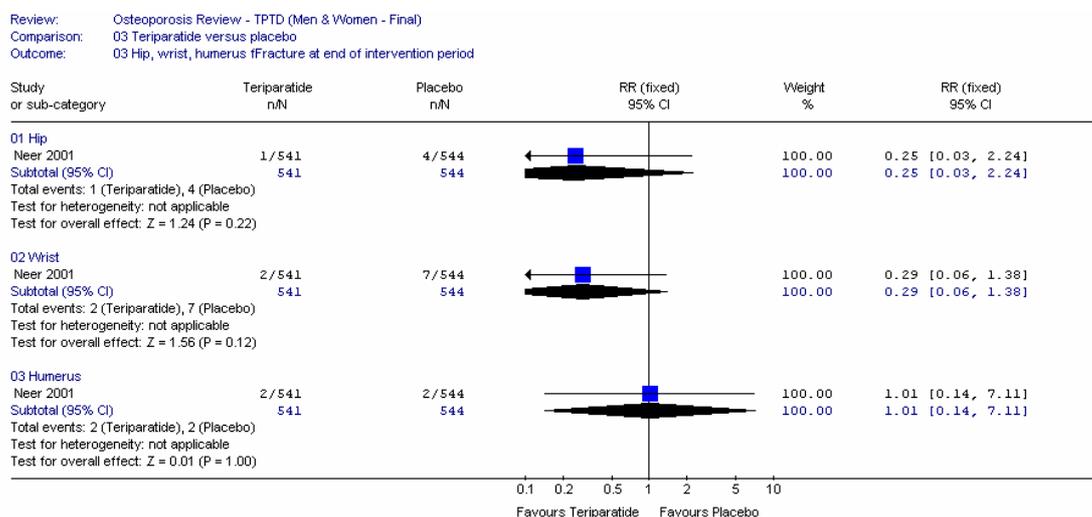
One study in 1085 patients (Neer 2001) reported the number of people with fragility hip fractures (figure 35). The CI was too wide to determine if there was a difference between interventions.

### Wrist fractures

One study in 1085 patients (Neer 2001) reported the number of people with fragility wrist fractures (figure 35). The CI was too wide to determine if there was a difference between interventions.

## Humerus fractures

One study in 1085 patients (Neer 2001) reported the number of people with fragility humerus fractures (figure 35). The CI was too wide to determine if there was a difference between interventions.



**Figure 35. Number of patients with hip fractures, wrist fractures and humerus fractures**

Use of other osteoporosis medication in follow-up after the end of the intervention period

As described above, the fracture outcomes are confounded for the three follow-up studies (Lindsay 2004 [FPT]; Prince 2005 [FPT]; Kaufmann 2005 [Orwoll 2003]) because of different use of osteoporosis medication, so the number of patients using other osteoporosis medication is recorded as an outcome measure. The three studies included volunteers from the main study as follows: Lindsay (2004) and Prince (2005) both included 77% of the participants in Neer (2001); and Kaufmann (2005) included 81% from Orwoll (2003). In the latter, the median duration of the main study was only 11 months.

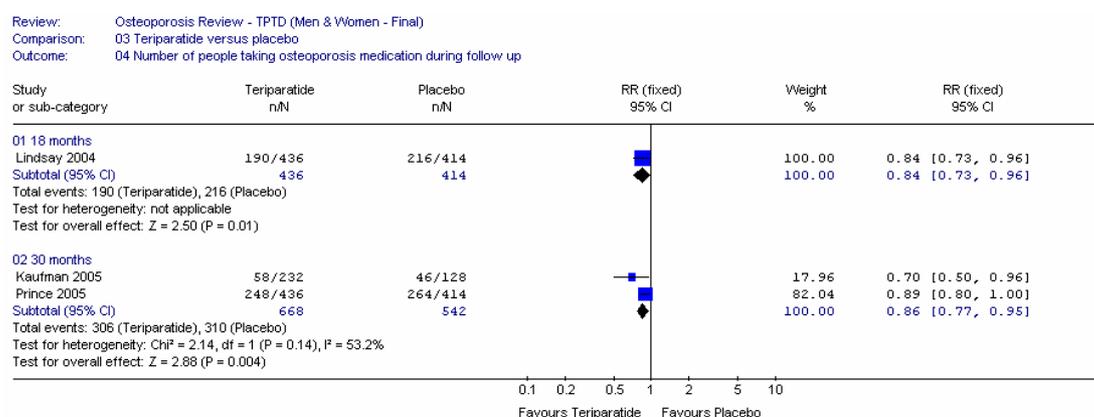
In the study by Kaufmann (2005) the denominators for the number of men in each group were calculated from percentages and the number taking medication. In addition, results were given only for the combined 20 and 40 microgram groups. Therefore this study is treated with caution.

### (a) At 18 months

One study (Lindsay 2004) in 850 patients showed there were significantly fewer patients requiring other osteoporosis medications in the follow-up period in the teriparatide group (figure 36); RR 0.84 (95% CI 0.73 to 0.96), which corresponds to a NNT of 12 (95% CI 7 to 50) for a control group rate of 52%.

### (b) At 30 months

A meta-analysis of two studies, one in men and one in postmenopausal women, totalling 1210 patients, showed there were significantly fewer patients requiring other osteoporosis medications in the follow-up period (figure 36); RR 0.86 (95% CI 0.77 to 0.95), which corresponds to a NNT of 13 (95% CI 8 to 34) for a control group rate range of 36–64%. There was some heterogeneity between studies ( $I^2 = 53\%$ ;  $p = 0.14$ ). In the absence of the study by Kaufmann (2005), which had more potential for bias, the RR was 0.89 (95% CI 0.80 to 1.00), that is, of borderline significance.



**Figure 36. Number of patients requiring other osteoporosis medication during follow up**

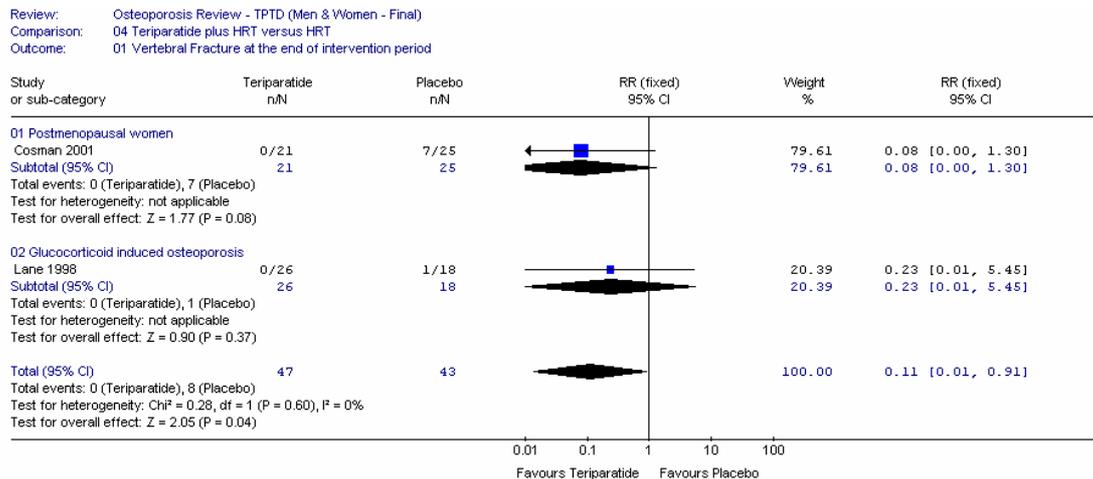
### Teriparatide plus HRT versus HRT

Two studies (Cosman 2001; Lane 1998) reported the adjunctive effect of teriparatide to HRT.

## Vertebral fractures

### (a) At the end of the intervention period

Two studies in 90 patients reported the number of patients with vertebral fractures at the end of the intervention period (Cosman 2001; Lane 1998). A meta-analysis showed there were significantly fewer patients with a fracture in the teriparatide group (figure 37); RR 0.11 (95% CI 0.01 to 0.91), but the CI is very wide. This corresponds to a NNT of 6 (95% CI 4 to 20) for a control group rate of 10–28%. There was no heterogeneity between groups ( $I^2 = 0\%$ ;  $p = 0.60$ ). There was some potential for bias in both studies because of large differences between groups in missing data, which may have confounded the results, and because of the lack of blinding.



**Figure 37. Number of patients with vertebral fractures**

NB forest plot scale 0.01 to 100

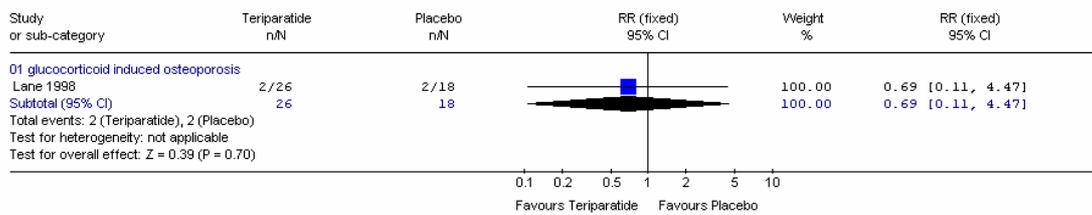
### (b) At follow-up of 12 months

In Lane (2000), 44 women were followed for 12 months after the end of the intervention period. The study reported that none of the women had new vertebral fractures.

## Nonvertebral fractures

One study in 44 women reported the number with nonvertebral fragility fractures at the end of the intervention period (Lane 1998). The CI was too wide to determine if there was a difference between interventions (figure 38).

Review: Osteoporosis Review - TPTD (Men & Women - Final)  
 Comparison: 04 Teriparatide plus HRT versus HRT  
 Outcome: 02 Non-Vertebral Fracture at end of intervention period



**Figure 38. Number of patients with nonvertebral fractures**

*All clinical fractures*

One study in 41 women (Cosman 2001) reported that no patients in either group had clinical fractures at the end of the intervention period.

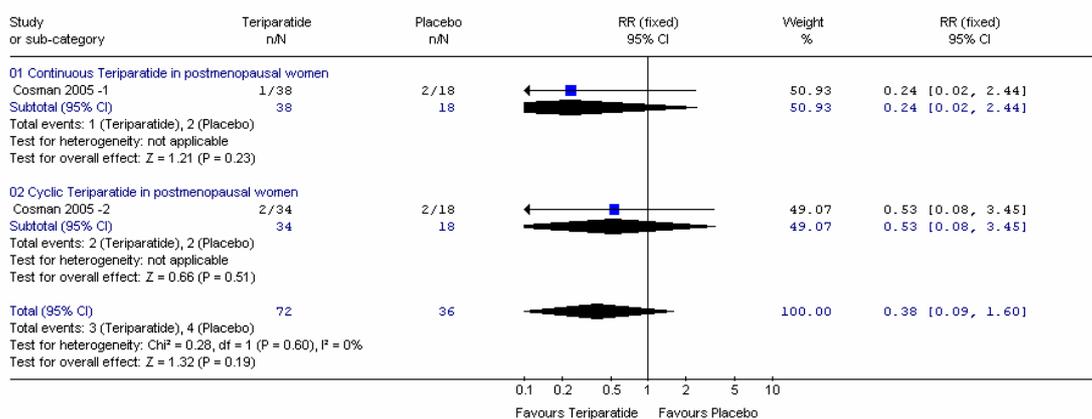
**Teriparatide plus alendronic acid versus alendronic acid**

One study (Cosman 2005) reported the adjunctive effect of teriparatide to alendronic acid. Two types of teriparatide intervention were used, daily and cyclic.

*Vertebral fractures*

One study in 108 women reported the number with vertebral fractures at the end of intervention period (Cosman 2005). The two comparisons were combined in a meta-analysis, but the CI was too wide to determine if there was a difference between interventions (figure 39).

Review: Osteoporosis Review - TPTD (Men & Women - Final)  
 Comparison: 05 Teriparatide plus alendronic acid versus alendronic acid  
 Outcome: 01 Vertebral Fracture at the end of intervention period

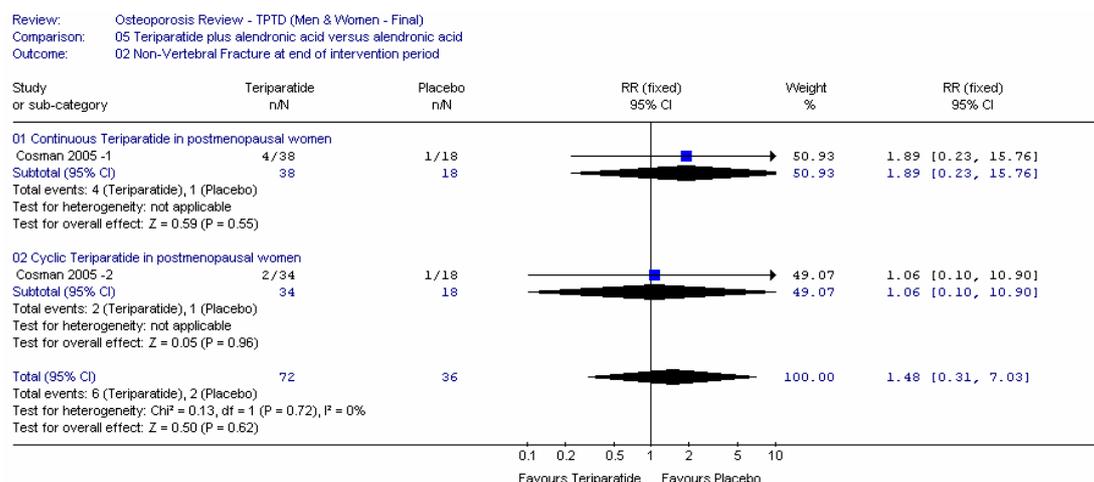


**Figure 39. Number of patients with vertebral fractures**

*Nonvertebral fractures*

One study in 108 women reported the number of patients with clinical nonvertebral fractures at the end of the intervention period (Cosman 2005).

The two comparisons were combined in a meta-analysis, but the CI was too wide to determine if there was a difference between interventions (figure 40).



**Figure 40. Number of patients with nonvertebral fractures**

### *Concordance in all studies*

Five studies reported on concordance; in most cases it was assessed by recording counts of medication returned at each visit. The following rates were reported: Neer (2001): 79–83%; Lane (1998): 80–90%; Cosman (2001): at least 90% of the distributed teriparatide was completed in both groups and concordance with hormone therapy was greater than 90% in both groups assessed; Kurland (2000): one patient in the treatment group had poor concordance in the first year and missed the equivalent of 4 months of therapy, but concordance improved in the final 6 months of the treatment protocol; Orwoll (2003): 79%.

Cosman (2005) assessed concordance with treatment by counts of teriparatide vials that were returned and by reviewing patient diaries; alendronic acid concordance was assessed by patient interview. The study reported that adherence to all treatment regimens exceeded 90%.

### **2.3.2 Parathyroid hormone (1-84)**

#### ***Description of studies included in the review***

Three papers were evaluated for inclusion. None of them reported only BMD as an outcome. One study was excluded and is described in appendix IV. One other study was excluded from this review (Black 2005), but was included in

other reviews. Two studies were included in this review (Black 2003 [PaTH]; Greenspan 2007 [Treatment of Osteoporosis with Parathyroid hormone [TOP]).

The PaTH trial (Black 2003; 2005) was designed to be of 2 years duration. Patients were randomised to 1 year of the following interventions: (1) PTH 1-84; (2) PTH 1-84 plus alendronic acid (combination group); and (3) alendronic acid. Then for the second year of the trial, patients in group (1) were randomised to alendronic acid (group 1a) or placebo (group 1b) and the patients in groups (2) and (3) were all given alendronic acid. A total of 94% of patients completed the 2 years. The first year of this study was reported in Black (2003) and the second in Black (2005). Both studies have previously been reported in the alendronic acid review: comparisons (2) versus (1) in Black (2003) and (1a) versus (1b) in Black (2005). For the PTH (1-84) review, there are no relevant comparisons in Black (2005), but the first year comparison of (2) versus (3) in Black (2003) is relevant to this review and is reported here.

### ***Study design***

The PaTH trial was conducted in four clinical centres in the USA. The TOP trial was conducted in 168 centres and nine countries, none of which included the UK (Greenspan 2007).

NPS pharmaceuticals supported, designed and conducted the TOP trial (Greenspan 2007). Most authors in this trial were employed by NPS; some were employed by NPS, Roche Labs, Glaxo Smith Kline, Merck, Procter and Gamble, Aventis, and Pfizer. Three authors had ownership options other than mutual funds in NPS and Pfizer. Black (2003) was supported by the National Institute of Arthritis and Musculoskeletal and Skin Disorders, but Merck provided supplementary funds to perform computed tomography. Greenspan (2007) stated that none of the pharmaceutical companies supporting the trial had any role in the design or interpretation of the findings.

Black (2003) had 238 patients and Greenspan (2007) was a large trial of 2679 patients. Both were carried out in primary care.

## **Population**

Participants in both studies were postmenopausal women only. The range of means for time since the menopause was 18–23 years (calculated for Black 2003).

Both studies included women with osteoporosis or osteopenia, with the overall T-score being (mean femoral neck)  $-2.2$  (SD  $0.7$ ). The inclusion criteria for Greenspan (2007) depended on age: for women aged 45–54 years, a T-score of  $-3.0$  SD or less at the lumbar spine, femoral neck or total hip with no prevalent vertebral fracture; or a T-score of  $-2.5$  SD with one to four vertebral fractures before enrolment. For those aged 55 years and older, the inclusion criteria were a T-score of  $-2.5$  SD with no vertebral fractures; or a T-score of  $-2.0$  SD if they had one to four vertebral fractures. The inclusion criteria for Black (2003) were a T-score of  $-2.5$  SD at the femoral neck, total hip or spine; or a T-score of  $-2.0$  SD at one of these sites plus at least a history of menopausal fracture, or maternal history of hip fracture or age of 65 years; or a combination of these risk factors.

Some patients in both studies had fractures at baseline: Greenspan (2007) had mean proportions of 18.7 % prevalent vertebral fractures, and Black (2003) reported that 50.8 and 41.7% of the patients had a clinical fracture (vertebral and nonvertebral) at baseline in the intervention and control groups respectively.

The age range of patients across studies was 45–89 years, with the mean age ranging from 64 to 71 years. In the TOP study, 85% of the patients were white (Greenspan 2007) while in Black (2003) the figure was 97%.

Both studies reported details on BMI. The reported mean BMI was 25.6 and 25.7 kg/m<sup>2</sup> for the treatment and control groups, respectively, in the TOP trial (Greenspan 2007) and 27.1 and 25.1 kg/m<sup>2</sup>, respectively, in Black (2003). Neither study reported on smoking status.

### ***Interventions***

The following interventions were used.

- PTH (1-84) subcutaneous 100 microgram self-injection once per day (Greenspan 2007).
- PTH (1-84) subcutaneous 100 microgram self-injection once per day and alendronic acid 10 mg tablet once per day (Black 2003).

### ***Comparisons***

The following comparisons were carried out.

- PTH (1-84) versus placebo for 18 months (Greenspan 2007).
- PTH (1-84) plus alendronic acid (10 mg/day) versus placebo plus alendronic acid for 12 months (Black 2003).

Black (2003) reported a run-in phase for 2 weeks before randomisation; it was not stated what the patients were given in this period. Greenspan (2007) reported the patients were stabilised for 2 weeks before randomisation, on supplemental calcium, citrate and vitamin D.

In Greenspan (2007), all patients received calcium and vitamin D; the regimen consisted of supplemental 700 mg/day calcium and 400 U/day of supplemental vitamin D3 during the intervention period (1.5 years). Black (2003) provided a regimen that consisted of 500 mg of elemental calcium carbonate and a multivitamin containing 400 IU of vitamin D for all patients.

None of the trials reported on other (or concurrent) osteoporosis drugs.

### ***Methodological quality***

The method of sequence generation was adequate in Greenspan (2007), which used a computer-generated algorithm. Sequence generation was not stated in Black (2003).

Allocation concealment was adequate in Greenspan (2007), in which treatment was uniquely numbered and randomly and sequentially assigned by

telephone to women at each site. Allocation concealment was not stated in Black (2003).

Both studies stated that patients self-administered their own regimen of daily injections of PTH (1-84), placebo, and/or oral alendronic acid, and were blinded to treatments. The outcome assessors were blinded in Greenspan (2007), however it was not stated if they were blinded in Black (2003).

Both studies reported power calculations: in Greenspan (2007) a sample of 1300 patients in each intervention group gave 90% power to detect a 60% or greater reduction in vertebral fracture incidence in the PTH (1-84) group compared with the control group. Given the SDs in Black (2003), there was a 90% power to detect a difference in the areal BMD of about 2.8% for the spine and 2.2% for the hip.

Both studies demonstrated baseline comparability of the groups. Patients in the TOP trial were comparable on age, years since menopause, mean BMI, prevalent vertebral fractures and mean BMD T-score. Black (2003) reported that the patients were comparable on age, age at menopause, height loss since 25 years old, mean BMI, clinical fractures since 45 years old, previous alendronic acid use, areal BMD at total hip, femoral neck and volumetric density on quantitative CT.

The TOP trial had more than 20% missing data in each intervention arm (Greenspan 2007): 36% (462/1286) of the patients in the PTH (1-84) group were lost to follow-up compared with 30% (369/1246) in the placebo group due to discontinuation of therapy treatment. Black (2003) had 5% (11/238) missing data overall, but did not provide further details for each intervention arm.

Both studies stated that they performed ITT analyses, so there were some imputed data; however, it was noted that there were missing data in the studies. Greenspan (2007) specifically stated that those lost to follow-up were assumed to have no fractures.

In the TOP trial, the method of vertebral fracture was defined radiographically by using a semiquantitative four-point grading scale (the Genant method) and an incident vertebral fracture was identified by a change in grade of one or more from baseline (Greenspan 2007). These details were not reported in Black (2003). Neither study reported the method used to assess hip fracture.

Greenspan (2007) reported nonvertebral fractures as adverse events and stated that they did not distinguish between traumatic and fragility nonvertebral fractures. Black (2003) only reported clinical fractures.

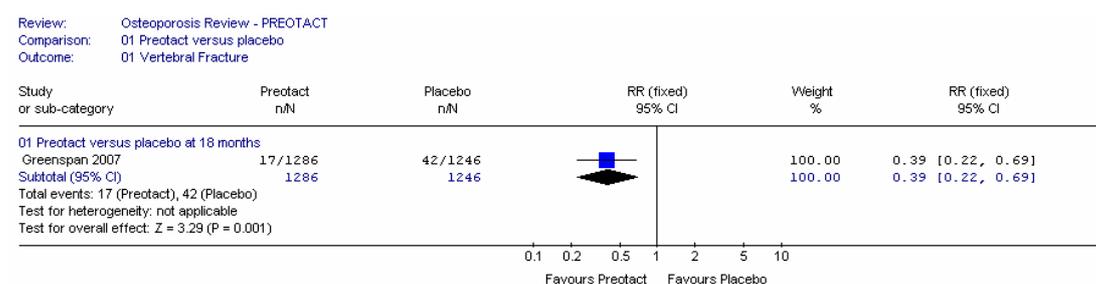
Overall, the study by Greenspan (2007) was considered to be at higher risk of bias because of the level of missing data, with some differences between groups (36% in the PTH [1-84] group compared with 30% in the placebo group), and the study did not distinguish between traumatic and fragility fractures. This study was treated with some caution.

## Results

Black (2003) reported that eight clinical fractures occurred during the trial and that the incidence was similar, approximately 3%, in all treatment groups. Given these limited data, we decided not to consider this study further and the results are focused on Greenspan (2007).

### Vertebral fractures

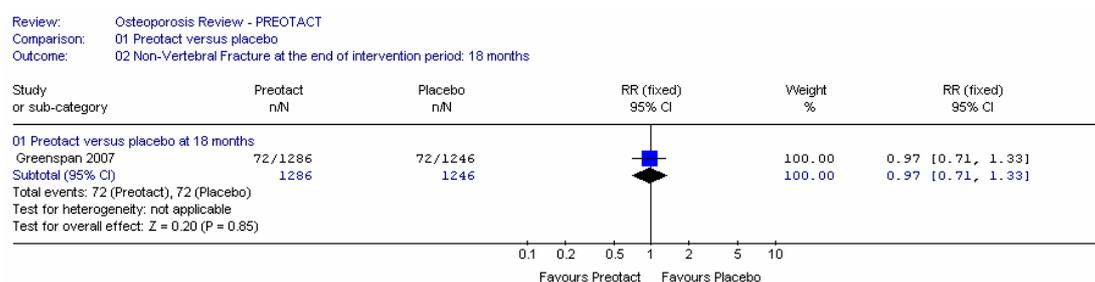
One study of 2532 patients recorded the number of patients with vertebral fractures at the end of the intervention period. There were significantly fewer patients with vertebral fractures in the PTH (1-84) group compared with the placebo group (figure 41); RR 0.39 (95% CI 0.22 to 0.69) at the end of the 18-month intervention period. This corresponds to a NT of 50 (95% CI 33 to 100) for a control group rate of 3%.



**Figure 41. Number of patients with vertebral fractures**

### Nonvertebral fractures

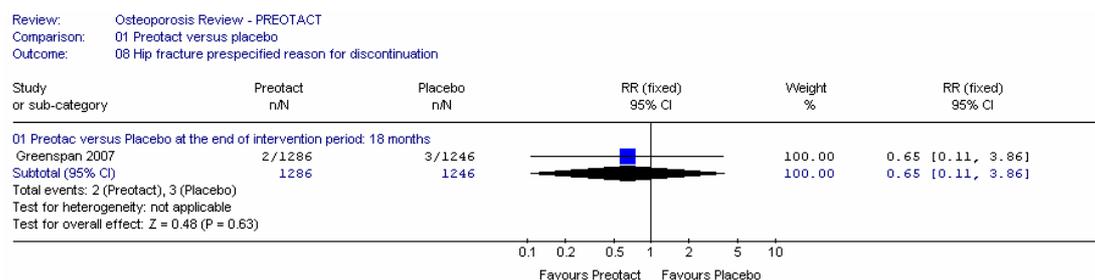
The number of patients with nonvertebral fractures at the end of the intervention period was reported under adverse events, but there was no distinction between fragility and traumatic nonvertebral fractures. One study in 2532 patients showed there was no significant difference between PTH (1-84) and placebo in the number of patients with nonvertebral fractures (figure 42); RR 0.97 (95% CI 0.71 to 1.33).



**Figure 42. Number of patients with nonvertebral fractures**

### Hip fractures as prespecified reason for discontinuation

The number of patients with hip fractures sufficiently bad to discontinue the study was also reported for the 18 months. An incident hip fracture was a specified reason for discontinuation in the trial and was reported as an adverse event. Traumatic and fragility nonvertebral fractures were not assessed distinguishably. Comparing PTH (1-84) with placebo, the CI was too wide to draw any conclusions (figure 43).



**Figure 43. Number of patients with a hip fracture leading to discontinuation**

### Concordance

In Black (2003), concordance was defined as the use of 80% of the treatment for at least the first 11 months. The average rate of concordance was assessed by means of the return of unused injection cartridges and tables.

Patients had 75% full adherence to injections (regardless of the content of the injections) and 81% full adherence to tablets. Concordance was not described in the TOP trial (Greenspan 2007).

### **2.3.3 Calcitonin**

#### ***Description of included studies***

A total of 31 studies were evaluated for inclusion. Four of these only reported BMD as an outcome and were excluded from the review (Adami 1995; Ellerington 1996; Kaskani 2005; Tekeoğlu 2005). Two more studies were excluded because the studies were in patients with comorbidities other than those treated with glucocorticoids (Välimäki 1999a; 1999b). Twenty other studies were excluded for reasons given in appendix IV. Five reports were included (Chesnut 2000 and Stock 1997 [abstract] [PROOF study]; Luengo 1994; Overgaard 1992; Tóth 2005), describing four trials.

#### ***Study design***

Of the four included RCTs, none were conducted in the UK, one was a multinational trial (Chesnut 2000), one was conducted in Denmark (Overgaard 1992), one in Spain (Luengo 1994) and one in Hungary (Tóth 2005).

One study (Chesnut 2000) was industry sponsored (Novartis Pharmaceuticals) and one received government funding: Overgaard (1992) was supported by a grant from the Danish Medical Research Council. Two studies did not state their source of funding (Luengo 1994; Tóth 2005).

One of the studies randomised at least 311 patients into each of the treatment arms (Chesnut 2000). The other studies were relatively small; one randomised 52 participants into each treatment arm (Overgaard 1992), one randomised 22 to each arm (Luengo 1994) and the other randomised 31 and 40 into each group (Tóth 2005).

Two of the included studies did not state whether the trials were conducted in primary or secondary settings (Chesnut 2000; Overgaard 1992). The study by Tóth (2005) was reported to have taken place in secondary care settings and the study by Luengo (1994) recruited outpatients.

## **Population**

Two studies (Chesnut 2000; Overgaard 1992) evaluated postmenopausal women, with a mean time since the menopause of 22–23 years. Luengo (1994) included patients taking oral glucocorticoids for asthma. Tóth (2005) evaluated men with idiopathic osteoporosis.

Average femoral neck T-scores could be calculated from data presented in one study (Tóth 2005). In Tóth (2005), the femoral neck BMD score was 0.777 g/cm<sup>2</sup> in the treatment group and 0.797 g/cm<sup>2</sup> in control group, suggesting that some of the men had osteopenia as well as osteoporosis. The study by Chesnut (2000) included participants with a lumbar spine BMD at least 2 SD below normal for women aged 30 years. In this study, the mean lumbar BMD was 0.85 (SD 0.12), suggesting that the participants had either osteoporosis or osteopenia. In Overgaard (1992), the mean bone mineral content of the lumbar spine ranged from 34.6 g (SD 7.5) to 36.2 g (SD 7.1). The study by Luengo (1994) did not report BMD values at baseline.

In Chesnut (2000), 71–75% of the women had one to five prevalent vertebral fractures at baseline, whereas in the study by Overgaard (1992) the proportion of women with a vertebral fracture at baseline ranged from 4 to 14%. Luengo (1994) had 18% prior fractures in each arm. Tóth (2005) specifically included men without vertebral fractures in their study.

The mean age of participants across studies ranged from 68 to 72 years for the postmenopausal women, the age range of the men in Tóth (2005) was 40–76 years and in Luengo (1994) the mean age of men was 59 years.

In the women, the average BMI was 25 kg/m<sup>2</sup> in Chesnut (2000); Overgaard (1992) did not report BMI, but did report that the participants were on average 157–160 cm tall, with a mean weight of 61 or 63 kg. The men in Tóth (2005) had a mean height of 169 and 170 cm in the control and treatment groups (respectively); the average weight was 75 kg in both groups (BMI was not reported).

None of the trials reported information on ethnicity, although Chesnut (2000) stated that white, Asian, or Hispanic women were eligible to participate in the trial.

Two studies reported whether participants consumed alcohol (Luengo 1994; Tóth 2005): Luengo (1994) stated that no patient consumed more than 20 g/day and Tóth (2005) did not provide any data. No participants smoked in Luengo (1994). Chesnut (2000) and Tóth (2005) reported that some of the participants smoked (although Tóth 2005 did not provide any data).

### ***Interventions***

The comparisons included the following.

- 200 IU/day calcitonin nasal spray versus placebo
  - for 2 years (Luengo 1994; Overgaard 1992)
  - for 5 years (Chesnut 2000).
- 200 IU/day calcitonin nasal spray versus no treatment for 2 years (Luengo 1994).
- 200 IU/day calcitonin nasal spray for alternate months versus no treatment for 18 months (Tóth 2005).

In Overgaard (1992), all of the participants received 500 mg/day calcium and those in Luengo (1994) received 1 g/day calcium. In Chesnut (2000) and Tóth (2005) all of the participants received 1 g/day calcium and 400 IU/day vitamin D.

The following additional doses of calcitonin were also examined, but have not been evaluated in this report.

- 50 IU/day calcitonin nasal spray (Overgaard 1992).
- 100 IU/day calcitonin nasal spray (Chesnut 2000; Overgaard 1992).
- 400 IU/day calcitonin nasal spray (Chesnut 2000).

### ***Methodological quality***

The method of sequence generation was adequate in one study (Chesnut 2000), which reported a computer-generated sequence was used; it was

partially adequate in another study (Overgaard 1992; the authors used random sampling numbers in blocks of four); and it was not clearly reported in the other studies. No studies reported allocation concealment methods. Two studies were reported to have been double blind (Chesnut 2000; Overgaard 1992). Tóth (2005) was an open-label study, and Luengo (1994) reported blinding of the outcome assessor, but the patients were not blinded.

One trial calculated an a priori sample size (Overgaard 1992). In this study, the authors estimated that 52 patients per group would be required to detect a difference of 2%, assuming a precision of 3% of the bone mineral content of the lumbar spine. This exact sample size was used in this study. Tóth (2005) calculated the power of the study based on their recruited sample size. They estimated that the study would have 80% power to detect a significant difference of at least 2% between the groups (based on the assumption of a SD of 3% for the changes in the spine BMD).

All studies reported that their groups were comparable at baseline for variables such as age, weight, height, years since menopause, bone mineral content, percentage of vertebral fractures, and biochemical markers.

Chesnut (2000) reported a very high level of missing data, with 58% missing in the 200 IU calcitonin group and 59% missing in the placebo group at 5 years. The authors analysed all participants with at least one follow-up radiograph, so that 89% of participants were included in the final analyses (no other details regarding ITT methods were reported). Overgaard (1992) reported that 13% of participants withdrew from the treatment group, and 15% withdrew from the placebo group, however 78% were included in the final analysis of these groups (available cases were included in the analysis, with no other details reported). Tóth (2005) reported no missing data, and all participants randomised were analysed. Luengo (1994) reported 23% missing data in each group.

Two studies reported methods to assess concordance involved counting used and unused bottles of study medication (Chesnut 2000) at each clinic visit and at the end of the study (Overgaard 1992).

In all studies, vertebral fractures were defined as a 20% or greater decrease in any vertebral height. Chesnut (2000) and Overgaard (1992) also evaluated nonvertebral fractures evaluated using hospital/medical records and radiographs. Two of the trials evaluated fracture as a primary outcome (Chesnut 2000; Overgaard 1992), Tóth (2005) evaluated fracture as a secondary outcome, and it was likely to be a secondary outcome in Luengo (1994). Overgaard (1992) included nontraumatic fractures and fractures caused by minor trauma in their analyses. Tóth (2005) only included low trauma fractures, and vertebral compression fractures in their analyses. Chesnut (2000) did not explicitly state what fractures were considered in their analysis, but it appears that all types of fractures were included (for example, vertebral and hip or femoral, humerus, radius, ulna or wrist fractures). Luengo (1994) did not describe the type of fractures.

Overall, the study by Chesnut (2000) was considered to be at high risk of bias because of the high proportion of missing data (59%), although 89% had at least one radiograph for vertebral fractures. The studies by Tóth (2005) and Luengo (1994) were also at higher risk of bias because the patients were not blinded. All studies were treated with caution and sensitivity analyses were carried out, when possible.

## ***Results***

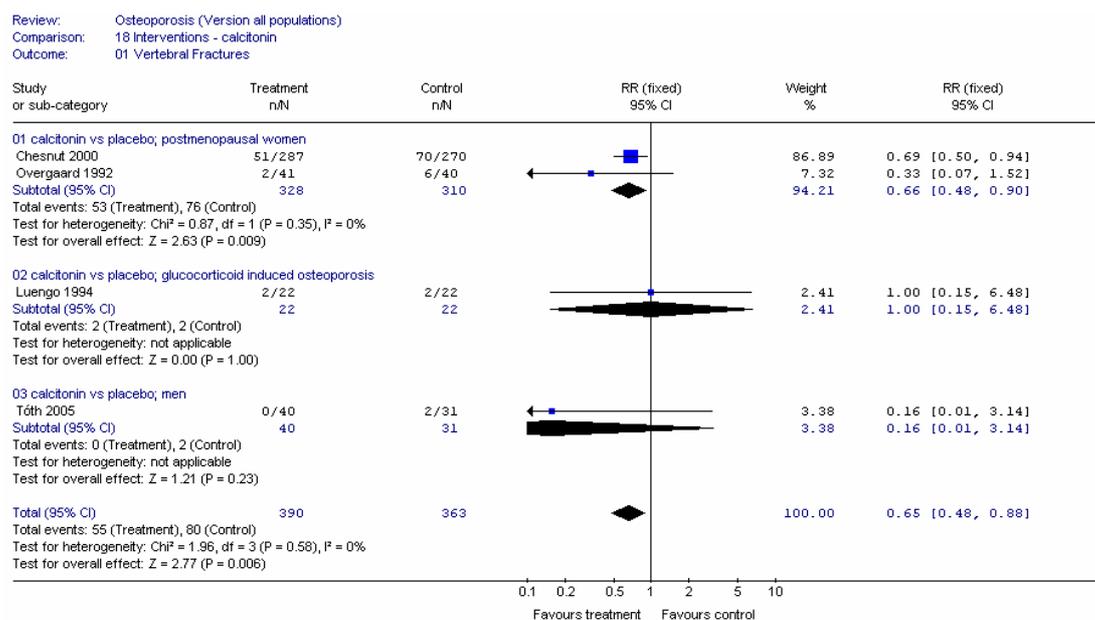
### **Calcitonin nasal spray versus placebo/no treatment**

Four studies compared 200 IU calcitonin nasal spray versus placebo or no treatment: Chesnut (2000) evaluated 627 patients for 5 years; Overgaard (1992) evaluated 104 patients for 2 years; and Luengo (1994) evaluated 44 patients for 2 years. Tóth (2005) evaluated 200 IU/day calcitonin nasal spray for alternate months versus no treatment in 71 patients for 18 months. It was unclear how many patients were included in the denominator for the study by Luengo (1994), so the randomised participants were assumed.

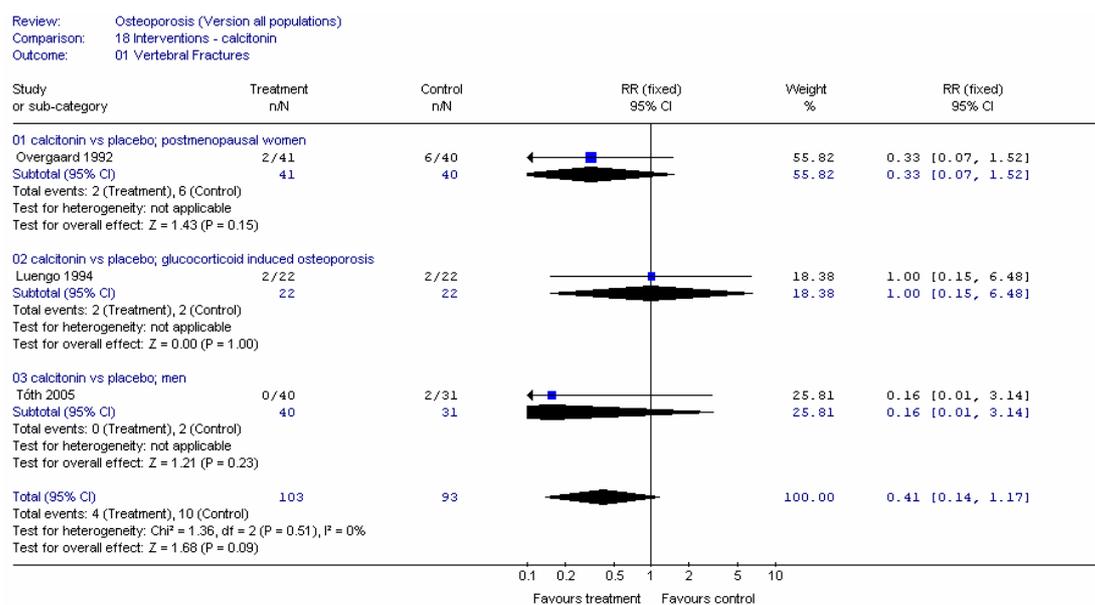
#### *Vertebral fractures*

Four studies reported the number of women with vertebral fractures. A meta-analysis of 753 women (figure 44) showed that significantly fewer had a vertebral fracture when taking 200 IU calcitonin nasal spray in comparison to

placebo or no treatment (RR 0.65; 95% CI 0.48 to 0.88); this corresponds to a NNT of 13 (95% CI 8 to 50) for a control group rate of 7–26%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.58$ ). The study by Chesnut (2000) had potential for bias because of high levels of missing data. In the absence of this study the CIs were too wide to determine if there was a difference between interventions (figure 45).



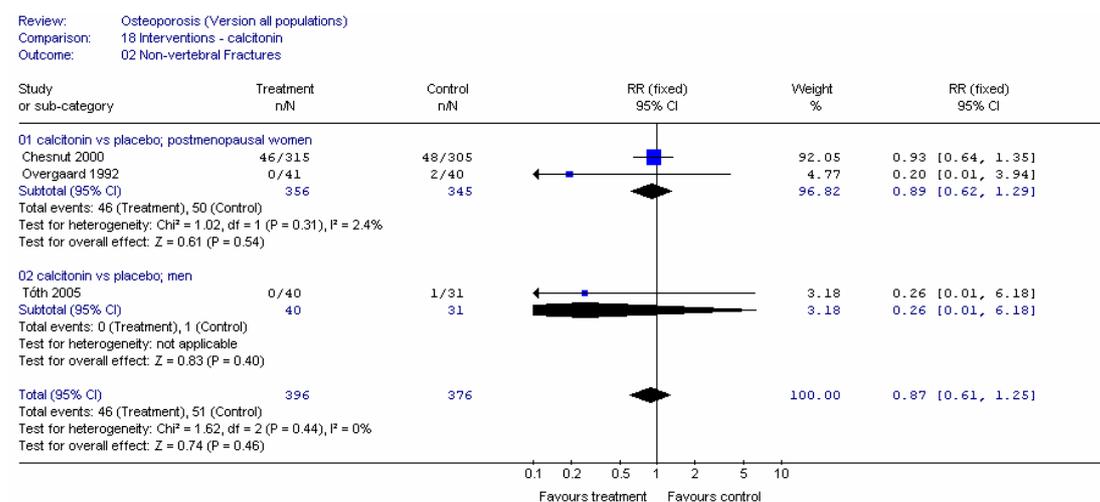
**Figure 44. Number of patients with vertebral fractures**



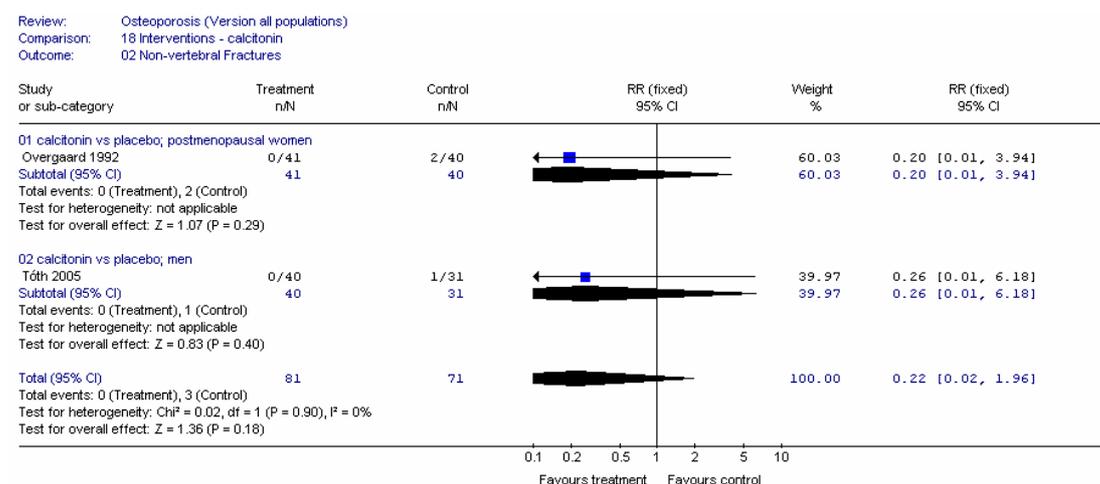
**Figure 45. Number of patients with vertebral fractures: sensitivity analysis without Chesnut (2000)**

## Nonvertebral fractures

Three studies evaluated the number of nonvertebral fractures (Chesnut 2000; Overgaard 2001; Tóth 2005) in 772 patients. There was no significant difference between interventions (figure 46); RR 0.87 (95% CI 0.61 to 1.25), and no heterogeneity was observed ( $I^2 = 0\%$ ;  $p = 0.44$ ). The study by Chesnut (2000) had potential for bias because of high levels of missing data. In the absence of this study the CIs were too wide to determine if there was a difference between interventions (figure 47).



**Figure 46. Number of patients with nonvertebral fractures**

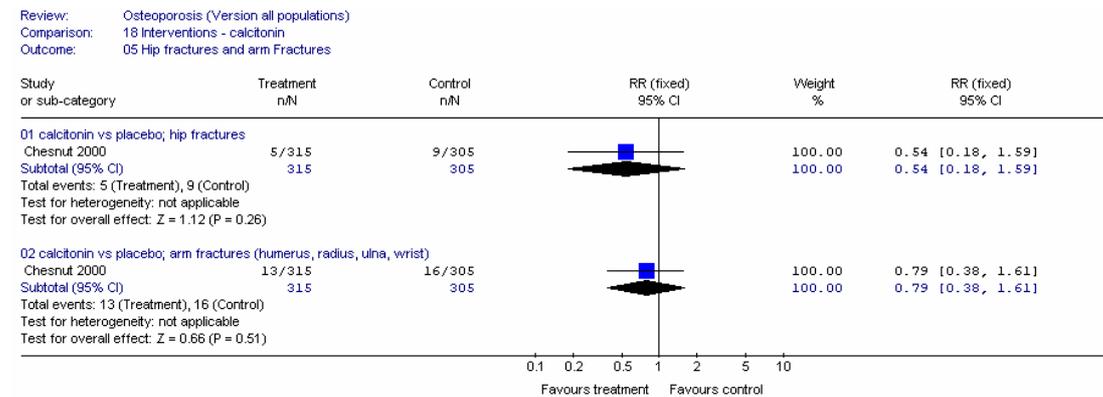


**Figure 47. Number of patients with nonvertebral fractures: sensitivity analysis**

## Hip fractures

One study of 620 women (Chesnut 2000) reported the number of hip fractures. There was no significant difference between interventions (figure

48), although the CI was fairly wide; RR 0.54 (95% CI 0.18 to 1.59). The study by Chesnut (2000) had high potential for bias because of high levels of missing data. However, because this was the only study reporting on hip fracture, we have included the data for completeness.



**Figure 48. Number of patients with a hip fracture or with an arm fracture**

### *Arm fractures (humerus, radius, ulna and wrist)*

One study of 620 women (Chesnut 2000) reported the number with arm fractures. There was no significant difference between interventions (figure 48), although the CI was fairly wide; RR 0.79 (95% CI 0.38 to 1.61). The study by Chesnut (2000) had potential for bias because of high levels of missing data.

## **2.4 Sex hormone therapies**

### **2.4.1 Hormone replacement therapy (oestrogen with or without progestogen)**

#### ***Description of studies***

A total of 51 papers were evaluated for inclusion. Ten of these only reported BMD as an outcome and were excluded from the review (Al-Azzawi 2005; Cheng 2000; Christiansen 2005; Eiken 1997; Greenwald 2005; Liu 2005; Nielsen 2004; Popp 2006; Warming 2004; Warming 2005). Eleven other studies were excluded from this review because they used unlicensed HRT regimens not listed in the BNF (Aitken 1973, Alexandersen 1999; Cauley 2001; Cauley 2003 [Women's Health Initiative trial; subset of women without hysterectomy]; Gallagher 2001 (59% of patients (those with an intact uterus)

had an unlicensed regimen); Greenspan 2003 (65% of patients (those with an intact uterus) had an unlicensed regimen); Ishida 2001; 2004; Komulainen 1998; Nachtigall 1979; Recker 1999). It is noted that Gallagher (2001) and Greenspan (2003) are included studies for other reviews in this document. Nine studies were excluded because they were in patients with other conditions, including primary hyperparathyroidism, and the GDG decided these would be unrepresentative. Eleven further studies were excluded. All excluded studies are described in appendix IV with reasons for exclusion.

Eleven reports were included (Anderson 2004 [Women's Health Initiative – WHI trial; subset of women with hysterectomy]; Bagger 2004; Bone 2000; Bush 1996 [Postmenopausal Estrogen/Progestin Interventions – PEPI]; Herrington 2000; Lufkin 1992; Mosekilde 2000 [Danish Osteoporosis Prevention Study]; Pacifici 1988; Ravn 1999 [Early Postmenopausal Intervention Cohort]; Wimalawansa 1998; Weiss 1999).

Anderson (2004) was the oestrogen-only part of the WHI trial, which was undertaken to determine if HRT would reduce cardiovascular events in mostly healthy postmenopausal women. The other arm using oestrogen with progestin was stopped early because of early increases in rates of coronary heart disease, stroke and venous thromboembolic disease in the HRT arm, and the oestrogen only arm was stopped 2 years later. Anderson (2004) reported the results up to the time of termination of the trial.

Bagger (2004) was a follow-up study of four earlier trials; because of missing data, only 55% of patients originally randomised entered the follow-up study. We judged this to be unrepresentative and the study was not considered further.

### ***Study design***

Of the remaining studies, one was conducted in the UK (Wimalawansa 1998); one in Denmark; seven in the USA; and one internationally (Ravn 1999).

Four studies received some funding from industry (Bone 2000; Lufkin 1992; Ravn 1999; Weiss 1999); two studies (Pacifici 1988; Wimalawansa 1998) did

not state their source of funding. The remaining studies were funded by nonindustry sources.

Three studies had fewer than 100 patients (Wimalawansa 1998: n = 72; Lufkin 1992: n = 75; Pacifici 1988: n = 93). One study had fewer than 20 patients in the intervention arm (Wimalawansa 1998: n = 18). Four studies (Bone 2000; Bush 1996; Herrington 2000; Ravn 1999) had between 100 and 200 patients in the intervention arm; Mosekilde (2000) had more than 500 and Anderson (2004) randomised more than 5000 women to HRT.

The majority of studies took place in secondary care (six); two (Anderson 2004; Mosekilde 2000) were in primary care; and two (Bush 1996; Ravn 1999) did not report the setting.

### ***Population***

All studies except Pacifici (1988) reported only on postmenopausal women. Pacifici (1988) included women from 26 to 80 years. All the women in three studies (Anderson 2004; Bone 2000; Weiss 1999) had had a hysterectomy; four studies reported that some of the patients had had a hysterectomy (Bush 1996; Herrington 2000; Lufkin 1992; Mosekilde 2000); one study appeared to have no patients who had had a hysterectomy (Ravn 1999); and Pacifici (1998) did not say. Two studies (Anderson 2004; Herrington 2000) included women with a surgical menopause. The time since the menopause varied from less than a year (Mosekilde 2000; Ravn 1999; Weiss 1999) to around 15 years (Lufkin 1992; Wimalawansa 1998).

Mean T-scores were reported (or calculated), and on this basis, one study (Pacifici 1988) was stated to be in patients with osteoporosis only. Three studies had patients with either osteoporosis or osteopenia (Bone 2000; Lufkin 1992; Wimalawansa 1998). Three studies had patients with either osteopenia or normal BMD (Bush 1996; Mosekilde 2000; Ravn 1999); two did not record BMD and patients were not selected for low BMD (Anderson 2004; Herrington 2000); and one study was unclear on BMD (Weiss 1999).

Five studies reported the proportions of patients who had had a prior fracture. This ranged from intermediate percentages (Anderson 2004: 14.5%) to 100%

(Lufkin 1992; Pacifici 1988; Wimalawansa 1998). The number of patients with a prior fracture was not reported in the remaining studies.

The age range of patients across the studies was 26 (Pacifici 1988) to 80 years, with the mean age, when given, ranging from 51 to 66 years. Two studies (Lufkin 1992; Pacifici 1988) reported that all the patients were white; Bone (2000) stated that 87–92% of their population were white; 83–91% and 75% of the patients were white in Weiss (1999) and Anderson (2004), respectively, and the rest did not report ethnicity.

### ***Interventions***

The following interventions were included.

- Conjugated oestrogen with progestogen:
  - oral HRT:
    - ◇ conjugated equine oestrogen 0.625 mg/day plus medroxyprogesterone acetate (MPA) 5 mg/day (Ravn 1999)
    - ◇ oestrogen 0.625 mg/day plus norgestrel 150 micrograms for 12 days each month (Wimalawansa 1998).
- Oestradiol with progestogen:
  - oral HRT:
    - ◇ oestradiol 2 mg/day for 12 days; oestradiol 2 mg/day plus norethisterone acetate (NETA) 1 mg/day for 10 days; oestradiol 1 mg/day for 6 days (part of Mosekilde 2000; part of Ravn 1999)
  - HRT patch:
    - ◇ 17 $\beta$  oestradiol (0.1 mg/day for days 1–21 of 28-day cycle) plus oral MPA (10 mg/day for days 11–21 of 28-day cycle) (Lufkin 1992).
- Oestradiol only:
  - oral HRT:
    - ◇ continuous oestradiol (2 mg/day) if hysterectomy (part of Mosekilde 2000)
  - transdermal patch:
    - ◇ transdermal weekly patch: 17 $\beta$  oestradiol 0.1 mg/day (Weiss 1999).
- Conjugated oestrogens only:

- oral HRT:
  - ◇ conjugated equine oestrogen (0.625 mg/day) (Anderson 2004; Bone 2000; Herrington 2000).

The Bush (1996) (PEPI) trial examined several HRT regimens (some licensed and others not) but the fracture results were only given for all the active treatment groups together, so licensed regimens could not be separated out. Therefore, the study was excluded from the analysis. Herrington (2000) randomised patients to a third arm, but this regimen (conjugated oestrogen 0.625 mg plus medroxyprogesterone acetate 2.5 mg) was not licensed for use in the UK.

### **Comparisons**

The following comparisons were carried out.

HRT versus placebo or no treatment

- Conjugated oestrogen with progestogen:
  - conjugated equine oestrogen 0.625 mg/day plus MPA 5 mg/day versus placebo (Mosekilde 2000; Ravn 1999):
    - ◇ Ravn (1999) gave different HRT regimens depending on the country concerned in this multicentre trial (this regimen was in the USA). Results were given for the HRT arm as a whole; duration was 4 years.
    - ◇ Mosekilde (2000) gave some patients this regimen and some continuous oestradiol; duration was 5 years.
  - Oestrogen 0.625 mg/day plus norgestrel 150 micrograms for 12 days each month versus no treatment for 4 years (Wimalawansa 1998).
- Oestradiol with progestogen:
  - oestradiol 2 mg/day for 12 days; oestradiol 2 mg/day plus NETA plus 1 mg/day for 10 days; oestradiol 1 mg/day for 6 days versus placebo for 4 years (Ravn 1999):
    - ◇ Ravn (1999) gave different HRT regimens depending on the country concerned in this multicentre trial (this regimen was in Europe). Results were given for the HRT arm as a whole.

- Oestradiol patches:
  - 17 $\beta$  oestradiol (0.1 mg/day for days 1–21 of 28-day cycle) plus oral MPA (10 mg/day for days 11–21 of 28-day cycle) versus placebo for 1 year (Lufkin 1992).
- Conjugated oestrogens only:
  - conjugated equine oestrogen (0.625 mg/day) (Anderson 2004; Bone 2000; Herrington 2000):
    - ◊ mean duration 2 years (Bone 2000)
    - ◊ mean duration 3.2 years (Herrington 2000)
    - ◊ mean duration 6.8 years (Anderson 2004).
- Oestradiol only:
  - continuous oestradiol (2 mg/day) for 5 years (Mosekilde 2000):
    - ◊ Mosekilde (2000) gave some patients this regimen and some conjugated oestrogens with progesterone.

#### Oral HRT plus other intervention versus other intervention alone

- Conjugated oestrogen with progestogen versus no treatment, with etidronate in both arms:
  - oestrogen 0.625 mg/day plus norgestrel 150 micrograms for 12 days each month plus etidronate 400 mg for 14 days every 12 weeks versus etidronate alone for 4 years (Wimalawansa 1998).
- Conjugated oestrogens versus placebo, with alendronate in both arms:
  - conjugated equine oestrogen (0.625 mg/day) plus alendronic acid 10 mg/day versus placebo plus alendronic acid 10 mg/day alone for 2 years (Bone 2000).

Anderson (2000) also randomised patients to a second HRT arm, but this was not included because the dose was not licensed for use in the UK.

#### ***Methodological quality***

The quality assessment for included studies is shown in appendix V.

The method of sequence generation was adequate (computer generated) in four of the studies (Anderson 2004; Bone 2000; Weiss 1999; Wimalawansa

1998). Herrington (2000) reported a partially adequate method (permuted random blocks). The remaining studies did not state the method of randomisation. Allocation concealment was partially reported in Weiss (1999), in which a central randomisation process was used, but there were no details given.

The majority of studies reported that the patients were blinded to the interventions. The exceptions were Ravn (1999), in which the patients were said to be randomised in a double-blind manner, but the HRT was open label; Wimalawansa (1998), in which patients could not be blinded; Mosekilde (2000), which did not blind the patients; and Pacifici (1988), which did not specify.

Three studies reported a sample size calculation: Anderson (2004) reported that 123,754 women would be required for a power of 81% to detect a reduction of 21% in rates of coronary heart disease over 9 years and that this sample size would give 65% power to detect a 20% reduction in hip fracture rates. The actual sample size achieved had 55% power for hip fracture. Herrington (2000) was designed to have an 80% power to detect a difference of 0.054 mm in the degree of change in the minimal luminal coronary artery diameter between active treatment and placebo groups. Weiss (2000) calculated a sample size of 176 patients based on detecting a difference of 4% in the change in BMD, corresponding to  $\alpha = 0.0022$  and a power of 60%. The remaining studies gave no details of an a priori sample size calculation.

All the studies included in the review demonstrated baseline comparability of the groups.

Four studies had missing data of more than 20% (Bone 2000; Pacifici 1988; Weiss 1999; Wimalawansa 1998). Pacifici (1988) had 61% missing data overall and this study was considered too biased to be included in the analyses. Weiss (1999) had 45% missing data overall; Bone (2000) had the following missing data: 21% HRT plus alendronate; 24% HRT; 26% alendronate; 32% placebo, that is, a differential dropout rate between HRT

and placebo. Wimalawansa (1998) reported missing data in each group: 3/18 (17%) HRT; 3/17 (18%) etidronate; 4/19 (21%) HRT plus etidronate; 4/18 (22%) placebo. One study had no missing data (Ravn 1999) and the rest had missing data of less than 20%.

Four studies used ITT analyses: in Ravn (1999) all the patients were followed, and the other three studies (Anderson 2004; Bone 2000; Weiss 1999) reported results for all the patients despite missing data; three used available case analysis (Lufkin 1992; Mosekilde 2000; Pacifici 1988).

None of the studies reported fracture as a primary outcome, although it was an important outcome in Mosekilde (2000); four studies reported fracture as a secondary outcome (Anderson 2004; Bush 1996; Lufkin 1992; Wimalawansa 1998) and the other studies recorded fractures as adverse events.

Of the seven studies reporting vertebral fractures, only one (Mosekilde 2000) reported using a 20–25% cut-off height reduction method, three (Lufkin 1992; Pacifici 1988; Wimalawansa 1998) used a 15% cut-off point in height reduction, and the rest did not state how vertebral fractures were measured.

Anderson (2004) reported the number of osteoporotic fractures separately. Mosekilde (2000) reported fragility fractures for all fractures and stated that the proportion (given) of high-energy fractures was the same at any skeletal site; we therefore calculated the number of fragility fractures from this information. Weiss (1999) reported that there were no nontraumatic fractures in any group; Bone (2000) reported fracture outcomes but most fractures occurred as a result of trauma and this study was therefore not included in the analyses of nontraumatic/osteoporotic fractures. The other studies did not describe whether fractures included trauma or not.

Overall, these studies had poor methodological quality for fracture outcomes. We considered that Bone (2000), Pacifici (1988) and Weiss (1999) had too high a potential for bias and these studies were not considered in the analyses. Of the remainder, Mosekilde (2000) and Wimalawansa (1998) had potential for bias: Mosekilde (2000) because the patients were not blinded and Wimalawansa (1998) because of some missing data and the lack of patient

blinding. Two of the studies had some potential for bias because of the use of a 15% reduction in height definition (Lufkin 1992; Wimalawansa 1998). We considered Mosekilde (2000) and Wimalawansa (1998) with caution and examined them in sensitivity analyses.

## **Results**

For this review, it was decided a priori to combine studies across all types of HRT, but to use subgroups by class of HRT. Within subgroups, studies with all modes of delivery of interventions were also combined. When there was heterogeneity, the following additional a priori subgroup analyses were considered: type of HRT, route of delivery (oral, transdermal patch, implant), whether or not patients had an intact uterus

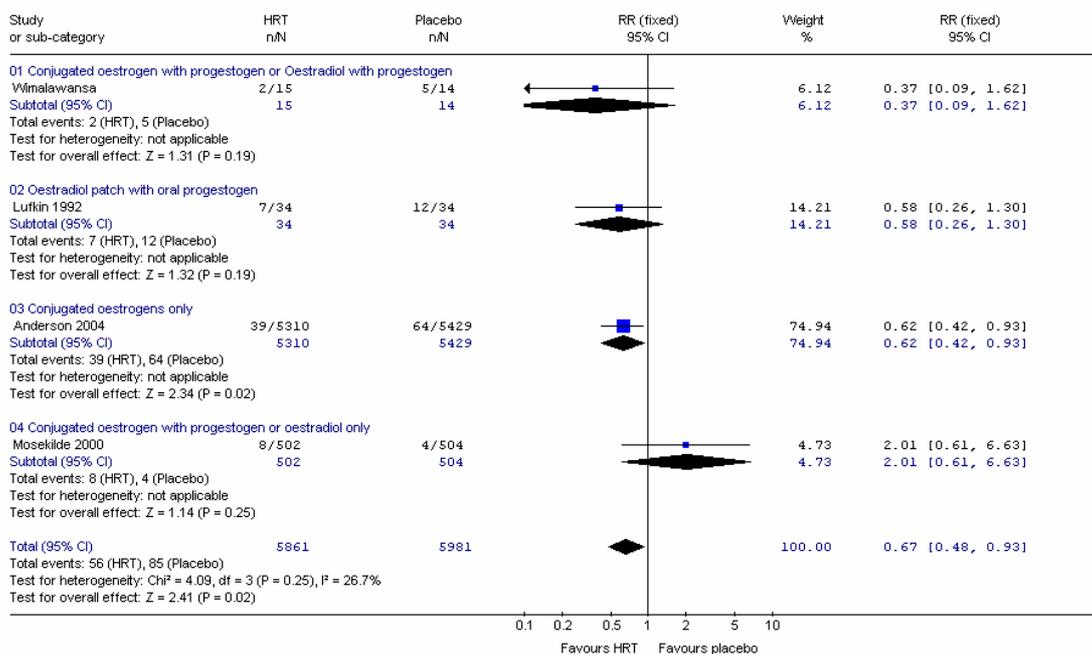
### **HRT versus placebo or no treatment**

Five studies compared HRT versus placebo (Anderson 2004; Herrington 2000; Lufkin 1992; Mosekilde 2000; Ravn 1999) and one study HRT versus no treatment (Wimalawansa 1998).

#### *Vertebral fractures*

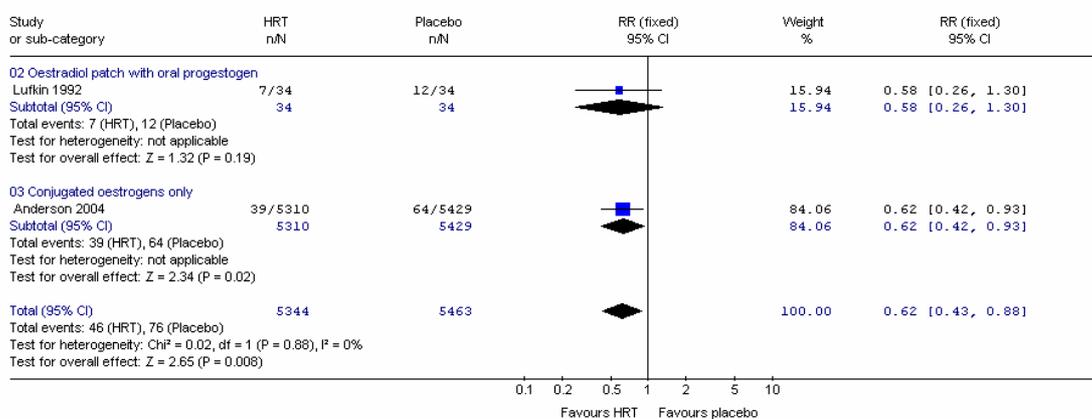
A meta-analysis of four studies in 11,842 patients showed there were significantly fewer patients with vertebral fractures in the HRT group compared with placebo (figure 49); RR 0.67 (95% CI 0.48 to 0.93). This corresponds to a number needed to treat of  $\infty$ . There was no significant heterogeneity ( $I^2 = 27\%$ ;  $p = 0.25$ ). Sensitivity analysis (figure 50) in the absence of Mosekilde (2000) and Wimalawansa (1998) gave a RR of 0.62 (95% CI 0.42 to 0.93), which corresponds to a NNT of 100 (95% CI 100 to  $\infty$ ); there was no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.88$ ). It was decided to take this result in preference.

Review: Osteoporosis (Version all populations)  
 Comparison: 19 HRT versus placebo  
 Outcome: 03 Vertebral fracture



**Figure 49. Number of patients with a vertebral fracture**

Review: Osteoporosis (Version all populations)  
 Comparison: 19 HRT versus placebo  
 Outcome: 07 Vertebral fracture sensitivity analysis

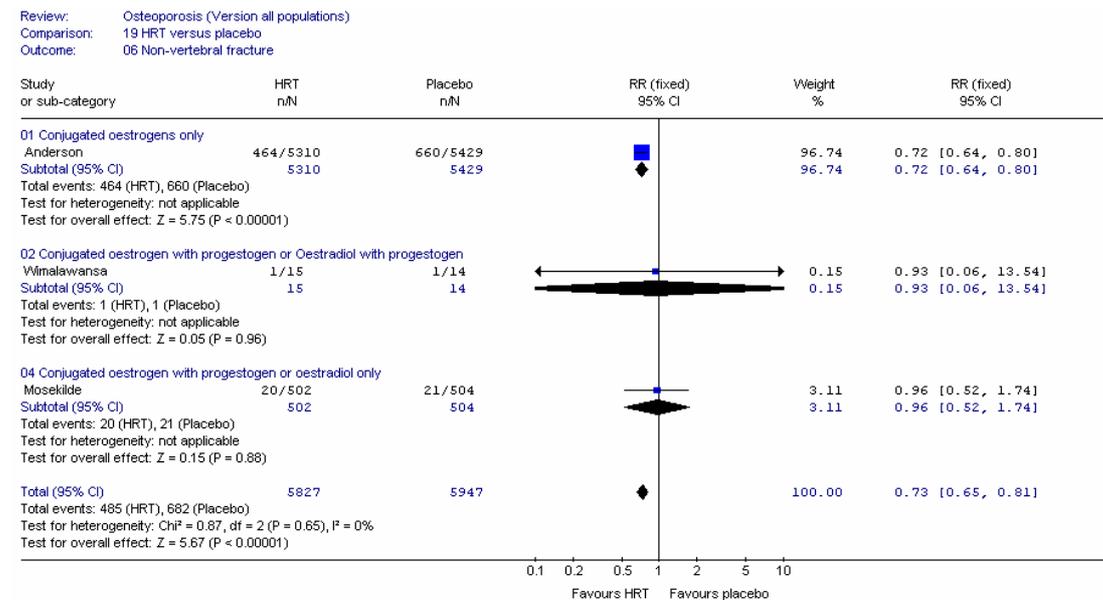


**Figure 50. Number of patients with a vertebral fracture: sensitivity analysis without Mosekilde (2000) and Wimalawansa (1998)**

### Nonvertebral fractures

Three studies (Anderson 2004; Mosekilde 2000; Wimalawansa 1998) reported the number of nonvertebral fractures. Anderson (2004) gave the total number of osteoporotic fractures and the number of vertebral fractures and we subtracted one from the other to obtain the number of nonvertebral fractures. Mosekilde (2000) gave the proportion of fragility fractures and we used this to calculate the number of fragility fractures. A meta-analysis of three studies in

11,774 patients showed there were significantly fewer patients with nonvertebral fractures in the HRT group compared with placebo (figure 51); RR 0.73 (95% CI 0.65 to 0.81). This corresponds to a NNT of 33 (95% CI 25 to 50) for a control group rate of 4–12%. There was no heterogeneity ( $I^2 = 0\%$ ,  $p = 0.65$ ). Sensitivity analysis in the absence of Mosekilde (2000) and Wimalawansa (1998) made little difference to the summary statistics.

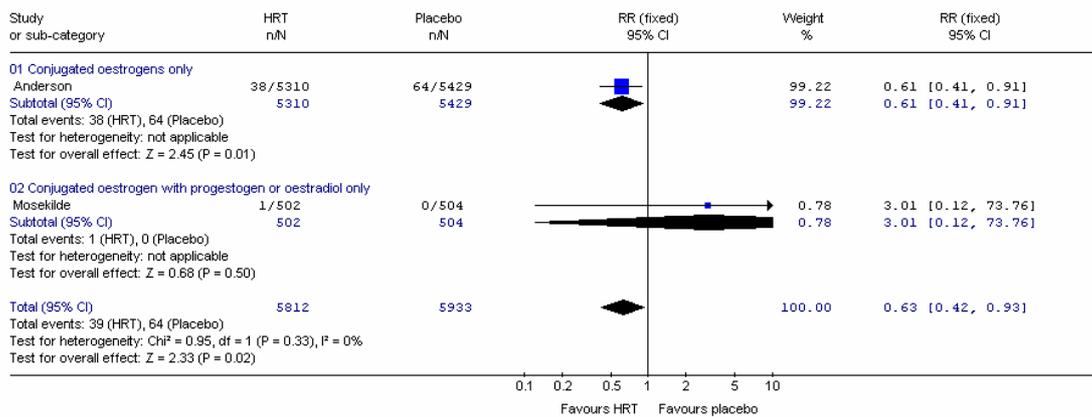


**Figure 51. Number of patients with a nonvertebral fracture**

### Hip fractures

A meta-analysis of two studies (Anderson 2004; Mosekilde 2000) in 11,745 patients showed there were significantly fewer patients with hip fractures in the HRT group compared with placebo (figure 52); RR 0.63 (95% CI 0.42 to 0.93). This gave a NNT of  $\infty$  (which was calculated to be 216 for Anderson 2004); there was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.33$ ). In the absence of Mosekilde (2000), there was little change in the summary statistics.

Review: Osteoporosis (Version all populations)  
 Comparison: 19 HRT versus placebo  
 Outcome: 04 Hip fracture

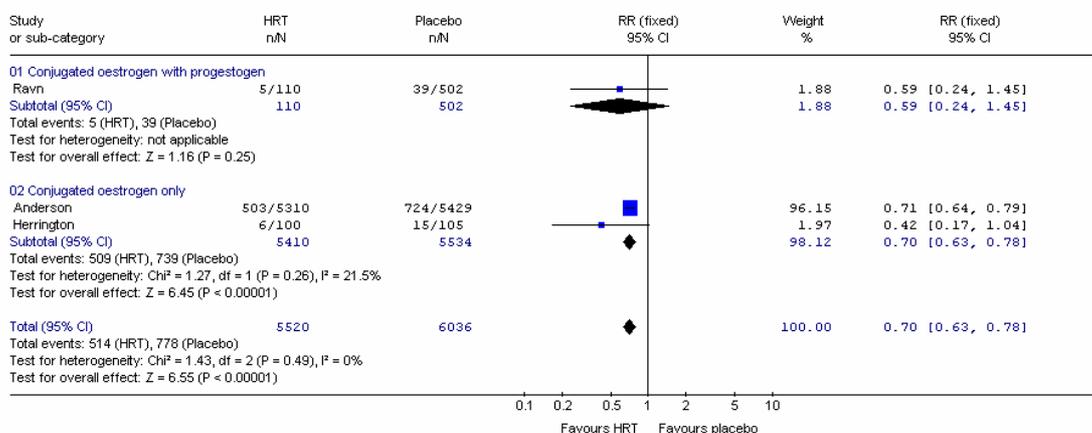


**Figure 52. Number of patients with a hip fracture**

*All fractures*

Three studies (Anderson 2004; Herrington 2000; Ravn 1999) reported the number of patients with any fracture; in two of these, the fractures were osteoporotic fractures, but Herrington (2000) only reported fractures as adverse events. A meta-analysis of the three studies in 11,556 patients (figure 53) showed there were significantly fewer patients having a fracture in the HRT group than in the placebo group; RR 0.70 (95% CI 0.63 to 0.78). This corresponded to a NNT of 25 (95% CI 20 to 34). There was no significant heterogeneity between studies ( $I^2 = 0\%$ ,  $p = 0.49$ ). Ravn (1999) was considered with caution due to lack of blinding, and Herrington (2000) may have included trauma fractures, but exclusion of these studies in a sensitivity analysis did not change the findings significantly.

Review: Osteoporosis (Version all populations)  
 Comparison: 19 HRT versus placebo  
 Outcome: 01 All fractures



**Figure 53. Number of patients with any fracture**

## Oral HRT plus other intervention versus other intervention alone

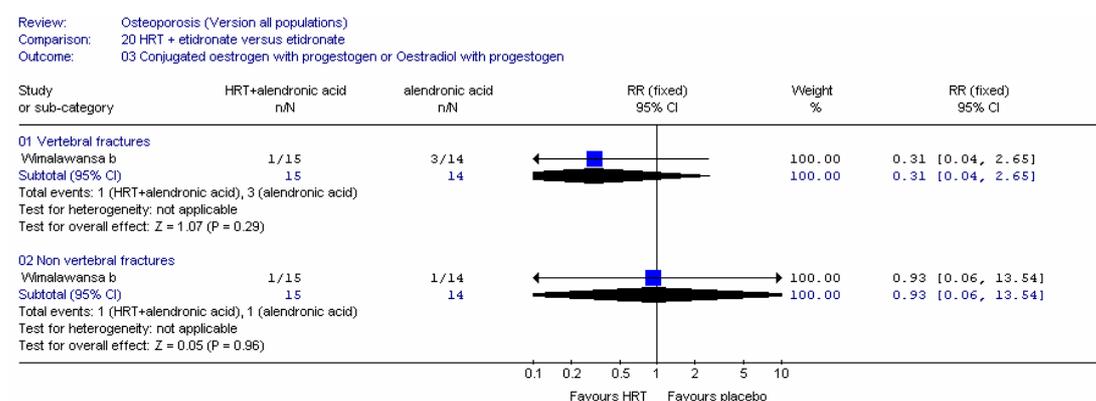
One study in 29 patients assessed oral HRT plus another intervention versus the other intervention alone (conjugated oestrogen with progestogen plus etidronate versus etidronate alone; Wimalawansa 1998).

### Vertebral fractures

The CI was too wide to determine if there was a difference between interventions (figure 54).

### Nonvertebral fractures

The CI was too wide to determine if there was a difference between interventions (figure 54).



**Figure 54. Number of patients with a vertebral fracture and number of patients with a nonvertebral fracture**

## 2.4.2 Selective estrogen (oestrogen) receptor modulators (SERMs) (raloxifene)

### Description of studies

Seven papers were evaluated for inclusion. Two of these (McClung 2006; Smith 2004) only reported BMD as an outcome and were excluded from the review. One study was excluded because it was in patients with other conditions that were not representative. Four reports describing four trials were included (Ettinger 1999 [Multiple Outcomes of Raloxifene Evaluation–MORE] ; Lufkin 1998; Michalska 2004; Reginster 2003). The trialists in all four studies received industry support from Eli Lilly.

### ***Study design***

One study (Reginster 2003) was conducted in Europe with multiple centres. Lufkin (1998) was conducted in the USA, and Ettinger (1999) undertook the study in 25 countries. Michalska (2004) did not state in which country the trial was conducted.

Ettinger (1999) was the largest study with 7705 participants randomised to treatment. All other studies were on a much smaller scale (Reginster 2003, n = 596; Lufkin 1998, n = 143; Michalska 2004, n = 100).

### ***Population***

Participants in all four studies were postmenopausal women. In Ettinger (1999) and Reginster (2003) the women were at least 2 years' postmenopausal. In Lufkin (1998) the women were 5 years without menses or had levels of serum estradiol below 73 pmol/litre and follicular stimulating hormone levels above 30 IU/litre. Michalska (2004) did not state the number of years participants were postmenopausal.

Mean baseline femoral neck T-scores were reported (or calculated). The participants in Lufkin (1998) had a T-score (lateral lumbar spine) of  $-3.4$  (SD 0.5). Michalska (2004) reported a mean T-score (femoral neck) of  $-2.6$  (SD 0.7) for the raloxifene group and  $-2.8$  (SD 0.6) for those allocated to placebo. Participants in Reginster (2003) were described as having osteopenia or osteoporosis, and the mean T-score (femoral neck) was  $-2.9$  (SD 0.6). Ettinger (1999) stratified women into two groups and then randomised them to treatments: group 1 included those whose T-score was below  $-2.5$  SD; group 2 were required to have low BMD and one or more moderate or severe vertebral fractures at baseline or two or more moderate fractures, regardless of their BMD. In practice, the T-scores (femoral neck) were  $-3.0$  (SD 0.5) for group 1 and  $-3.2$  (SD 0.6) for group 2. Overall, two studies (Ettinger 1999; Lufkin 1998) were assessed as being in patients with osteoporosis only; and the other two studies (Michalska 2004; Reginster 2003) were in patients with osteoporosis or osteopenia.

In Lufkin (1998) all participants had at least one fracture at baseline. Some participants in Ettinger (1999) had one or more fractures at baseline: 10–11% in study group 1 and 88–90% in study group 2. In Reginster (2003) about 25% of women had fractures at baseline, and Michalska (2004) did not report if any women had prior fractures. Only Ettinger (1999) reported on smoking history, with about 17% of participants being smokers. Alcohol consumption was not reported in any of the studies.

The mean age of participants in the four studies ranged from 61 to 68 years. Limited data were available for ethnicity: in Reginster (2003) 99% of participants were white.

### ***Interventions***

Raloxifene was given as 60 mg/day, the licensed dose, although Ettinger (1999) and Lufkin (1998) randomised a third arm whereby raloxifene was given at 120 mg/day. Ettinger (1999) pooled the results of the two dosages for nonvertebral fractures, hip and wrist fractures and these combined data were used in the analyses; otherwise only the 60 mg dose was included.

Calcium and vitamin-D were given concomitantly in Lufkin (1998), Michalska (2004) and Reginster (2003). Ettinger (1999) gave participants calcium and cholecalciferol.

Reginster (2003) also gave both groups 20 mg/day fluoride ions. In Michalska (2004), one of the entry criteria was that participants should have been treated with alendronate 10 mg/day for at least 3 years prior to entry to the trial. We therefore treated both Reginster (2003) and Michalska (2004) as separate comparisons.

### ***Comparisons***

The following comparisons were carried out.

- Raloxifene versus placebo: two studies:
  - 1 year intervention period (Lufkin 1998)
  - 3 years (Ettinger 2000).
- Raloxifene plus fluoride versus fluoride: one study (Reginster 2003):

- 18 months raloxifene plus 152 mg monofluorophosphate versus placebo plus monofluorophosphate.
- Raloxifene versus placebo, following 3 years of alendronic acid: one study (Michalska 2004):
  - 1 year blinded intervention study
  - 1 year extension study with extra year being open-label same treatments (100% participants continued into this phase).

### ***Methodological quality***

All four studies were stated to be randomised. Lufkin (1998) used block randomisation. Ettinger (1999), Michalska (2004), and Reginster (2003) were all stated to be randomised, but did not provide details of the techniques used. Allocation concealment was partially adequate in one study (Ettinger 1999): randomisation was performed by Eli Lilly's clinical trials group, which was not involved in patient monitoring; allocation concealment was not reported in any of the other studies.

All four studies used double-blind methodology, although it was not made clear which of the investigators were blinded, except for Reginster (2003), which reported that the radiologist assessing the fractures was blinded to treatment allocation. The study by Michalska (2006) was partially blinded: the first year was double blind, but the second year consisted of open-label interventions; fracture results were reported after 2 years.

Sample size calculations were carried out by Ettinger (1999), Michalska (2004) and Reginster (2003). The study by Ettinger (1999) was designed to provide at least 90% power to detect a 40% reduction in vertebral fractures between raloxifene and placebo. Michalska (2004) calculated that 33 patients in each group would be needed to detect a 2.4% difference in lumbar spine BMD between groups with 80% power. Reginster (2003) calculated that 192 participants in each group were needed to provide a 90% power to detect a change in BMD of 1.5%. Lufkin (1998) gave no details of an a priori sample size calculation.

Ettinger (1999) reported missing data in each group, with 1152 people (22%) leaving the study in the raloxifene group and 652 (25%) dropping out from the placebo group. Lufkin (1998) stated that 13 out of 143 (9%) patients did not complete the study, but did not give further details. In Reginster (2003), 89 people (30%) did not complete the study in the raloxifene group, and 76 (26%) in the placebo arm. Michalska (2004) reported one person (3%) dropping out of the raloxifene group and none from the placebo group.

Reginster (2003) used available case analyses while Ettinger (1999), Lufkin (1998) and Michalska (2004) used ITT analyses. The first two of these carried forward the observations from the last assessment and included these interim results in the analysis.

Vertebral fractures were assessed radiographically by Ettinger (1999) using a cut-off point of at least a 20% decrease in height. Lufkin (1998) also used radiographic methods to determine vertebral fracture using cut-offs of 15% and 30% in height reduction; the latter values were used here. Reginster (2003) only stated that fractures were determined radiographically. Michalska (2004) did not report vertebral fracture incidence.

Fractures were primary outcomes in Ettinger (1999), Reginster (2003) and Lufkin (1998). In Michalska (2004) fractures appeared to be reported as safety data. In Ettinger (1999) fractures from trauma were excluded; Reginster (2003) reported that nonvertebral fractures were osteoporotic; Michalska (2004) did not define the type of nonvertebral fracture; and Lufkin (1998) did not report nonvertebral fracture data.

Details recorded at baseline in Reginster (2003) found participants comparable on age, weight, height, BMI and BMD. In Lufkin (1998), participants were comparable on age, years since menopause and BMI. In Michalska (2004) participants were comparable on age, weight, years since menopause and BMD. Participants in the study by Ettinger (1999) were comparable on age, BMI, smoking, BMD and previous oestrogen therapy.

Overall, two studies (Ettinger 1999; Reginster 2003) were considered to have potential for bias because of more than 20% missing data. These were treated

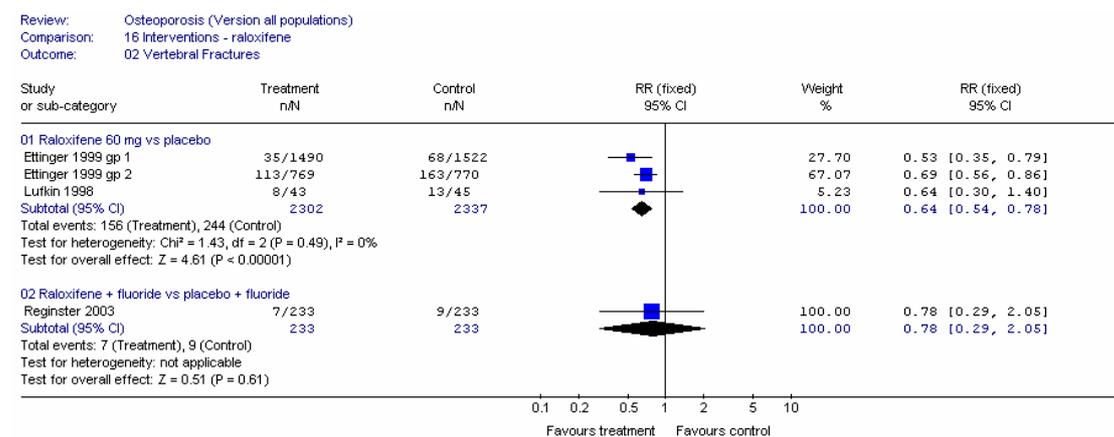
with caution and investigated in sensitivity analyses, when possible. For the nonvertebral fracture outcomes, the study by Ettinger (1999) also included 50% of patients taking an unlicensed dose; these results should be treated as partially indirect evidence. One further study (Michalska 2006) was considered to have some potential for bias because half of the study was unblinded.

## Results

### Raloxifene versus placebo

#### Vertebral fractures

Two studies reported the number of patients with vertebral fractures (Ettinger 1999; Lufkin 1999). The study by Ettinger (1999) stratified patients into one group with fractures at baseline and one group with mainly no fractures at baseline. These two groups were recorded separately and separate results were also given for the 60 mg arm for this outcome. A meta-analysis of the three comparisons in 4639 patients showed that significantly fewer patients had vertebral fractures in the raloxifene group compared with placebo (figure 55); RR 0.64 (95% CI 0.54 to 0.78). This corresponds to a NNT of 25 (95% CI 20 to 50) for a control group rate range of 4–29%. There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.49$ ). There was more than 20% missing data in the large study by Ettinger (1999), but sensitivity analysis left one unpowered study (Lufkin 1998) and the full analysis was taken.

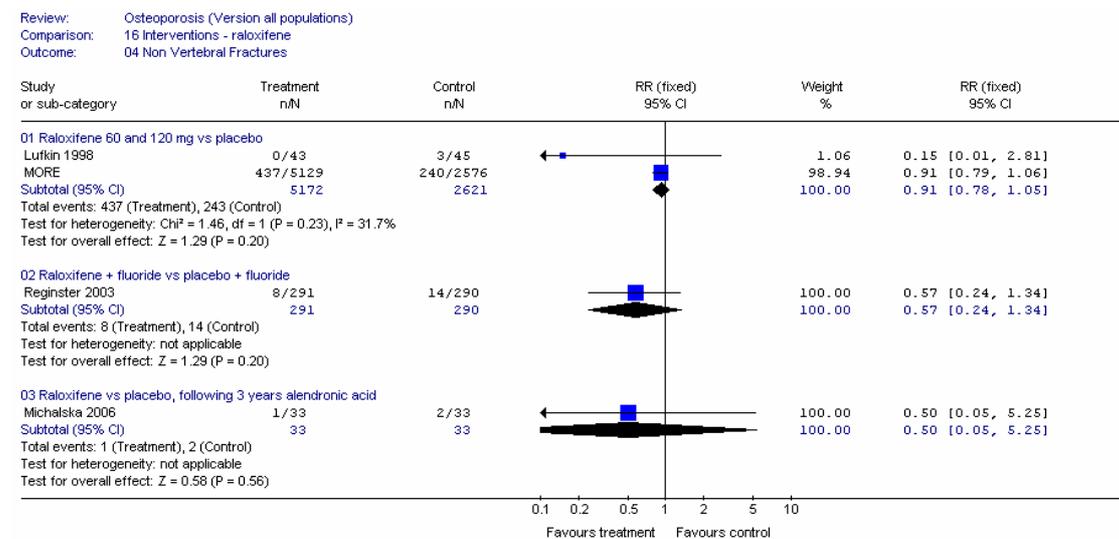


**Figure 55. Number of patients with vertebral fractures**

#### Nonvertebral fractures

Two studies reported the number of patients with nonvertebral fractures (Ettinger 1999; Lufkin 1999). A meta-analysis of two studies in 7793 patients

(figure 56) showed that there was no significant difference in the number of patients with nonvertebral fractures for raloxifene compared with placebo; RR 0.91 (95% CI 0.78 to 1.05), for a control group rate range of 7–9%. There was some heterogeneity but this was not significant ( $I^2 = 32\%$ ;  $p = 0.23$ ). For this outcome, the study by Ettinger (1999) reported combined results for both randomised doses of raloxifene (60 and 120 mg), and also combined the results for the two strata (fracture and nonfracture arms). There was also more than 20% missing data in this dominant study.

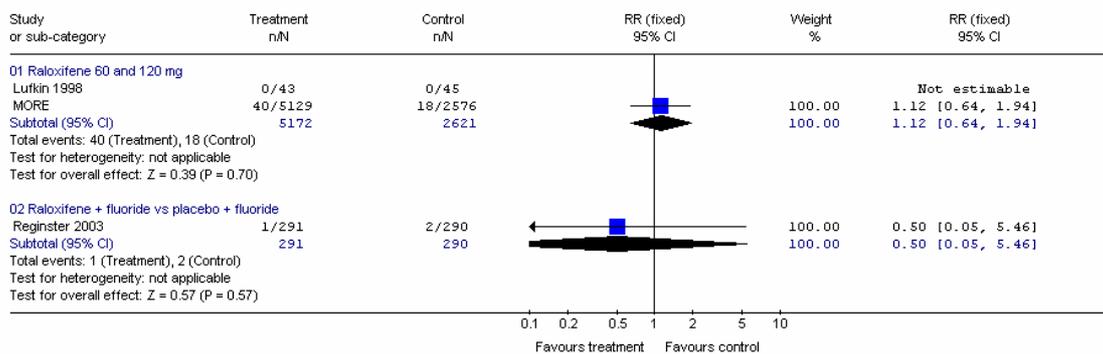


**Figure 56. Number of patients with nonvertebral fractures**

### *Hip fractures*

A meta-analysis of two studies in 7793 patients (figure 57) showed that there was no significant difference in the number of patients with hip fractures for raloxifene compared with placebo; RR 1.12 (95% CI 0.64 to 1.94), for a control group rate of 0.7%. The study by Ettinger (1999) included patients receiving the 120 mg unlicensed dose, and there was more than 20% missing data in this large study.

Review: Osteoporosis (Version all populations)  
 Comparison: 16 Interventions - raloxifene  
 Outcome: 05 Hip Fractures

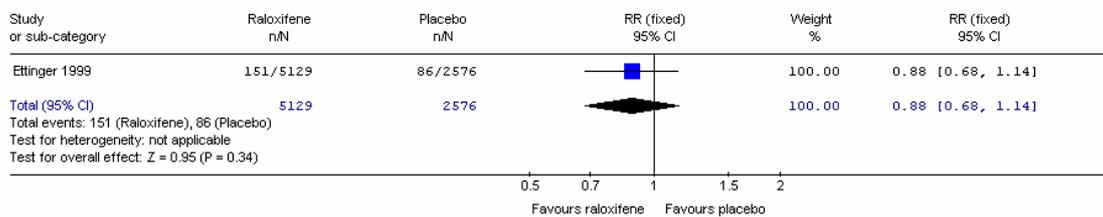


**Figure 57. Number of patients with hip fractures**

### Wrist fractures

One study in 7705 patients (figure 58) showed that there was no significant difference in the number of patients with a wrist fracture for raloxifene compared with placebo; RR 0.88 (95% CI 0.68 to 1.14) for a control group rate of 3%. The CI was fairly wide, however, and the study included patients receiving the 120 mg unlicensed dose.

Review: Raloxifene versus placebo for osteoporosis  
 Comparison: 01 RALOXIFENE versus PLACEBO  
 Outcome: 04 Wrist fracture



**Figure 58. Number of patients with wrist fractures**

### Raloxifene plus monofluorophosphate versus placebo plus monofluorophosphate

One study (Reginster 2003) compared raloxifene versus placebo in patients who were each given monofluorophosphate for an intervention period of 18 months.

### Vertebral fractures

One study in 466 patients, showed no significant difference in the number of people with a vertebral fracture, although the CI was fairly wide (figure 55); RR 0.78 (95% CI 0.29 to 2.05).

### Nonvertebral fractures

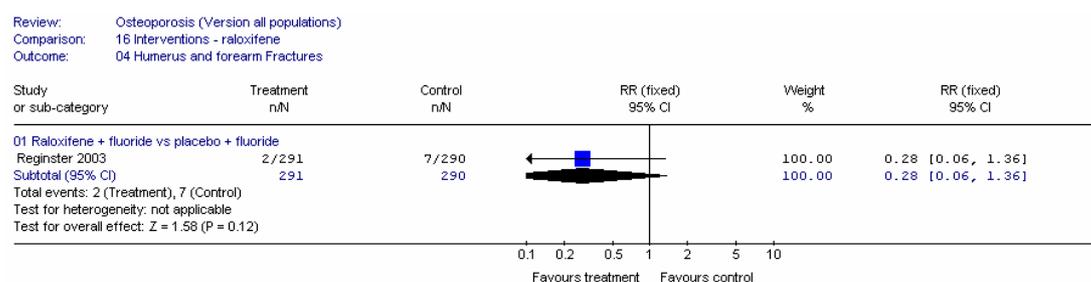
One study in 581 patients, showed no significant difference in the number of people with a vertebral fracture, although the confidence interval was fairly wide (figure 56); RR 0.57 (95% CI 0.24 to 1.34).

### Hip fractures

In one study of 581 patients, the CI was too wide to determine if there was a difference between interventions in the number of patients with a hip fracture (figure 57).

### Humerus or forearm fractures

In one study of 581 patients, the CI was too wide to determine if there was a difference between interventions in the number of patients with a humerus or forearm fracture (figure 59).



**Figure 59. Number of patients with humerus or forearm fractures**

### Raloxifene versus placebo, following 3 years of alendronic acid

One study (Michalska 2006) compared raloxifene with placebo in patients who had received alendronic acid for 3 years prior to the commencement of the trial. The trial consisted of 1 year of blinded allocation, then 1 year of open-label interventions. Fractures were reported as adverse events.

### Nonvertebral fractures

One study in 66 patients reported the number of patients with nonvertebral fractures. The CI was too wide to determine if there was a difference between interventions (figure 56).

### **2.4.3 Hormone replacement therapy (other): nandrolone**

#### ***Description of studies***

Seven studies were evaluated for inclusion. Four studies (Birkenhäger 1992; Erdtsieck 1994; Leidig-Bruckner 1997; Szücs 1992) were excluded because the dose was not licensed for use in the UK: the intervention in each of the excluded studies was nandrolone given at a dose of 50 mg every 4 weeks rather than every 3 weeks. Three studies were included (Frisoli 2005; Guesens 1986; Passeri 1993).

#### ***Study design***

None of the studies were conducted in the UK. One study took place in Belgium (Geusens 1986), one in Brazil (Frisoli 2005) and one in Italy (Passeri 1993).

Frisoli (2005) was part funded by a public body; Geusens (1986) had drugs supplied by Organon, Leo Pharmaceuticals and Sandoz; and Passeri (1993) did not provide details on funding.

All three studies had fewer than 100 participants (Frisoli 2005, n = 65; Geusens 1986, n = 60; Passeri 1993, n = 46). Two of these studies had 25 or fewer patients in the intervention arm (Passeri 1993, n = 25; Geusens 1986, estimated to be 20).

#### ***Population***

Two studies (Frisoli 2005; Passeri 1993) aimed to evaluate the effectiveness of nandrolone solely in postmenopausal women, whereas the population in Geusens (1986) included both men and postmenopausal women (overall 48 women and 12 men). Only one study (Passeri 1993) indicated how long the women had been postmenopausal, but this was only reported for women who completed the trial. The mean time since the menopause was 10.6 (SD 6) and 15.3 (SD 9) years in the nandrolone and placebo completers, respectively.

Mean baseline femoral neck T-scores were calculated for one study (Frisoli 2005) and calculated from lumbar spine BMD measurements in completers only in Passeri (1993). On this basis, both studies were assessed as being in

patients with osteoporosis or osteopenia. The other study, Geusens (1986), did not report BMDs.

Frisoli (2005) reported that some of the women had a history of fracture, but this was reported only for completers: 68 and 52% in the nandrolone and placebo groups had fractures at baseline; while in Passeri (1993) and Geusens (1986) all the women had fractures at baseline.

The age range, when given, across the studies was 25-76 years, with the mean age, when given, ranging from 63 to 77 years. Descriptions of ethnicity were only reported by Frisoli (2005), which described the women as white. None of the studies reported on smoking history.

### ***Interventions***

The intervention for all studies was a 50 mg nandrolone injection every 3 weeks.

Passeri (1993) gave all participants 1000 mg/day of calcium and Frisoli (2005) gave the women 500 mg/day of calcium. Geusens (1986) did not appear to give calcium supplements, and the comparator intervention was calcium. This study also had a third arm, alfacalcidol, but this was not included because this intervention is not licensed for osteoporosis in the UK.

### ***Comparisons***

The following comparisons were carried out.

- Nandrolone versus placebo:
  - 18-month intervention period (Passeri 1993)
  - 2 years (Frisoli 2005).
- Nandrolone versus calcium for 2 years (Geusens 1986).

### ***Methodological quality***

None of the studies reported the method of sequence generation and only one reported on allocation concealment: Frisoli (2005) reported that the women were randomised independently by the medical board of Organon/AKZO-Nobel (that is, an adequate method).

All three studies stated that their studies were double blind: Frisoli (2005) added that the participants and the principal investigators were blind to treatment allocation. The other studies did not provide further details on whether blinding also included the assessors. None of the studies tested the success of blinding for both participants and assessors.

In the study by Frisoli (2005), sample sizes were calculated on BMD outcome only. It was estimated that the sample size needed was 25 in the nandrolone group and 25 in the placebo group for a 7.5% increase in mean BMD at the femoral neck in the intervention group by 2 years. Geusens (1986) and Passeri (1993) did not report if sample size calculations were estimated.

Participants in Frisoli (2005) had baseline characteristics comparable for weight, height, BMI, blood counts and BMD, but were not comparable on age (the placebo group was significantly older). The other two studies only gave baseline characteristics for those completing the study.

One study had more than 50% missing data: Geusens (1986) had 34/60 (57%) dropouts overall, which was stated to be equal across groups. The other two studies had more than 20% missing data: in Frisoli (2005), four women dropped out of the nandrolone group (12.5%) and 12 were missing from the placebo group (36%); that is, a differential dropout rate. In Passeri (1993) there were 12 (48%) and 9 (43%) missing from the nandrolone and placebo groups, respectively. It was unclear in Frisoli (2005) if an ITT analysis was carried out, but the other two studies reported available case analyses.

Two of the studies (Frisoli 2005; Geusens 1986) stated that vertebral fractures were assessed radiographically with at least a 20% height reduction as a cut-off criterion. Passeri (1993) assessed vertebral fracture radiographically, but did not define fracture according to height reduction. Nonvertebral fractures were not reported in any of the studies. Fractures were reported as secondary outcomes in two studies (Frisoli 2005; Geusens 1986), and as adverse events in Passeri (1993). None of the studies reported nonvertebral fractures.

Overall, all the studies were considered to be at higher risk of bias because of missing data. Geusens (1986) had more than 50% missing data and this was

considered to be too unreliable to use in the analysis. The other studies had less than 50% missing data, but Frisoli (2005) also had differential missing data. Frisoli (2005) had a significant difference in age between the nandrolone and placebo groups. It was decided to include Frisoli (2005) and Passeri (1993) in the analysis, but both were considered to have higher potential for bias, and therefore their evidence grading should be lowered.

## Results

### Nandrolone versus placebo

#### Vertebral fractures

Two studies reported the number of patients with vertebral fractures (Frisoli 2005; Passeri 1993). In the former study, results were presented as the number of patients with a fracture and as percentages; we used these data to calculate the denominators, which corresponded to the numbers completing the study. A meta-analysis of these two studies in 95 participants showed there was no significant difference in the number of patients with vertebral fractures in the nandrolone group compared with placebo (figure 60); RR 0.89 (95% CI 0.61 to 1.31). There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.43$ ).

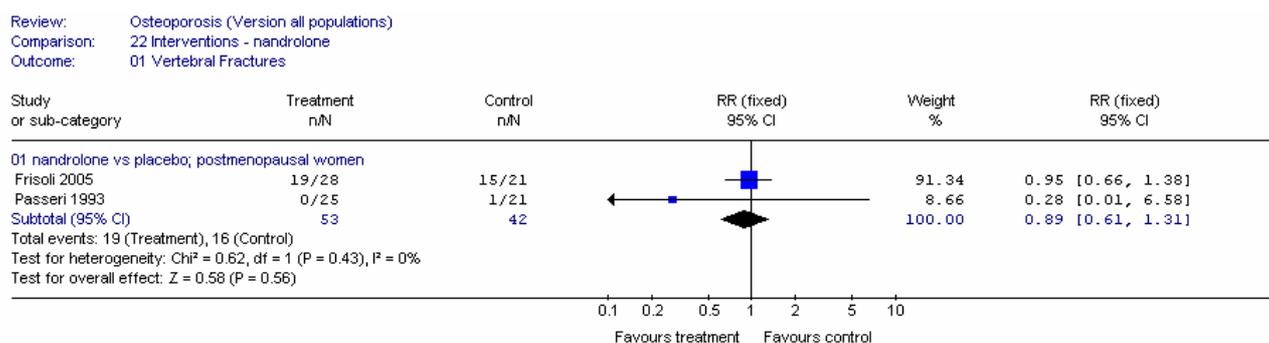


Figure 60. Number of patients with vertebral fractures

## 2.5 Vitamin and mineral supplements

### 2.5.1 Vitamin D

#### Description of studies

Seventy-one papers were evaluated for inclusion. Seven of these papers reported only on BMD as an outcome and were excluded from the review

(Baeksgaard 1998, Barone 2007, Bernstein 1996, Buckley 1996, Lambrinoudaki 2000, McDonald 2006, Mikati 2006). Five further studies were excluded because the population was a higher-risk group (Bauman 2005; El-Husseini 2004; Sato 2005; Van Cleemput; Wissing 2005). Thirty-two studies had interventions that were not in the BNF (Aloia 1988; Dukas 2004; Dykman 1984; Gallagher 1990; Garay Lillo 1997; Geusens 1986 (included in the nandrolone review); Gorai 1999; Ishida 2004; Lakatos 2000; Menczel 1994; Mizunuma 2006; Munno 1989; Orino 1987; Orino 1994; Orwoll 1989; Ott 1989; Peacock 2000 (included in the calcium review); Reginster 1999; Ringe 1999; Sambrook 1993; Sambrook 2000; Sato 1997; Sato 1999; Shiraki 1996; Shiraki 2004; Sosa 2000; Stempfle 1999; Stempfle 2002; Talalaj 1996; Ushiroyama 2001; Yamada 1989; Zhang 2005) and five other studies were excluded for reasons given in Appendix IV. The remaining 22 studies were included (Adachi 1996; Avenell 2004; Chapuy 1992 (Decalyos); Chapuy 2002 (Decalyos II); Dawson-Hughes 1997; Ebeling 2001; Falch 1987; Gallagher 1989; Gallagher 2001 (STOPIT); Harwood 2004 [Nottingham Neck of Femur (NoNOF)]; Jackson 2006 (Women's Health Initiative; WHI); Komulainen 1998; Larsen 2004; Lips 1996; Lyons 2007; Meyer 2002; Pfeifer 2000; Porthouse 2005; RECORD 2005; Smith 2007; Tilyard 1992; Trivedi 2003).

Two studies (Gallagher 2001; Komulainen 1998) also included comparisons of vitamin D with placebo as adjuncts to HRT regimens that were not in the BNF. These comparisons were not included in the review.

All studies but one (Pfeifer 2000) gave the participants the interventions on a regular basis. Pfeifer (2000) was designed to investigate the effect of a short term intervention on subsequent progress. Therefore the trialists gave an 8-week intervention and then followed the participants for two years.

### ***Study design***

Seven studies were conducted in the UK (Avenell 2004; Harwood 2004; Lyons 2007; Porthouse 2005; RECORD 2005; Smith 2007; Trivedi 2003); four in the United States (Dawson-Hughes 1997; Gallagher 1989; Gallagher 2001; Jackson 2006); two in France (Chapuy 1992; Chapuy 2002); two in Norway (Falch 1987; Meyer 2002); one each in Germany (Pfeifer 2000); Holland (Lips

1996); Denmark (Larsen 2004); Finland (Komulainen 1998); Canada (Adachi 1996); Australia (Ebeling 2001) and New Zealand (Tilyard 1992).

Twelve studies received some funding from industry (Chapuy 2002; Ebeling 2001; Gallagher 1989; Gallagher 2001; Harwood 2004; Komulainen 1998; Larsen 2004; Meyer 2002; Pfeifer 2000; Porthouse 2005; Smith 2007; Tilyard 1992); three studies had the drugs provided by industry (Avenell 2004; Chapuy 1992; RECORD 2005); two studies (Adachi 1996; Falch 1987) did not state the source of funding. Five studies were funded by non-industry sources (Dawson-Hughes 1997; Jackson 2006; Lips 1996; Lyons 2007; Trivedi 2003).

Four studies had fewer than 100 participants (Adachi 1996 n = 62; Ebeling 2001 n = 41; Falch 1987 n = 86; Gallagher 1989 n = 62). Three studies had between 100 and 200 participants (Avenell 2004; Harwood 2004; Pfeifer 2000). Five studies had between 400 and 1000 (Chapuy 2002; Dawson-Hughes 1997; Gallagher 2001; Komulainen 1998; Tilyard 1992). Nine studies had over 1000 participants (Chapuy 1992 n = 3270; Larsen 2004 n = 9605; Lips 1996 n = 2578; Lyons 2007 n = 3440; Meyer 2002 n = 1144; Porthouse 2005 n = 3314; RECORD 2005 n = 5292; Smith 2007 n = 9440; Trivedi 2003 n = 2686). Finally, Jackson (2006) had 36,282 participants.

Three studies had more than two arms: Avenell (2004) and the RECORD trial (2005) had four arms: vitamin D, calcium, vitamin D plus calcium and placebo. Harwood (2004) had 4 arms: oral vitamin D plus oral calcium; injected vitamin D and oral calcium; single injection vitamin D and no treatment; and Larsen (2004) had four arms: vitamin D plus calcium plus a leaflet, an educational intervention plus a leaflet, both interventions and no treatment. Overall, 30 comparisons were included in this review.

Seven studies took place in secondary care (Adachi 1996; Avenell 2004; Falch 1987; Gallagher 1989; Harwood 2004; RECORD 2005; Tilyard 1992); ten took place in primary care (Dawson-Hughes 1997; Gallagher 2001; Jackson 2006; Komulainen 1998; Larsen 2004; Lips 1996; Pfeifer 2000; Porthouse 2005; Smith 2007; Trivedi 2003). Four studies recruited people

from nursing homes or sheltered housing (Chapuy 1992; Chapuy 2002; Lyons 2007; Meyer 2002). One study did not report the setting (Ebeling 2001).

### ***Population***

Eleven studies included only postmenopausal women (Chapuy 1992; Chapuy 2002; Falch 1987; Gallagher 1989; Gallagher 2001; Harwood 2004; Jackson 2006; Komulainen 1998; Pfeifer 2000; Porthouse 2005; Tilyard 1992). One study (Ebeling 2001) was in men; one (Adachi 1996) was in participants receiving glucocorticoids and included people with polymyalgia rheumatica, temporal arteritis, asthma, vasculitis or SLE); this study included men and women. Nine other studies included men and women (Avenell 2004; Dawson-Hughes 1997; Larsen 2004; Lips 1996; Lyons 2007; Meyer 2002; RECORD 2005; Smith 2007; Trivedi 2003). Of the studies that included postmenopausal women, only three recorded the time the women had been postmenopausal and this ranged from 6 months (Komulainen 1998) to 15 years (Tilyard 1992).

Mean baseline femoral neck T-scores were reported (or calculated) for seven studies (Chapuy 2002; Dawson-Hughes 1997; Ebeling 2001; Gallagher 2001; Harwood 2004; Jackson 2006; Komulainen 1998); one reported at the spine (Adachi 1996) and the remaining studies did not report BMD values. In two studies (Falch 1987; Gallagher 1989) the authors' classification of osteoporosis status was taken. On this basis, three studies were assessed as being in participants with osteoporosis only (Falch 1987; Gallagher 1989; Tilyard 1992); four studies were assessed to include participants with osteoporosis or osteopenia (Chapuy 2002; Ebeling 2001; Harwood 2004; Jackson 2006); one study was in people with osteopenia or normal BMD (Dawson-Hughes 1997); one study included participants with normal BMD (Komulainen 1998) and the rest did not record BMD.

Fifteen studies reported the proportions of participants who had a prior fracture. In seven, all participants had a prior fracture (Avenell 2004; Ebeling 2001; Falch 1987; Gallagher 1989; Harwood 2004; RECORD 2005; Tilyard 1992). For the remaining eight studies, the percentage of participants with at least one vertebral fracture at baseline varied from 6 to 70% across studies; (Adachi 1996 23%; Gallagher 2001 28% vitamin D and 14% placebo; Jackson

2006 34%; Komulainen 1998 15%; Larsen 2004 26%; Meyer 2002 26-29%; Porthouse 2005: around 58%; Smith 2007 37-39%). The number of participants with a prior fracture was not reported in five studies (Chapuy 1992; Chapuy 2002; Dawson-Hughes 1997; Lyons 2007; Trivedi 2003). Lips (1996) reported that people with history of hip fracture were excluded and Pfeifer (2000) excluded people with osteoporotic fractures of the extremities.

The age range of participants across the studies ranged from 27 (Ebeling 2001) to 107 years (Lyons 2007), with the mean age, where given, ranging from 52 (Komulainen 1998) to 85 years (Chapuy 2002).

RECORD (2005) reported the ethnicity of participants in the trial: around 99% were white. Jackson (2006) reported that 83% were white, 9.2% black, 4.2% Hispanic, 0.4% American Indian or Native American, 2% Asian or Pacific Islander and 1.2% ethnicity not identified. Gallagher (2001) reported 98.2% were white, 1.2% black, and 0.4% Asian. The remaining studies did not report ethnicity.

Five studies reported the BMI of participants (Harwood 2004; Jackson 2006; Komulainen 1998; Meyer 2002; Trivedi 2003). The mean ranged from 22 to 29 kg/m<sup>2</sup>.

The proportion of smokers was stated in eight studies and ranged from around 4% (Trivedi 2003) to 12% (RECORD 2005); it was not reported in the remaining studies.

Alcohol consumption was reported in two studies (Jackson 2006; Trivedi 2003). In one study (Trivedi 2003) around 10% of participants never drank alcohol, 15% were occasional drinkers and 75% regular drinkers and in Jackson (2006): 10% never consumed alcohol; 18% consumed alcohol in the past; 14% consumed less than 1 drink per month; 21% consumed less than 1 drink per week; 26% consumed 1 to less than 7 drinks per week and 11% consumed 7 or more drinks per week. Smith (2007) reported that 4 of 200 participants consumed more than 20 units alcohol per week.

Most studies disallowed prior treatment with osteoporosis drugs (Adachi 1996; Chapuy 2002; Dawson-Hughes 1997; Falch 1987; Gallagher 1989; Gallagher 2001; Harwood 2004; Pfeifer 2000; RECORD 2005; Tilyard 1992). The exceptions were: Chapuy (1992) which allowed prior oestrogen; Jackson (2006), which allowed calcium up to 1000mg/day and vitamin D up to 1000IU/day in addition to the study drug. In this study, at baseline, around 30% in each group had calcium supplementation of at least 500mg/day and about 40% in each group had 400 IU or more of vitamin D (supplements and diet) per day. In addition, 52% of the women in each group also received HRT during the trial; bisphosphonate and calcitonin were allowed and these were taken by 1% of women at baseline but this increased to 17% during the trial (it was unclear if this differed between groups). Komulainen (1998) and Lips (1996) did not exclude on grounds of prior treatments, but Lips (1996) allowed the participants to have a vitamin supplement containing vitamin D during the trial; this happened for only 3% (same in each group). In the Porthouse (2005) control group, 6% used supplements at 18 months into the trial. In Larsen (2004), three of the randomised programmes included revision of the patient's medication.

### ***Interventions***

The following interventions were used:

#### **Vitamin D**

- hydroxylated Vitamin D: calcitriol (1,25-dihydroxycholecalciferol) 250 nanograms twice daily (Ebeling 2001; Falch 1987; Gallagher 1989; Gallagher 2001; Tilyard 1992)
- native vitamin D3 (cholecalciferol) in 13 studies:
  - 300 IU/day, extending to 100 IU/day for the 5th year (Komulainen 1998)
  - 400 IU/day (Falch 1987; Jackson 2006; Larsen 2004; Lips 1996)
  - native vitamin D3 in form of cod liver oil, containing 400 IU /day (Meyer 2002)
  - 700 IU/day (Dawson-Hughes 1997)
  - 800 IU/day (Avenell 2004; Chapuy 1992; Chapuy 2002; Harwood 2004; Porthouse 2005; RECORD 2005)

- 800 IU/day given for 8 weeks then followed for 1 year (Pfeifer 2000)
- 100,000 IU orally every 4 months (Trivedi 2003)
- native vitamin D2 (ergocalciferol) in 3 studies
  - 100,000 IU orally every 4 months (Lyons 2007)
  - 300,000 IU intramuscular injection every year (Harwood 2004; Smith 2007)
- native Vitamin D, type not specified
  - 50,000 IU/day (Adachi 1996)

## Calcium

Twelve studies also gave calcium interventions concurrently with vitamin D:

- 500mg/day as calcium lactate for 10 months (i.e. 93mg elemental Ca/day) (Komulainen 1998)
- 500 mg/day as calcium citrate malate (Dawson-Hughes 1997)
- 1 g/day as calcium carbonate (Adachi 1996; Avenell 2004; Harwood 2004; Jackson 2006; Larsen 2004; Porthouse 2005; RECORD 2005)
- 1 g/day type of calcium not specified (Pfeifer 2000)
- 1.2 g/day as tricalcium phosphate (Chapuy 1992; Chapuy 2002).

For the studies in which calcium was a comparator, the following calcium regimens were used:

- 500mg/day as calcium lactate for 10 months (i.e. 93mg elemental Ca/day) (Komulainen 1998)
- 1 g/day calcium gluconate (5.2 g calcium salt) (Tilyard 1992)
- 1 g/day as calcium carbonate (Avenell 2004; RECORD 2005)
- 1 g/day calcium salt not stated (Ebeling 2001)
- 1.2 g/day as calcium carbonate (Pfeifer 2000)

Of these interventions, all but Chapuy (1992), Chapuy (2002) and Komulainen (1998) corresponded to the type and dose in the BNF; the two Chapuy studies had a different calcium formulation but the same dose. Komulainen (1998) and Dawson-Hughes (1997) appeared to have a lower dose. It was decided to include all the studies, nevertheless.

## **Comparisons**

Comparisons included were:

- Hydroxylated vitamin D versus placebo
  - Calcitriol versus placebo
    - ◇ 1 year intervention period (Gallagher 1989)
    - ◇ 3 years (Gallagher 2001)
- Hydroxylated vitamin D versus calcium
  - calcitriol versus calcium
    - ◇ 2 years intervention period (Ebeling 2001)
    - ◇ 3 years (Tilyard 1992)
- Native vitamin D versus placebo/no treatment
  - Colecalciferol (vitamin D3) versus placebo
    - ◇ 2 to 5 years (RECORD 2005)
    - ◇ 3 years (Meyer 2002)
    - ◇ 3.5 years (Lips 1996)
    - ◇ 5 years (Trivedi 2003)
  - Vitamin D3 versus no treatment
    - ◇ Up to 46 months (Avenell 2004)
  - Vitamin D2 versus placebo
    - ◇ Oral route, 5 years (Lyons 2007)
    - ◇ IM route, 5 years (Smith 2007)
  - Vitamin D2 versus no treatment
    - ◇ IM route, 5 years (Harwood 2004)
- Native vitamin D plus calcium versus calcium
  - Vitamin D3 plus calcium versus calcium
    - ◇ 2 to 5 years (Avenell 2004; RECORD 2005)
    - ◇ 5 years (Komulainen 1998)
    - ◇ 8 weeks intervention, then followed for 1 year (Pfeifer 2000)
- Native vitamin D versus calcium
  - Vitamin D3 versus calcium
    - ◇ 2 to 5 years (Avenell 2004; RECORD 2005)
- Native vitamin D plus calcium versus placebo/no treatment

- Vitamin D3 plus calcium versus placebo
  - ◇ 18 months (Chapuy 1992)
  - ◇ 2 years (Chapuy 2002)
  - ◇ 2 to 5 years (RECORD 2005)
  - ◇ 3 years (Dawson-Hughes 1997)
  - ◇ 7 years (Jackson 2006) (52% participants received HRT, 30% had additional calcium; 40% had additional vitamin D above 400 IU/day)
- Vitamin D (unspecified) plus calcium versus placebo
  - ◇ 3 years intervention period (Adachi 1996)
- Vitamin D3 plus calcium versus no treatment
  - ◇ 1 year (Harwood 2004)
  - ◇ 18 months (Porthouse 2005) both groups also had leaflet on falls prevention and dietary advice
- 42 months (Larsen 2004) all participants received nurse-led educational/environmental home visit
  - ◇ up to 46 months (Avenell 2004)
- Vitamin D2 plus calcium versus no treatment
  - ◇ 1 year (Harwood 2004)
- Comparison of different types of vitamin D
  - Calcitriol versus colecalciferol
    - ◇ 3 years (Falch 1987)

Five studies reported baseline vitamin D levels: the serum 25-hydroxyvitamin D level was 8.9 ng/ml in Chapuy (2002); 29 nmol/l in Harwood (2004); 78 nmol/l in Gallagher (2001); and less than 10 microg/day in Meyer (2002). In Smith (2007), people taking 400 IU/day (10 microg/day) or more of vitamin D were excluded. Recommended levels are 30 ng/ml and 75 nmol/ml or more, and 15 microg/day for people aged over 70 years. Thus, all the studies giving vitamin D levels except Gallagher (2001) had people who were vitamin D deficient.

Twelve studies reported the baseline calcium levels: in three studies the calcium serum level was 2.36 mmol/l (Adachi 1996; Falch 1987; Harwood 2004), i.e. normal levels. Chapuy (2002) reported a baseline level of 9.2mg/dl.

Mean calcium daily intakes at baseline were: 675-697 mg/day (Ebeling 2001); 859 and 876 mg/day (Lips 1996); 885-899 mg/day (Tilyard 1992). During the study, Jackson (2006) had a mean intake of 1150 mg/day; Meyer (2002) had a mean calcium dietary intake of 446-456 mg/day and Smith (2007) 625 mg/day. Gallagher 1989 allowed a free calcium intake. The recommended intake for people over 50 years is 1.2g/day, i.e. there was calcium deficiency in all studies except Jackson (2006).

### ***Methodological quality***

The method of sequence generation was adequate in eight studies, in which a computer-generated sequence was employed in seven (Avenell 2004; Harwood 2004; Komulainen 1998; Lips 1996; Lyons 2007; Porthouse 2005 and RECORD 2005) and minimisation was used in one (Adachi 1996); the remaining studies did not report the method.

Six studies reported adequate allocation concealment: in two (Avenell 2004; RECORD 2005) a centralised facility was used; five described randomisation by an independent third party (Gallagher 2001; Lips 1996; Lyons 2007; Porthouse 2005; Trivedi 2003) and Smith (2007) used numbered identical ampoules prepared by a third party and distributed sequentially. Allocation concealment was partially reported in three studies (Harwood 2004; Komulainen 1998; Meyer 2002). The other studies did not describe allocation concealment.

Participants were not blinded in seven studies: Komulainen (1998) was described as being open for all treatment groups. The participants were not blinded in Avenell (2004), no placebos were used in four other studies (Harwood 2004; Larsen 2004; Porthouse 2007; Tilyard 1992) and Falch (1987) was unlikely to be blinded because different frequencies of dosing were used. The remaining studies had either double blinding or single blinding of participants.

Blinding of the outcome assessors was reported in the majority of studies; it was unclear in six studies (Chapuy 1992; Ebeling 2001; Lips 1996; Meyer

2002; Porthouse 2005; Pfeifer 2000). Outcome assessors were not blinded in Harwood (2004).

Nine studies (Avenell 2004; Chapuy 1992; Jackson 2006; Lyons 2007; Meyer 2002; Porthouse 2005; RECORD 2005; Smith 2007; Tilyard 1992) carried out a priori power calculations based on fracture outcomes, three studies were powered for BMD (Ebeling 2001; Gallagher 2001; Komulainen 1998) and power calculation was not reported in the remaining studies.

Chapuy (1992) reported that “the sample size for this study was chosen so that a reduction of 30 percent in the annual hip fracture rate, which was estimated at 3.5 percent on the basis of the mean age of the women, could be detected”, but no actual figures for sample size required were given; 3,270 women were recruited. Jackson (2006) reported 85% power to detect an 18% reduction in hip fracture in 35,000 women at a placebo rate of 33.6 per 10,000 person-years (=99% power for total fractures); 36,282 women were recruited. Lyons (2007) reported a power calculation suggesting that 4,000 participants per group would give 82% power to detect a fracture reduction from 20% to 17.5% after 3 years, but only 3,440 people were actually recruited (i.e. the study was underpowered). Meyer (2002) stated that a sample size of 1113 in each group would give 80% power to detect reduction in hip fractures of 30; the study recruited 1144 participants in total, i.e. the study was underpowered. Porthouse (2005) reported that to observe a 34% reduction in the fracture rate (from a control rate of 10%) with 80% power would need 2,855 participants in a 2:1 ratio in favour of the control group; 3,314 women were recruited. RECORD (2005) reported that 4,200 participants were required to have 80% power ( $p < 0.05$ ) to detect a decrease in incidence of fractures from 15% in controls to 12%; 5,292 people were recruited. Smith (2007) stated that there was 80% power to detect a difference in nonvertebral fractures of 30% at 5% significance with 5000 participants in each arm; the study had 9440 participants across two arms. Tilyard (1992) reported that 554 women were required to detect a 50% difference in fracture rates (2-tailed  $p < 0.05$ ) with 90% power; 622 women were recruited.

All the studies but three (Dawson-Hughes 1997; Gallagher 2001; Harwood 2004) were comparable on several baseline characteristics. Adachi (1996) stated that the groups were comparable on age, gender, prednisone dose and BMD. Chapuy (1992) stated that the groups were comparable on age, weight, height, dietary calcium, and the proportion of fallers. Chapuy (2002) stated that the groups were comparable on age, weight, height, calcium and vitamin D intake, serum calcium and vitamin D and BMD. Falch (1987) reported that the groups were comparable on age, weight, height, years since menopause and bone mineral content. Gallagher (1989) reported that the groups were comparable on age and the number of prior vertebral fractures, but the number of characteristics given was low. The groups in Jackson (2006) were comparable on age, race, family and personal history of fracture, falls, BMI, physical activity, calcium intake, vitamin D, sunlight, alcohol, smoking and use of HRT. The groups in Komulainen (1998) were comparable on age, BMI, years since menopause, prior fracture, smoking, alcohol, physical activity, calcium intake and BMD. The groups in Lips (1996) were comparable on age, gender, residence, amount of time outdoors/in sunshine, mobility and calcium intake. The groups in Lyons (2007) were comparable on age, gender and residence. The groups in Porthouse 2005 were comparable on age, weight, prior fracture, smoking, health, maternal hip fracture, falls and calcium intake. The groups in RECORD (2005) were comparable on age, gender, race, type of fracture, time since fracture, weight, smoking, physical activity, thyroxine, steroids and thiazide diuretics. The groups in Tilyard (1992) were comparable on age, number of prior vertebral fractures, weight, height, years since menopause and dietary calcium. The groups in Trivedi (2003) were comparable on age, BMI, comorbidity, smoking, steroids, HRT, alcohol and physical activity.

Dawson-Hughes (1997) stated that the groups were comparable on age, weight, height, calcium and vitamin D intake, smoking, physical activity and BMD, but were not comparable for dietary calcium (placebo group was higher). In Gallagher (2001) participants were comparable on age, weight, height, BMD, but not comparable in the percentage of women with vertebral fractures (higher in the group randomised to calcitriol). Harwood (2004) stated

that the groups were comparable on age, smoking, dietary Ca, alcohol, mobility, biochemistry, BMD; but were not comparable for the number of participants with no prior fracture (control 76%, intervention 50%); and the vitamin D alone group had a higher proportion with hypovitaminosis.

Eight studies (Adachi 1996; Chapuy 1992; Chapuy 2002; Harwood 2004; Lyons 2007; Meyer 2002; Smith 2007; Tilyard 1992) reported missing data in excess of 20%:

In Adachi (1996) 64% and 61% in the treatment and placebo groups, respectively refused follow-up. In Chapuy (1992), deaths accounted for 16% of the vitamin D group and 17% of the placebo group, while loss to follow up was 3 and 4% respectively. The total missing data was 46% in each group; the analysis excluded only death and loss to follow up. In Chapuy (2002), missing data was largely due to death in this sample of elderly participants; missing data were similar in each group (28% on treatment and 36% placebo; of which deaths accounted for 18% and 24%). In Harwood (2004), missing data was 50/150 (33%) overall, of whom 23% died (14-31% between groups) and the rest withdrew from follow up, but it was not stated from which groups. In Lyons (2007) (an elderly population), 47% either died or were lost to follow up in each of the vitamin D and placebo groups. In Meyer (2002) there were: 30% and 28% deaths respectively in the vitamin D and placebo groups. Only 68% and 65% did not complete their treatments. In Smith (2007), the missing data in each group was 33% on vitamin D and 35% on placebo: these participants failed to return their fracture questionnaires. All participants randomised were included in the analysis. In Tilyard (1992), missing data seemed to be 32% in the vitamin D group and 28% in the calcium group. Only 69% completed 3 years.

Avenell (2004) reported a differential missing data, with vitamin D at 43%; calcium at 14%; no tablets at 17%. The other studies reported missing data of less than 20%. Of these studies, two had differential missing data between groups: Falch (1987) reported missing data of 17% on calcitriol and 5% on colecalciferol; Gallagher (2001) reported 9% on placebo and 18% on calcitriol.

The analysis was by intention to treat (ITT) with all cases followed in one study (Larsen 2004) and ITT with all included in the analysis or fracture incidence carried forward in ten other studies (Avenell 2004; Ebeling 2001; Komulainen 1998; Lips 1996; Lyons 2007; Meyer 2002; Porthouse 2005; RECORD trial 2005; Smith 2007; Trivedi 2003). The analysis was per protocol in five studies (Adachi 1996; Chapuy 1992; Chapuy 2002; Dawson-Hughes 1997; Jackson 2006) and available case analysis in the other studies apart from Gallagher (2001) which was unclear.

The method of vertebral fracture assessment was reported in six studies (Adachi 1996; Ebeling 2001; Falch 1987; Gallagher 1989; Gallagher 2001; Tilyard 1992). Two studies reported radiographic assessment with 15% reduction height (Adachi 1996, Falch 1987), one study defined fracture as a decrease of 15% or more during any one year in the anterior or posterior height of the body of a vertebra from T4 through L4 (Tilyard 1992) and two studies (Ebeling 2001; Gallagher 2001) defined new fracture as a 20% reduction in any vertebral height compared with baseline in a previously normal vertebra. In one study (Gallagher 1989) baseline vertebral fracture was diagnosed if anterior height of the vertebra was 25% less than the posterior height and new fractures were noted if the anterior height of the vertebrae decreased by 15% from the baseline measurement. Two studies reported that a radiographic method was used but gave no details (Chapuy 2002; Jackson 2006). One study (Trivedi 2003) reported that incidences of fracture were based on self-report or death certificates. Five studies reported clinical vertebral fractures (Avenell 2004; Harwood 2004; Larsen 2004; Lyons 2007; RECORD 2005).

Sixteen studies (Avenell 2004; Chapuy 1992; Chapuy 2002; Falch 1987; Gallagher 1989; Jackson 2006; Komulainen 1998; Larsen 2004; Lips 1996; Lyons 2007; Meyer 2002; Porthouse 2005; RECORD 2005; Smith 2007; Tilyard 1992; Trivedi 2003) reported fracture data as the primary outcome; four studies (Adachi 1996; Dawson-Hughes 1997; Ebeling 2001; Gallagher 2001) reported fracture data as a secondary outcome and two studies reported fractures as an adverse event (Harwood 2004; Pfeifer 2000).

Three studies (Jackson 2006; Porthouse 2005; RECORD 2005) excluded traumatic fractures; Harwood (2004) and Pfeifer (2000) reported fractures as a consequence of a fall; Komulainen (1998) stated that they did not exclude trauma fractures, however the study also reported that there were no fractures associated with car accidents or other major trauma.

Dawson-Hughes (1997) reported that 76% of the participants had fractures classified as osteoporotic, but did not give this by group or site. They did, however, report the RR for the number of participants with nonvertebral osteoporotic fractures. The rest of the studies did not say if the fractures were osteoporotic.

Overall, we considered that 16 studies (Adachi 1996; Avenell 2004; Chapuy 1992; Chapuy 2002; Dawson-Hughes 1997; Falch 1987; Gallagher 1989; Gallagher 2001; Harwood 2004; Komulainen 1998; Larsen 2004; Lyons 2007; Meyer 2002; Porthouse 2007; Smith 2007; Tilyard 1992) had higher potential for bias.

Of these studies, Adachi (1996) had more than 60% missing data and it was decided not to include this study in the analysis; this study also used a decrease in height of 15% as a marker for vertebral fracture.

Five studies had more than one methodological problem:

- Avenell (2004) had a lack of patient blinding as well as a differential loss to follow up (vitamin D 43% versus no tablets 17%).
- Falch (1987) had a lack of patient blinding and a differential loss to follow up (17% calcitriol versus 5% colecalciferol) plus a 15% cut off for vertebral fractures.
- Gallagher (2001) had baseline differences (more vertebral fractures in the calcitriol group) and differential loss to follow up (18% calcitriol versus 9% placebo)
- Harwood (2004) had a lack of patient blinding, and differences in baseline characteristics (the number of participants with no prior fracture and the number with hypovitaminosis), plus 33% missing data

- Tilyard (1992) had a lack of patient blinding and missing data of 32 and 28%, plus a 15% cut off for vertebral fractures.

All of these studies were treated with caution and considered in sensitivity analyses.

Five additional studies had more than 20% missing data: Chapuy 1992 (46% in each group); Chapuy 2002 (28% and 36%); Lyons 2007 (47% in each group); Meyer 2002 (32 and 35%); Smith 2007 (33 and 35%). These studies were all treated with caution, especially Chapuy (1992) and Lyons (2007).

Five other studies had only one methodological problem: Dawson-Hughes (1997) had differences in patient characteristics (dietary calcium was higher for the placebo group); Gallagher (1989) had a cutoff of 15% height reduction. Three studies had no patient blinding (Komulainen 1998; Larsen 2004; Porthouse 2007). These studies also had some potential for bias and were considered in sensitivity analyses.

## **Results**

### *Concordance*

This was reported in 17 studies:

- nine had more than 75% concordance (Chapuy 1992; Chapuy 2002; Komulainen 1998; Lips 1996; Lyons 2007; Pfeifer 2000; Smith 2007; Tilyard 1992; Trivedi 2003)
- Gallagher (2001) had 70 and 78% respectively for calcitriol and placebo
- Dawson Hughes (1997) reported 92-93% concordance amongst completers
- Four studies had between 50 and 75% concordance (Avenell 2004 had 65%; Jackson 2006 had 59%; Meyer 2002 had cessation of treatment of 36% vitamin D and 39% placebo; Porthouse 2005 had 59%).
- In Larsen (2004) those recruited were offered the interventions and participation was: 48% in the Environment programme; 56% in Vitamin D plus calcium group and 45% for the combined group

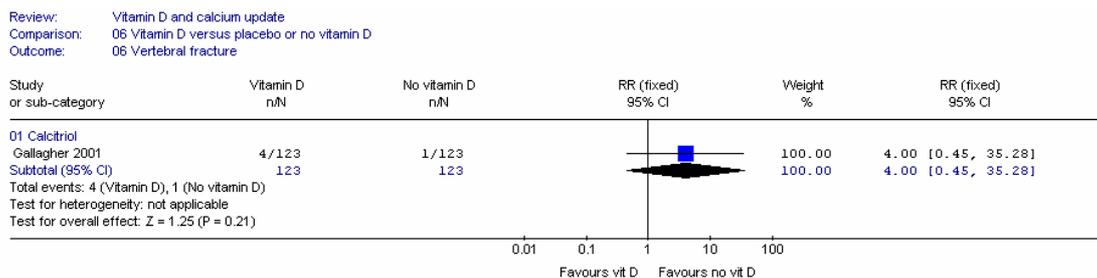
- In RECORD (2005), 42 to 53% took tablets on more than 80% of days (vitamin D plus calcium and calcium alone groups had 42%; vitamin D alone had 53% and the placebo group was at 50%).

The other studies did not report concordance.

## Hydroxylated vitamin D versus placebo or no treatment

### Vertebral fractures

Two studies evaluated vertebral fractures for the comparison of calcitriol versus placebo, but one of these, Gallagher (1989), did not give the number of participants with a fracture. The other study (Gallagher 2001) reported the numbers of participants with fractures, but the denominators were unclear, so the numbers randomised were assumed. The single study in 246 participants, gave a CI that was too wide to determine if there was a difference between interventions (figure 61).



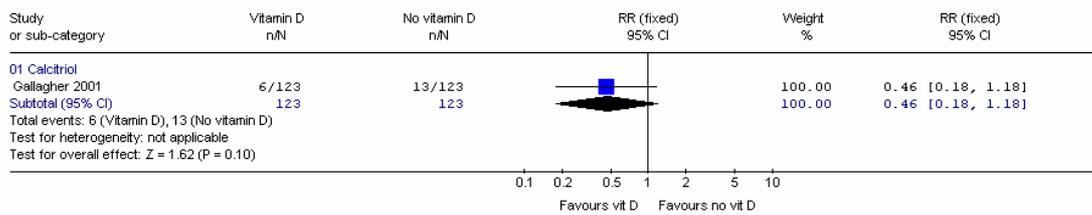
**Figure 61. Number of participants with vertebral fractures**

NB: Forest plot scale 0.01 to 100

### Nonvertebral fracture

One study (Gallagher 2001) in 246 participants evaluated nonvertebral fractures. The study reported only percentages and the number randomised was assumed. There was no significant difference between interventions (figure 62); RR 0.46 (95% CI 0.18 to 1.18). The confidence interval was fairly wide.

Review: Vitamin D and calcium update  
 Comparison: 06 Vitamin D versus placebo or no vitamin D  
 Outcome: 02 Non-vertebral fracture any site



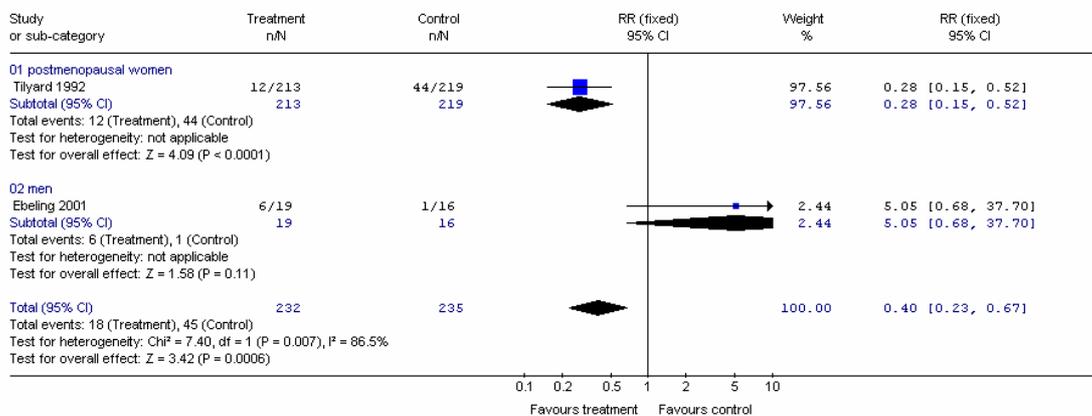
**Figure 62. Number of participants with nonvertebral fractures**

## Hydroxylated vitamin D versus calcium

### Vertebral fracture

Two studies reported the number of people with vertebral fractures for the comparison calcitriol versus calcium (Ebeling 2001; Tilyard 1992). Meta-analysis of the two studies in 467 participants (figure 63) showed significant heterogeneity ( $I^2 = 87\%$ ;  $p = 0.007$ ). Sensitivity analysis in the absence of Tilyard (1992), which had a higher potential for bias, left one study in 35 participants, showing results that were too uncertain to determine if there was a difference between interventions.

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Calcitriol vs calcium  
 Outcome: 01 Vertebral Fractures

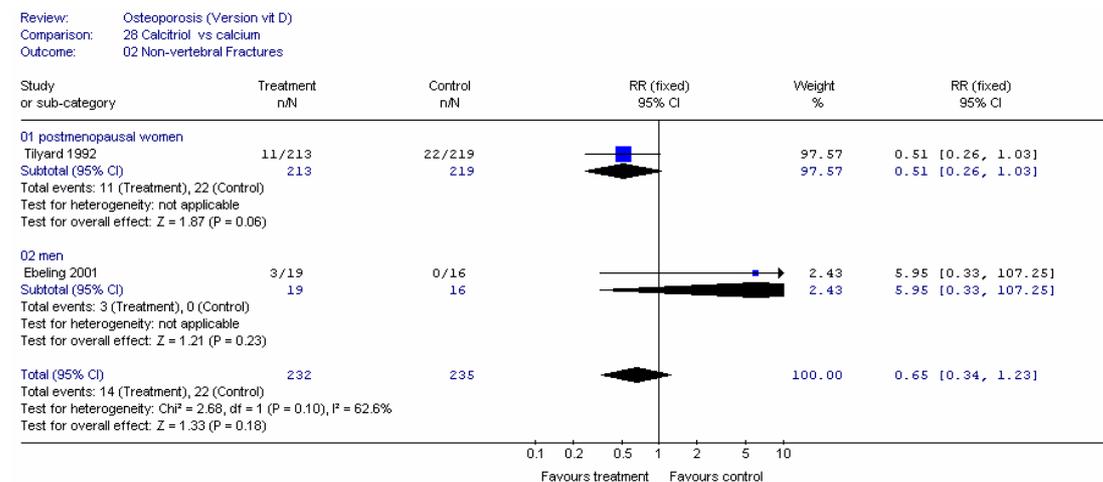


**Figure 63. Number of participants with vertebral fractures**

### Non vertebral fractures

Two studies reported the number of people with vertebral fractures for the comparison calcitriol versus calcium (Ebeling 2001; Tilyard 1992). Meta-analysis of the two studies in 467 participants (figure 64) showed some heterogeneity ( $I^2 = 63\%$ ;  $p = 0.10$ ). There was no significant difference between interventions; RR 0.65 (95% CI 0.34 to 1.23)]. A sensitivity analysis,

in the absence of Tilyard (1992) gave a wide confidence interval and conclusions were not drawn.



**Figure 64. Number of participants with nonvertebral fractures**

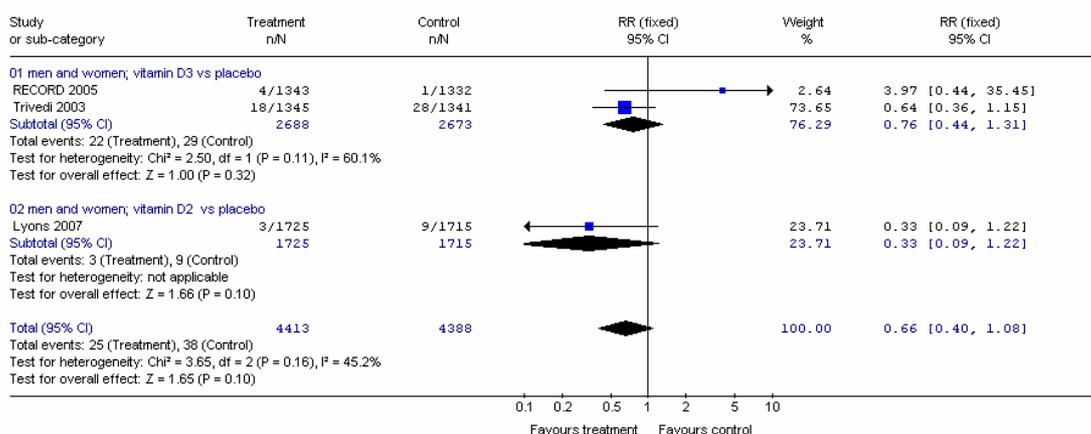
## Native vitamin D versus placebo

### *Clinical vertebral fracture*

Two studies compared vitamin D3 versus placebo (RECORD 2005; Trivedi 2003) and one vitamin D2 versus placebo (Lyons 2007). All studies reported clinical fractures. These studies were combined as subgroups. Meta-analysis of the three studies in 8801 participants showed there was no significant difference between interventions (figure 65); RR 0.66 (95%CI 0.40 to 1.08). There was no significant heterogeneity between studies ( $I^2 = 45\%$ ;  $p = 0.16$ ). In Trivedi (2003) fracture incidences were based on self report or death certification.

Sensitivity analysis in the absence of Lyons (2007), which had 47% missing data, showed no significant difference between interventions, RR 0.76 (95%CI 0.44 to 1.31), but there was significant heterogeneity ( $I^2 = 60\%$ ;  $p = 0.11$ ). It was noted that Trivedi (2003) gave the participants 100,000 IU orally every 4 months, whereas RECORD (2005) gave 800 IU daily; concordance was much higher in the former study.

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Native vitamin D vs placebo/no treatment  
 Outcome: 01 Vertebral Fractures (clinical)



**Figure 65. Number of participants with vertebral fractures**

*Nonvertebral fracture*

Eight studies (Avenell 2004; Harwood 2004; Lips 1996; Lyons 2007; Meyer 2002; RECORD 2005; Smith 2007; Trivedi 2003) reported nonvertebral fractures in 22,098 participants. In three of the studies (Harwood 2004; Lyons 2007; Smith 2007), the intervention was vitamin D2 and in the rest it was vitamin D3. Lips (1996) reported separately hip and other peripheral fractures and these were summed to give the number of nonvertebral fractures. The number of participants with nonvertebral fractures was calculated for Trivedi (2003) by subtracting the number of participants with vertebral fractures from the total number with fractures.

Meta-analysis showed there was no significant difference between interventions (figure 66); RR 1.01 (95%CI 0.94 to 1.10); there was no heterogeneity between studies ( $I^2 = 0\%$ ,  $p = 0.62$ ). The funnel plot showed asymmetry (figure 67). In the absence of the studies at higher risk of bias (Avenell 2004; Harwood 2004; Lyons 2007; Meyer 2002; Smith 2007), the three remaining studies showed similar results to the full meta-analysis (figure 68); RR 1.02 (95%CI 0.91 to 1.14), with no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.57$ ).

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Native vitamin D vs placebo/no treatment  
 Outcome: 02 Non-vertebral Fractures

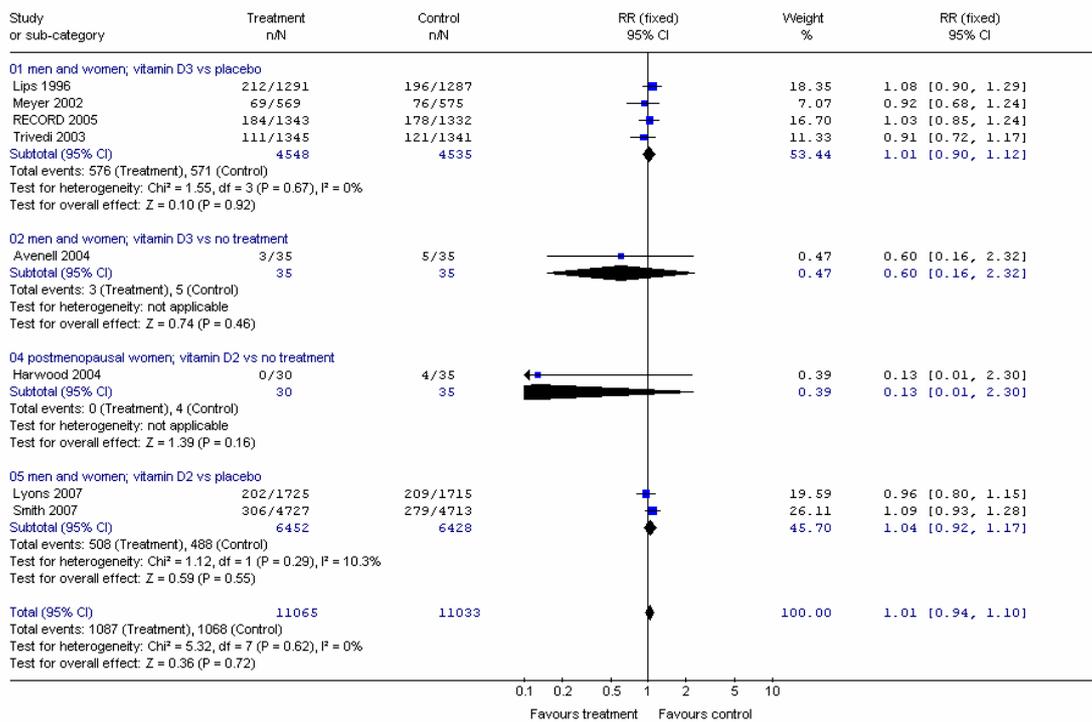


Figure 66. Number of participants with nonvertebral fractures

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Native vitamin D vs placebo/no treatment  
 Outcome: 02 Non-vertebral Fractures

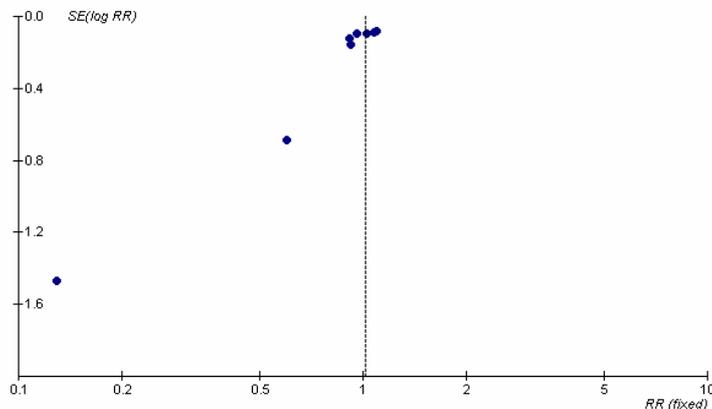
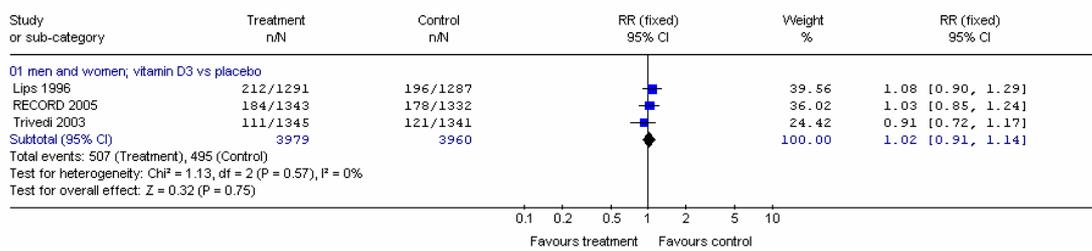


Figure 67. Funnel plot for number of participants with nonvertebral fractures

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Native vitamin D vs placebo/no treatment  
 Outcome: 07 Non-vertebral Fractures sensitivity analysis

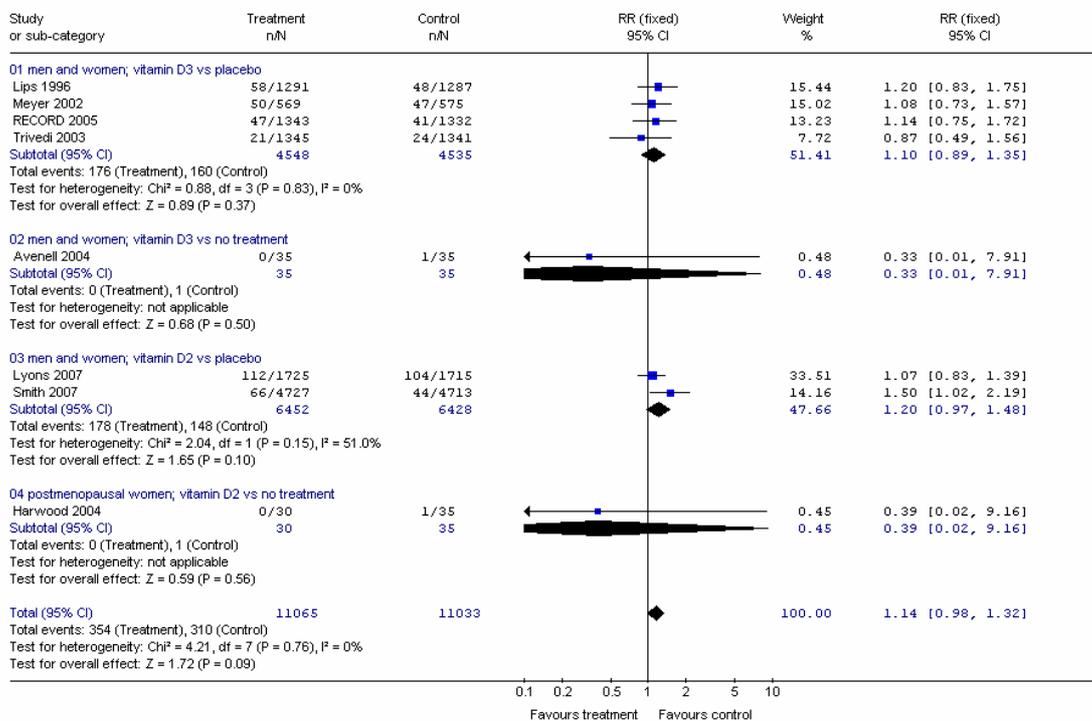


**Figure 68. Sensitivity analysis: number of participants with nonvertebral fractures**

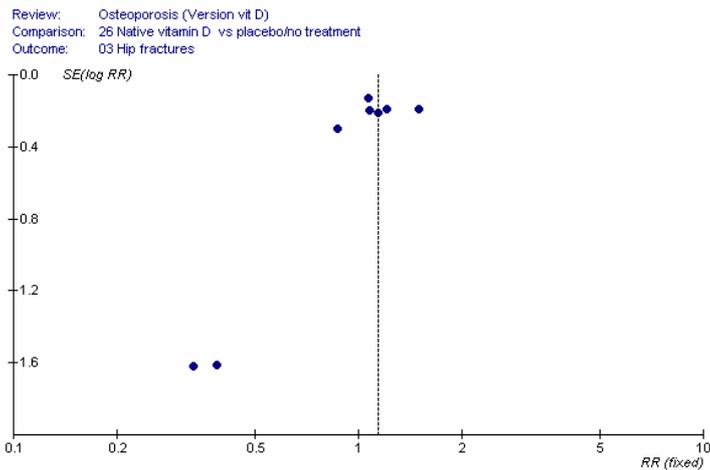
### Hip fracture

Meta-analysis of the same eight studies in 22,098 participants showed no significant difference between interventions (figure 69); RR 1.14 (95%CI 0.98 to 1.32)]. There was no heterogeneity between studies ( $I^2=0\%$ ;  $p=0.76$ ). The funnel plot was also asymmetrical (figure 70). In the absence of studies with higher potential for bias, the summary statistics were very similar to the full meta-analysis: RR 1.11 (95%CI 0.86 to 1.42), with no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.65$ ), so the full meta-analysis was preferred.

Review: Osteoporosis (Version vit D)  
 Comparison: 26 Native vitamin D vs placebo/no treatment  
 Outcome: 03 Hip fractures



**Figure 69. Number of participants with hip fractures**

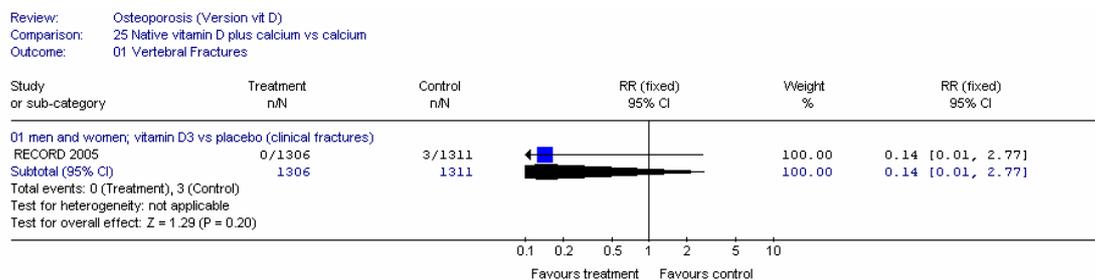


**Figure 70. Funnel plot for number of participants with hip fractures**

**Native vitamin D plus calcium versus calcium**

*Vertebral fractures*

One study (RECORD 2005) in 2617 participants recorded the number of people with new clinical vertebral fractures (figure 71). The confidence interval was too wide to determine if there was a difference between interventions.

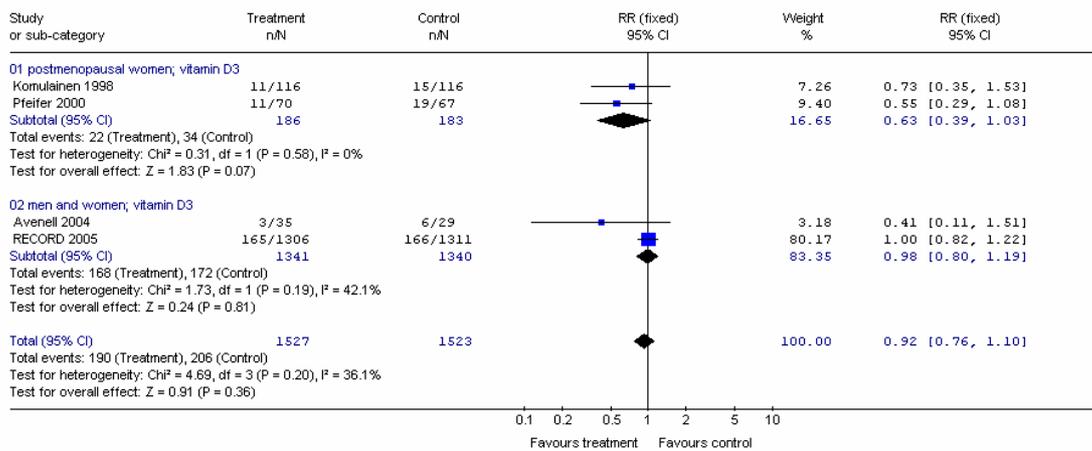


**Figure 71. Number of participants with vertebral fractures**

*Nonvertebral fracture*

Meta-analysis of four studies (Avenell 2004; Komulainen 1998; Pfeifer 2000; RECORD 2005) in 3050 participants showed there was no significant difference in the number of participants with nonvertebral fractures (figure 72); RR 0.92 (95%CI 0.76 to 1.10), with no significant heterogeneity between studies ( $I^2 = 36\%$ ;  $p = 0.20$ ). A sensitivity analysis in the absence of the studies at higher risk of bias showed a similar result, with increased heterogeneity ( $I^2 = 64\%$ ;  $p = 0.10$ ).

Review: Osteoporosis (Version vitamin D)  
 Comparison: 25 Native vitamin D plus calcium vs calcium  
 Outcome: 02 Non-vertebral Fractures

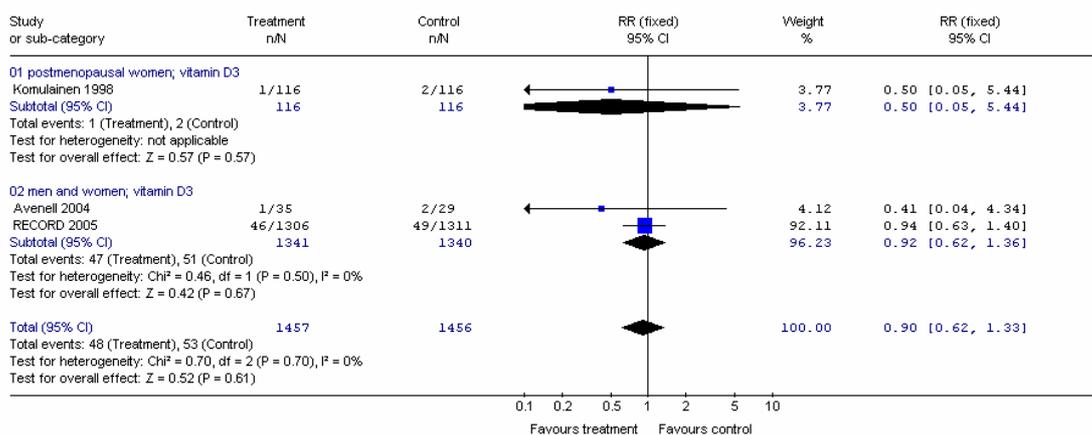


**Figure 72. Number of participants with nonvertebral fractures**

### Hip fractures

Three studies (Avenell 2004; Komulainen 1998; RECORD 2005) in 3103 participants reported the number with hip fractures (figure 73). Meta-analysis showed there was no significant difference between interventions; RR 0.90 (95%CI 0.62 to 1.33); there was no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.70$ ). In the absence of the studies at higher risk of bias (Avenell 2004 and Komulainen 1998), a very similar result was obtained.

Review: Osteoporosis (Version vit D)  
 Comparison: 25 Native vitamin D plus calcium vs calcium  
 Outcome: 03 Hip fractures

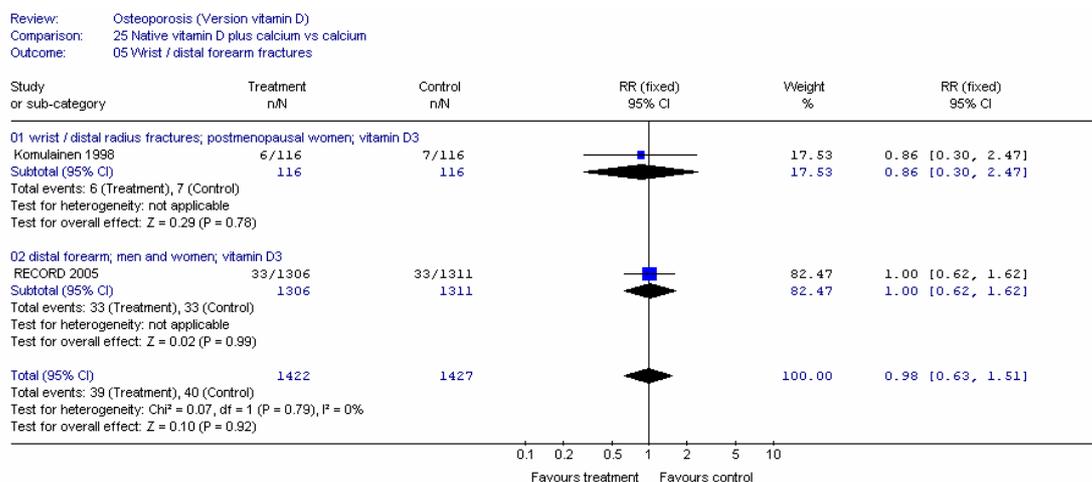


**Figure 73. Number of participants with hip fractures**

### Wrist/distal radius fractures

Two studies (Komulainen 1996; RECORD 2005) in 2849 participants reported the number with wrist or distal radius fractures (figure 74). There was no significant difference between interventions; RR 0.98 (95%CI 0.63 to 1.51)

and no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.79$ ). In the absence of the study at higher risk of bias (Komulainen 1998), a very similar result was obtained.

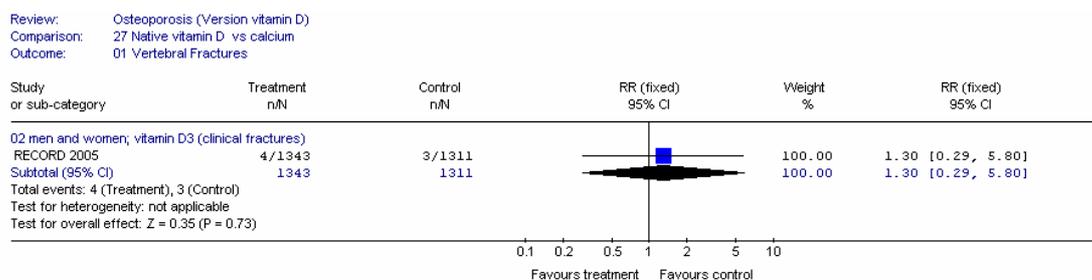


**Figure 74. Number of participants with wrist and distal radius fractures**

## Native vitamin D versus calcium

### Vertebral fractures

One study (RECORD 2005) in 2654 participants recorded the number of people with new clinical vertebral fractures (figure 75). The confidence interval was too wide to determine if there was a difference between interventions.

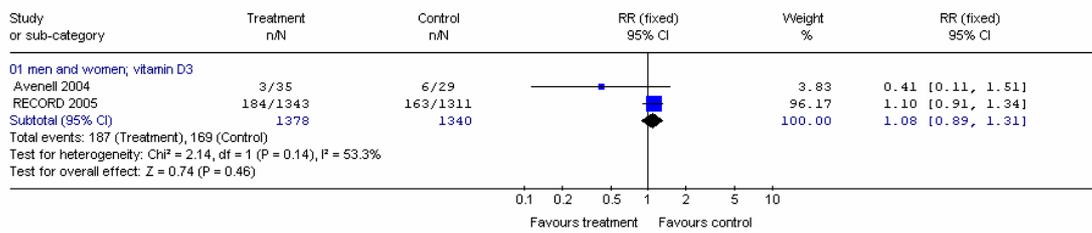


**Figure 75. Number of participants with vertebral fractures**

### Nonvertebral fracture

Meta-analysis of two studies (Avenell 2004; RECORD 2005) in 2718 participants showed there was no significant difference in the number of participants with nonvertebral fractures (figure 76); RR 1.08 (95%CI 0.89 to 1.31), with some heterogeneity between studies ( $I^2 = 53\%$ ;  $p = 0.14$ ). It is noted that participants in the Avenell (2004) study were not blinded and there was differential missing data.

Review: Osteoporosis (Version vitamin D)  
 Comparison: 27 Native vitamin D vs calcium  
 Outcome: 02 Non-vertebral Fractures

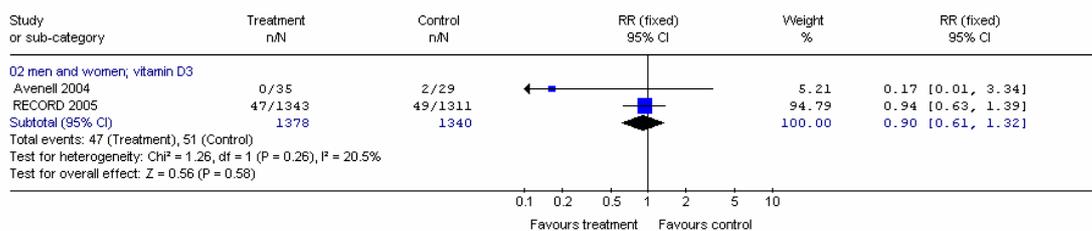


**Figure 76. Number of participants with nonvertebral fractures**

### Hip fractures

Two studies (Avenell 2004; RECORD 2005) in 2718 participants reported the number with hip fractures (figure 77). Meta-analysis showed there was no significant difference between interventions; RR 0.90 (95%CI 0.61 to 1.32); there was no significant heterogeneity ( $I^2 = 21\%$ ;  $p = 0.26$ ).

Review: Osteoporosis (Version vitamin D)  
 Comparison: 27 Native vitamin D vs calcium  
 Outcome: 03 Hip fractures

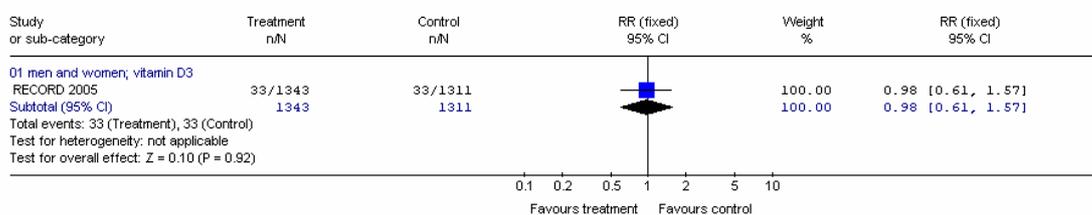


**Figure 77. Number of participants with hip fractures**

### Distal forearm fractures

One study (RECORD 2005) in 2654 participants reported the number with distal forearm fractures (figure 78). There was no significant difference between interventions; RR 0.98 (95%CI 0.61 to 1.57).

Review: Osteoporosis (Version vitamin D)  
 Comparison: 27 Native vitamin D vs calcium  
 Outcome: 05 Distal forearm fractures



**Figure 78. Number of participants with distal forearm fractures**

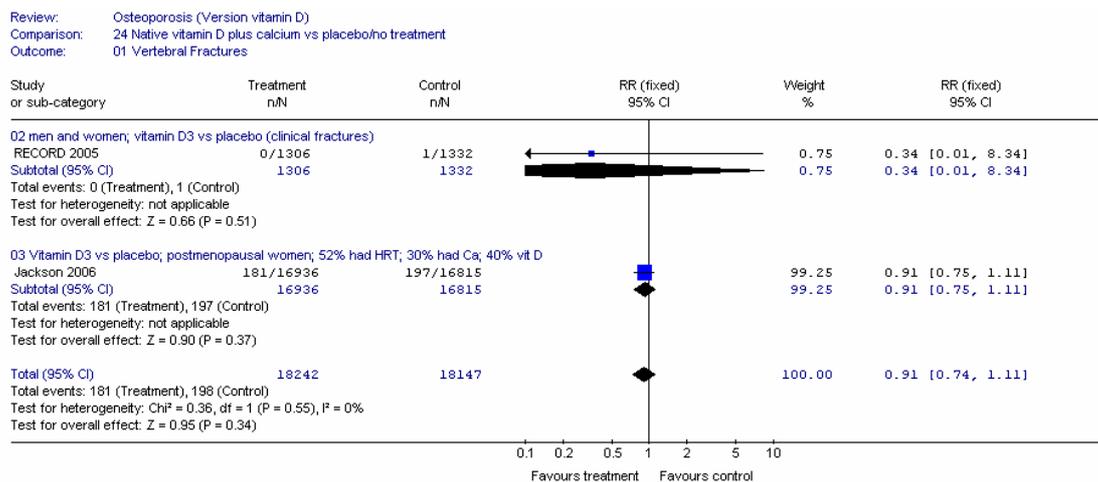
## Vitamin D plus calcium versus placebo or no treatment

Nine studies with 10 comparisons investigated the effect of combined vitamin D and calcium versus placebo or no treatment. One arm of Harwood (2004) gave the participants vitamin D2 and the rest of the studies used vitamin D3.

### Vertebral fracture

In Jackson (2006) it was unclear which denominators were used, but it was assumed to be the number who were alive and had outcomes data submitted in the last 18 months. 52% of the women in each group received HRT and 30% in each group had additional supplementary calcium and 40% in each group had additional vitamin D above 400 IU/day. 17% of participants took osteoporosis drugs during the trial period, but it was unclear which groups had these drugs, therefore it is possible that this trial was confounded.

Meta-analysis of two studies (Jackson 2006; RECORD 2005) in 36,389 participants showed there was no significant difference in the number of participants with vertebral fractures (figure 79); RR 0.91 (95%CI 0.74 to 1.11), with no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.55$ ). In the absence of the Jackson study, the confidence interval was too wide to determine if there was a difference between interventions.



**Figure 79. Number of participants with vertebral fractures**

### Nonvertebral fracture

Seven studies with 8 comparisons investigated nonvertebral fractures (Avenell 2004; Chapuy 1992, Chapuy 2002; Dawson-Hughes 1997, Harwood 2004 (2 comparisons); Jackson 2006; RECORD 2005). The Jackson (2006) study

allowed the women to have personal calcium and vitamin D supplements and 52% were on HRT; during the trial osteoporosis medication was allowed and 17% received this. The number of nonvertebral fractures for this study was calculated by subtracting the number of vertebral fractures from the total number. Chapuy (2002) stated that the denominators were for the population that were assessed at least once. Dawson-Hughes (1997) reported percentages for the cumulative incidence of a first fracture (osteoporotic 76% and trauma) and gave both types of fracture by site. The latter study also gave the relative risk for osteoporotic fractures (RR 0.4; 95%CI 0.2 to 0.8), but not the number of people with an osteoporotic fracture. For the Harwood (2004) study, the Avenell Cochrane review reported (via author communication) the number of participants with hip fractures; these were reported in the number of people who had a fall, so fractures were assumed to be osteoporotic. RECORD (2005) reported the number of osteoporotic fractures, which included the number of clinical vertebral fractures. We therefore calculated the number of nonvertebral fractures by subtraction.

Meta-analysis of these studies in 40,310 participants showed a borderline significant difference between interventions, favouring vitamin D plus calcium (figure 80); RR 0.95 (95%CI 0.90 to 1.00);  $p = 0.04$ . There was some heterogeneity ( $I^2 = 40\%$ ,  $p = 0.11$ ). A funnel plot (figure 81) showed some asymmetry. Analysis using a random effects model gave a relative risk of 0.88 (95%CI 0.77 to 1.01).

In a sensitivity analysis in the absence of the studies at higher risk of bias (Avenell 2004; Chapuy 1992; Chapuy 2002; Dawson Hughes 1997; Harwood 2004) and Jackson (2006) which may have been confounded, left the RECORD (2005) study, which showed similar summary statistics to the main meta-analysis.

Review: Osteoporosis (Version vitamin D)  
 Comparison: 24 Native vitamin D plus calcium vs placebo/no treatment  
 Outcome: 02 Non-vertebral Fractures

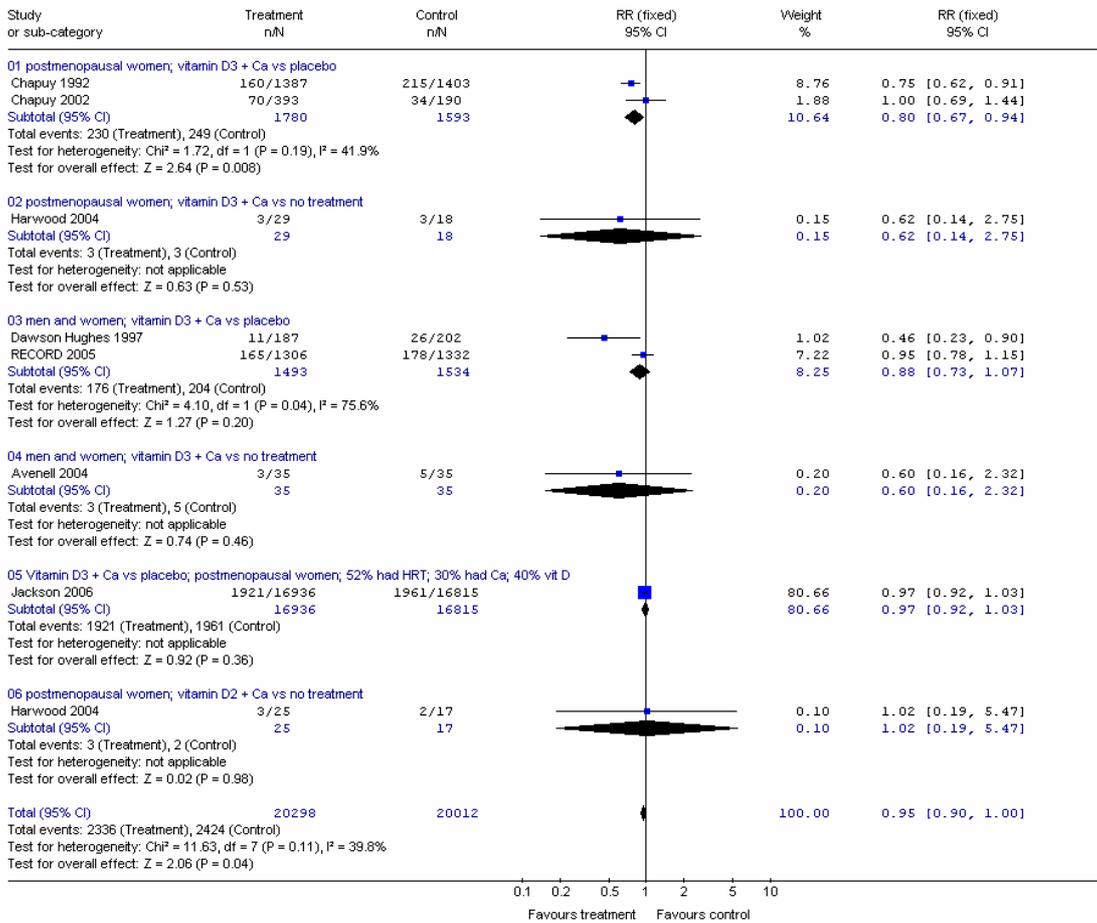


Figure 80. Number of participants with nonvertebral fractures

Review: Osteoporosis (Version vitamin D)  
 Comparison: 24 Native vitamin D plus calcium vs placebo/no treatment  
 Outcome: 02 Non-vertebral Fractures

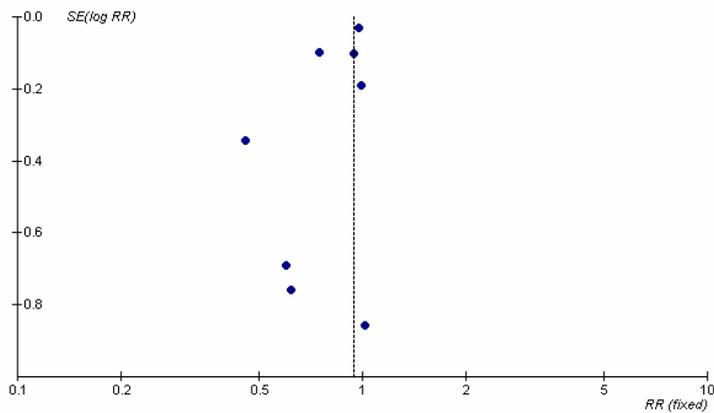
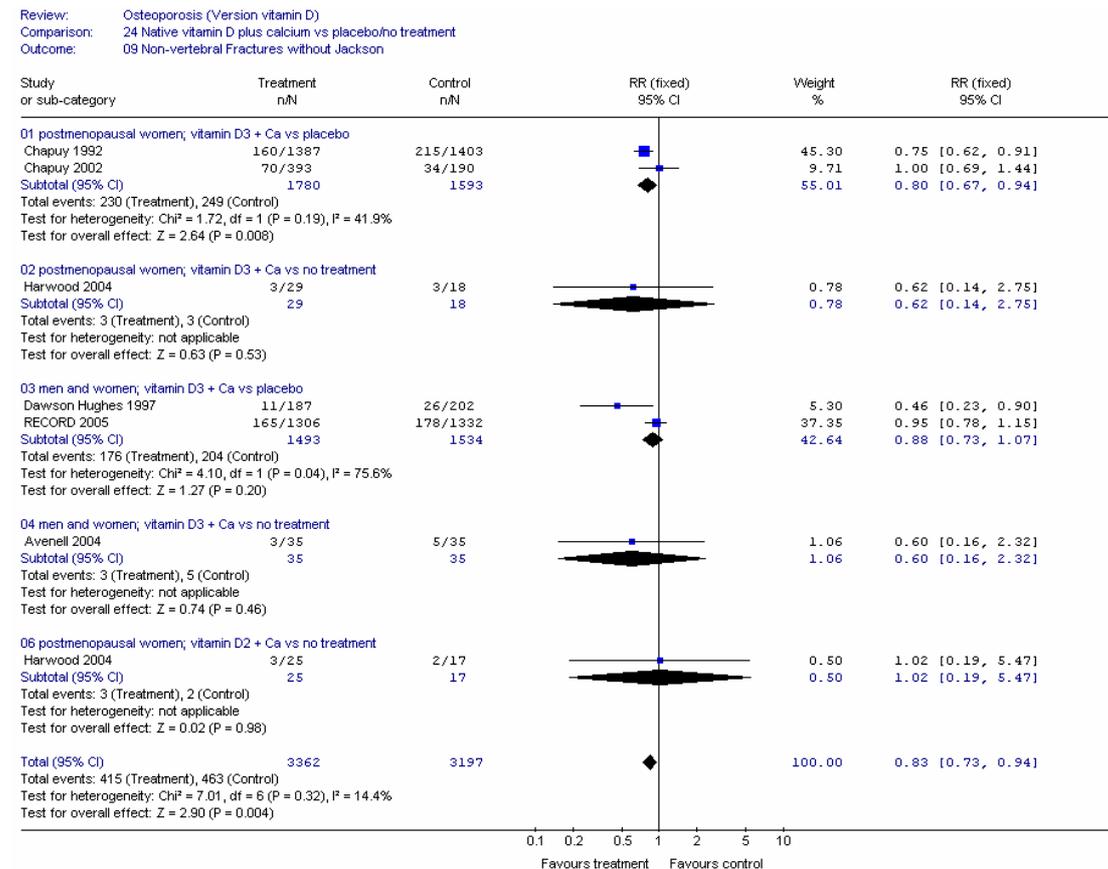
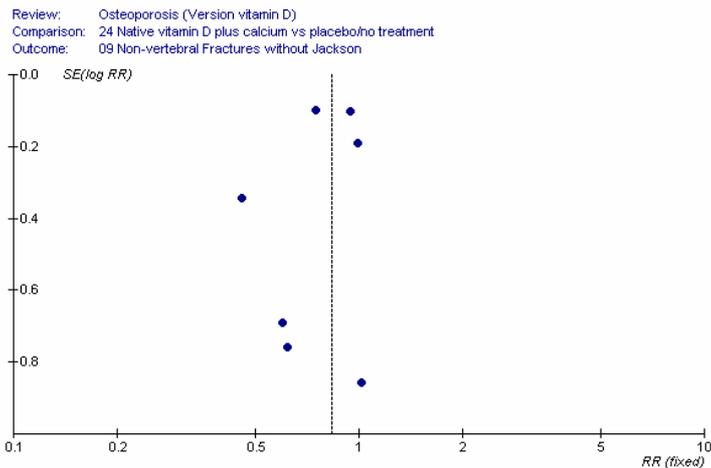


Figure 81. Funnel plot for number of participants with nonvertebral fractures

In the absence of the Jackson (2006) study, which was partly a different comparison (Vitamin D plus calcium plus HRT versus placebo plus HRT), there were significantly fewer participants with nonvertebral fractures (figure 82) in the vitamin D plus calcium group; RR 0.83 (95%CI 0.73 to 0.94); there was less heterogeneity ( $I^2 = 14\%$ ,  $p = 0.32$ ) and the funnel plot was approximately symmetrical (figure 83). This corresponds to a number needed to treat of 50 (95%CI 25 to 100) for a control group rate of 11% to 18%.



**Figure 82. Sensitivity analysis: number of participants with nonvertebral fractures (without Jackson 2006)**



**Figure 83. Funnel plot for number of participants with nonvertebral fractures without Jackson (2006)**

*Hip fracture*

Eight studies with 9 comparisons (Avenell 2004; Chapuy 1992, Chapuy 2002; Dawson-Hughes 1997, Harwood 2004 (2 comparisons); Jackson 2006; Porthouse 2005; RECORD 2005) evaluated the effect of vitamin D plus calcium on hip fracture incidence. The Avenell Cochrane review reported for Harwood (2004) (via author communication) the number of participants having hip fractures; fractures were reported in the number of people who had a fall, so these were assumed to be osteoporotic. Jackson (2006) allowed the women to have personal calcium and vitamin D supplements and 52% were on HRT; during the trial osteoporosis medication was allowed and 17% received this. The denominators were assumed to be the per protocol numbers.

Meta-analysis of these studies in 43,623 participants showed a small significant difference between interventions, favouring the vitamin D and calcium combination (figure 84); RR 0.83 (95%CI 0.72 to 0.96). There was no heterogeneity between studies ( $I^2 = 0\%$ ,  $p = 0.57$ ). A funnel plot (figure 85) showed a little asymmetry. The corresponding number needed to treat is  $\infty$ .

Review: Osteoporosis (Version vitamin D)  
 Comparison: 24 Native vitamin D plus calcium vs placebo/no treatment  
 Outcome: 03 Hip fractures

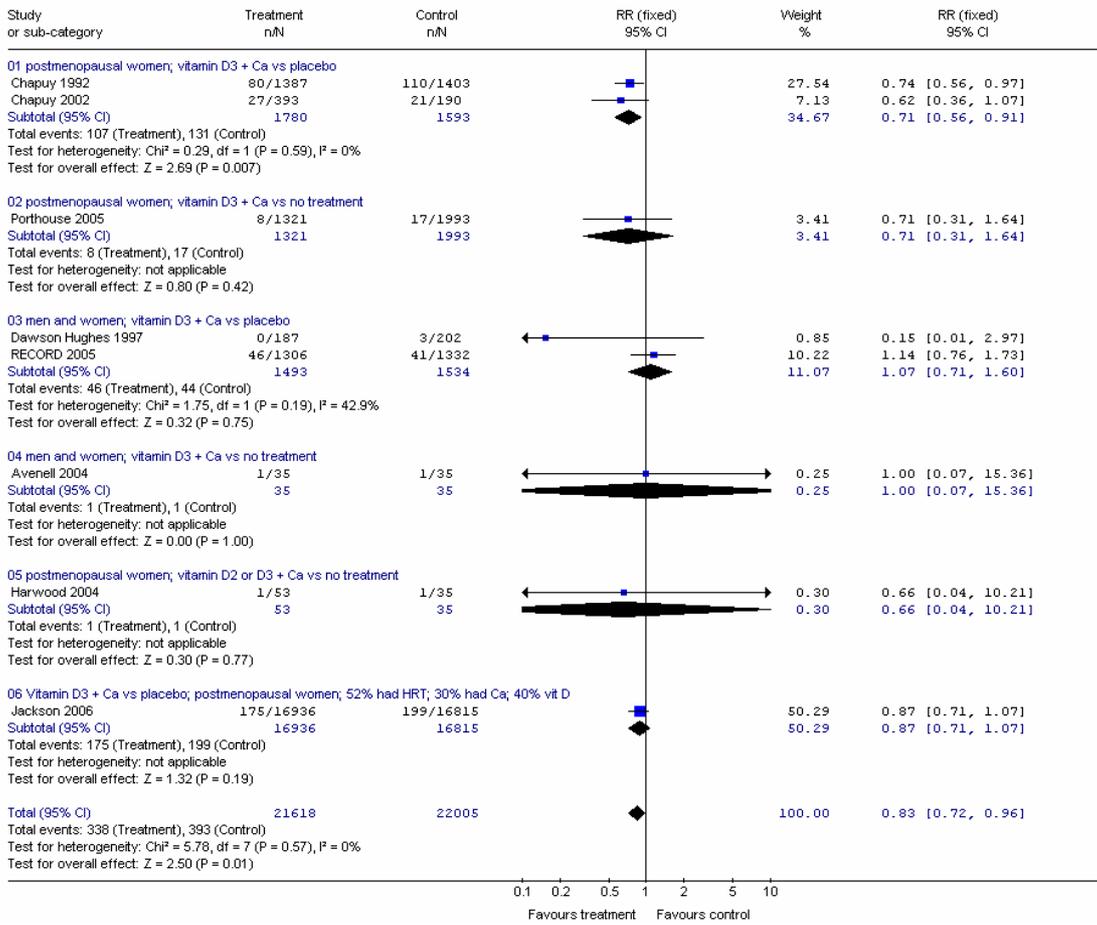


Figure 84. Number of participants with hip fractures

Review: Osteoporosis (Version vitamin D)  
 Comparison: 24 Native vitamin D plus calcium vs placebo/no treatment  
 Outcome: 03 Hip fractures

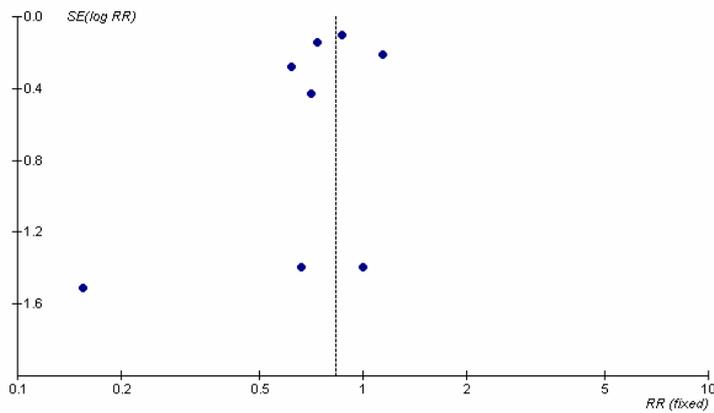
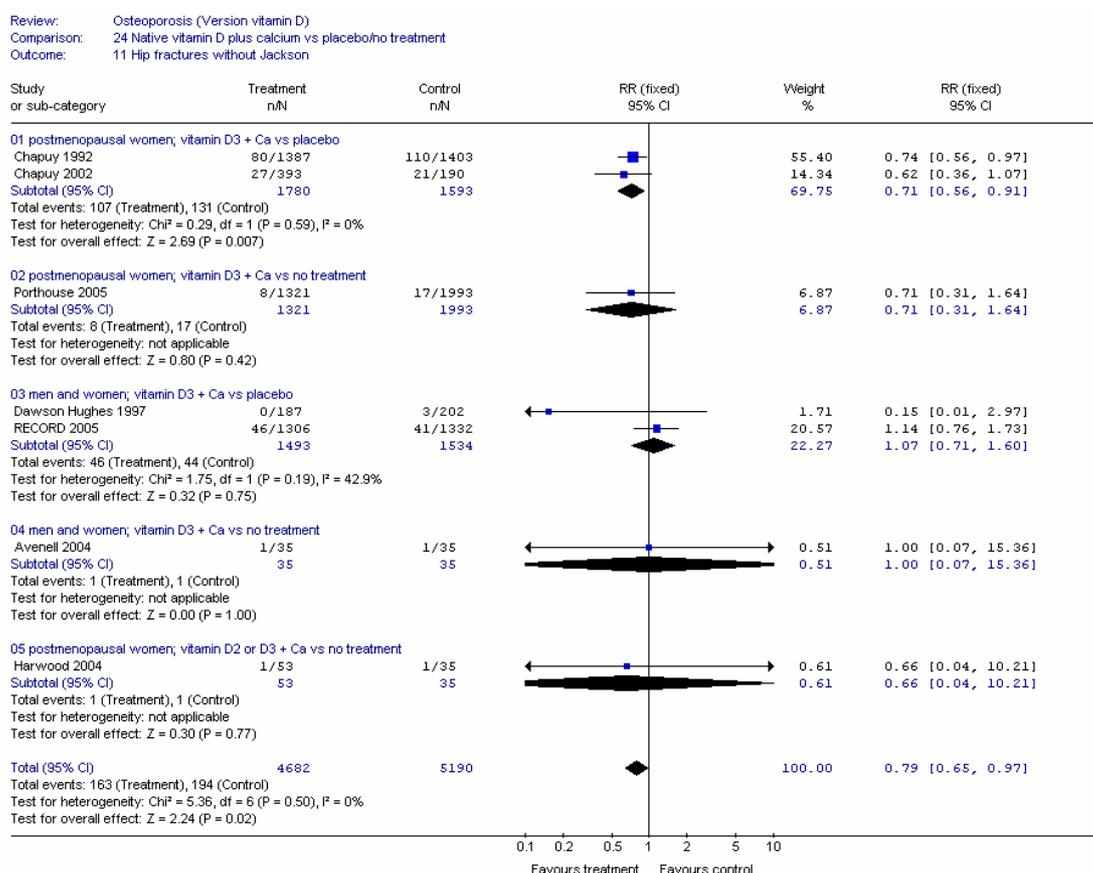


Figure 85. Funnel plot for number of participants with hip fractures

Sensitivity analysis in the absence of the studies at higher risk of bias left the RECORD (2005) study, which showed no significant difference between interventions; RR 1.14 (95% CI 0.76 to 1.73). It is noted that this study had low concordance levels (43%), which may have diluted any effect.

In a second sensitivity analysis in the absence of Jackson (2006), meta-analysis showed significantly fewer people with fractures in the vitamin D plus calcium group (figure 86), RR 0.79 (95%CI 0.65 to 0.97). The funnel plot was similar. This corresponds to a number needed to treat of 100 (95%CI 50, ∞).

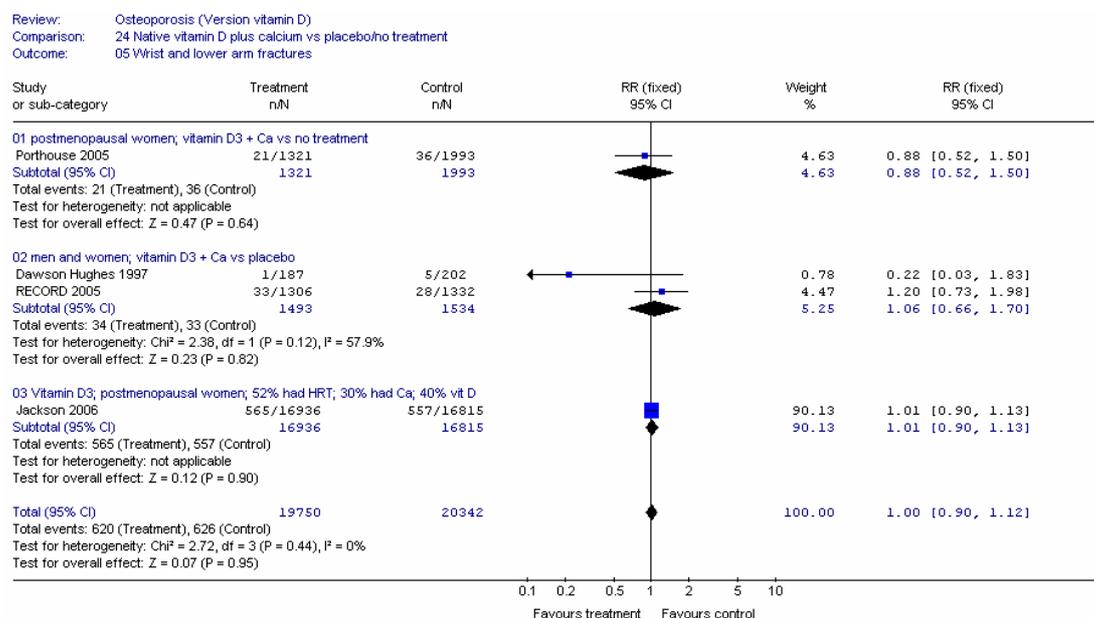


**Figure 86. Sensitivity analysis: number of participants with hip fractures (without Jackson 2006)**

*Lower arm or wrist fracture*

Four studies reported wrist or lower arm fractures (Dawson Hughes 1997; Jackson 2006; Porthouse 2006; RECORD 2005). Meta-analysis in 40,092 participants showed there was no significant difference between interventions (figure 87); RR 1.00 (95%CI 0.90 to 1.12), with no heterogeneity between studies (I<sup>2</sup> = 0%, p = 0.44). A sensitivity analysis in the absence of the two

studies which had higher potential for bias (Dawson Hughes 1997; Porthouse 2006) and Jackson (2006), which may have been confounded and was partly measuring the additional effect in people given HRT, left RECORD (2005), which had summary statistics of RR 1.20 (95%CI 0.73 to 1.98).

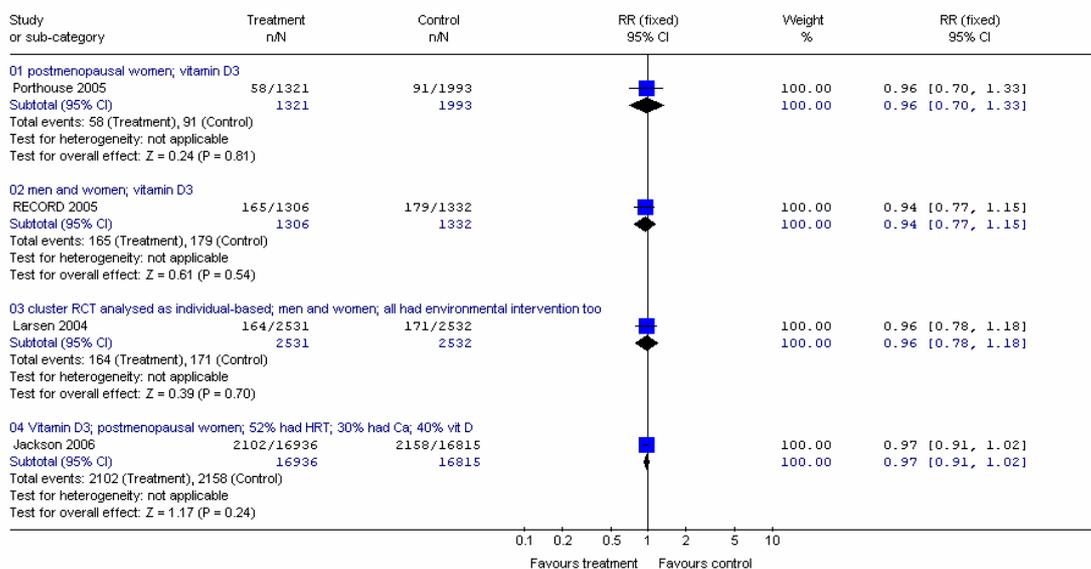


**Figure 87. Number of participants with wrist and lower arm fractures**

### All clinical fractures

Four studies reported the outcome, all clinical fractures (Jackson 2006; Larsen 2004; Porthouse 2005; RECORD 2005). Larsen (2004) was a cluster randomised study in which 4 geographical regions were randomised to 4 interventions. The study compared vitamin D plus calcium versus no intervention in participants who all were offered an environmental /education intervention as well. This trial did not take account of intra-cluster correlation, but analysed the data as if it were in individual participants. Its results are shown in Figure 88. There was no significant difference between interventions in any of the studies. A sensitivity analysis in the absence of the studies more likely to be at risk of bias left RECORD (2005), which showed no significant difference between interventions and similar summary statistics to the main meta-analysis.

Review: Osteoporosis (Version vitamin D)  
 Comparison: 24 Native vitamin D plus calcium vs placebo/no treatment  
 Outcome: 06 All Fractures



**Figure 88. Number of participants with all clinical fractures**

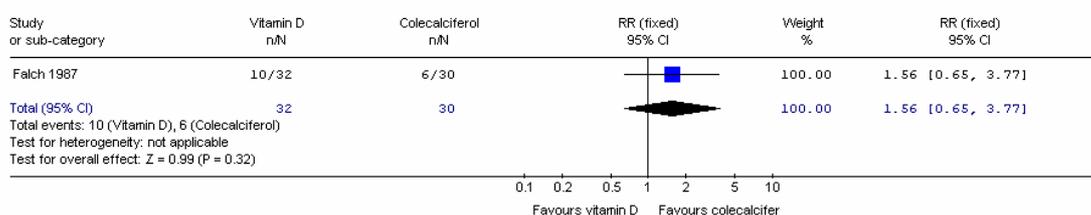
**Comparison of different types of vitamin D**

**Calcitriol versus colecalciferol**

*Vertebral fracture*

One study (Falch 1987) in 62 participants reported vertebral fractures, comparing calcitriol versus colecalciferol. The confidence interval was fairly wide and showed no significant difference between interventions (figure 89); RR 1.56 (95%CI 0.65 to 3.77). This study should be treated with caution because of differential drop-outs.

Review: Vitamin D and calcium update  
 Comparison: 09 Vitamin D versus colecalciferol  
 Outcome: 01 Vertebral fractures

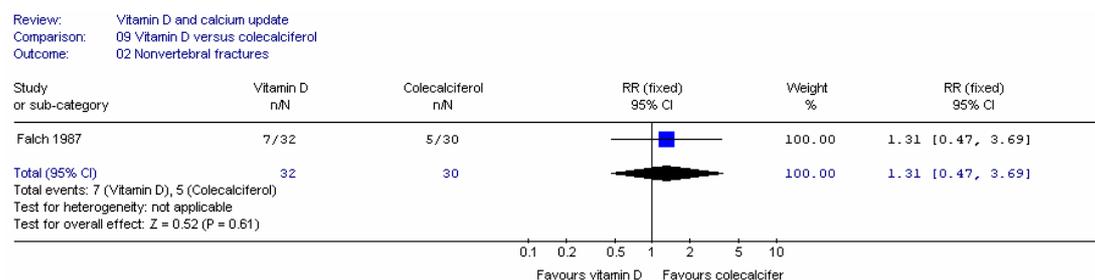


**Figure 89. Number of participants with vertebral fractures**

*Nonvertebral fracture*

One study (Falch 1987) with 62 participants reported the number of participants with nonvertebral fractures at any site. The confidence interval is fairly wide and there was no significant difference between interventions

(figure 90); RR 1.31 (95%CI 0.47 to 3.69). This study should be treated with caution because of differential drop-outs.



**Figure 90. Number of participants with nonvertebral fractures**

## 2.5.2 Calcium

### *Description of studies*

A total of 14 papers were evaluated for inclusion. One study (Stellon 1985) was excluded because the calcium formulation was not in the BNF, however we decided to include studies using calcium formulations in the BNF, even if the incorrect doses were used because calcium is not listed as an indication for osteoporosis. Two other studies were excluded for reasons given in appendix IV. Eleven reports were included (Avenell 2004; Chevalley 1994; Hansson 1987; Peacock 2000; Prince 2007; Recker 1996; RECORD 2005; Reid 1993; 1995; 2006; Riggs 1998), describing ten trials.

Reid (1995) was a 2-year extension study of Reid (1993), in which 86/135 (64%) of patients agreed to continue their randomised treatments. Although the later study stated that none of the patient characteristics in the extension study differed significantly from those of the original study, we decided that there was too much missing data to consider Reid (1995) further.

### *Study design*

Two studies were conducted in the UK (Avenell 2004; RECORD 2005). One study was carried out in Switzerland (Chevalley 1994), one in Sweden (Hansson 1987), three in the USA (Peacock 2000; Recker 1996; Riggs 1998), one in Australia (Prince 2007), and two in New Zealand (Reid 1993; 2006).

One study received funding from industry: Chevalley (1994) received grant support from Robapharm AG. Recker (1996) received a grant from the

National Dairy Promotion and Research Board. Two studies (Avenell 2004; RECORD 2005) received some support from Shire Pharmaceuticals, but their representatives did not take part in the analyses; some of the authors received consultancy fees. Three studies received grants from public bodies (Hansson 1987; Prince 2007; Riggs 1998). Three studies (Peacock 2000; Reid 1993; 2006) did not provide details of funding.

One study had fewer than 100 patients overall (Chevalley 1994, n = 93); four studies had between 100 and 200 patients (Hansson 1987, n = 100; Reid 1993, n = 122; Avenell 2004 n=134; Recker 1996, n = 197); Riggs (1998) had 236 women and Peacock (2000) included 438 men and women. The other studies had more than 1000 patients (Prince (2007), 1460 women; RECORD (2005) 5292 women (2643 in calcium part); Reid (2006), 1471 women). Hansson (1987) had only 25 patients in each group.

### ***Population***

Six studies included only postmenopausal women (Hansson 1987; Prince 2007; Recker 1996; Reid 1993; 2006; Riggs 1998). The mean time since the menopause ranged from 9 to 26 years. Recker (1996) did not state how long the women had been postmenopausal. Avenell 2004, Chevalley (1994), Peacock (2000) and RECORD (2005) included men and postmenopausal women.

Mean baseline T-scores at the femoral neck, proximal femur or total hip were reported (or calculated) for five studies; they were not reported in Avenell (2004), Recker (1996) or RECORD (2005) and no standard deviations were given in Riggs (1998). Hansson (1987) stated that the women had osteoporosis, but no BMD details were given. On this basis, one study (Hansson 1987) was assessed as being in patients with osteoporosis only; two studies were assessed to include patients with osteoporosis or osteopenia (Prince 2007; Reid 2006); one study (Reid 1993) was assessed to include patients with osteopenia or normal BMD; and two studies (Chevalley 1994; Peacock 2000) were considered to cover all baseline BMDs.

Two studies (Peacock 2000; Reid 1993) did not report whether any of the patients had vertebral fractures at baseline. Riggs (1998) reported that none of the patients had fractures at baseline and all patients had at least one fracture in Avenell (2004), Hansson (1987) and RECORD (2005). Of the four remaining studies, the percentage of patients with at least one vertebral fracture at baseline varied from 28% (Reid 2006) to 59% (Chevalley 1994; Recker 1996). Prince (2007) reported that 25–32% of the patients had had prevalent fractures since age 50 years.

The mean age, when given, ranged across studies from 58 (Reid 1993) to 77 years (RECORD 2005). Five studies had lower age limits for their inclusion criteria (Reid 2006: over 55 years; Peacock 2000 and Recker 1996: over 60 years; Avenell 2004, Prince 2007 and RECORD 2005: over 70 years).

Peacock (2000) stated that participants were white, RECORD (2005) stated that 99% were white and Recker (1996) reported that the women were white of European ancestry. None of the other studies reported ethnicity data. Prince (2007) and Reid (1993) reported that some of the women were smokers. Chevalley (1994) stated that less than 5% of participants were smokers and none of the other studies reported on comorbidities or smoking history.

Seven studies reported the dietary calcium intake for the calcium and placebo groups as follows: Recker (1996): estimated at 386 and 442 mg/day; Chevalley (1994): 619 mg/day; Peacock (2000): 670 and 629 mg/day; Riggs (1998): 711 and 717 mg/day; Reid (1993): 750 mg/day; Reid (2006): 861 and 853 mg/day; Prince (2007): 915 and 897 mg/day. Avenell (2004), Hansson (1987) and RECORD (2005) did not report on dietary intakes of calcium. For this age group, the recommended calcium level is 1200 mg/day, so in all studies patients had inadequate levels (National Academy of Sciences Food and Nutrition Board [NFB] recommendations).

## ***Interventions***

The interventions included the following:

- calcium citrate, 750 mg elemental calcium per day (Peacock 2000)
- calcium citrate, 1 g elemental calcium per day (Reid 2006)
- calcium citrate, 1.6 g elemental calcium per day, which was decreased if calcium levels were too high (Riggs 1998)
- calcium carbonate, 800 mg elemental calcium per day (Chevalley 1994)
- calcium carbonate, 1 g calcium per day (Avenell 2004; RECORD 2005)
- calcium carbonate, 1.2 g elemental calcium per day (Prince 2007; Recker 1998)
- calcium (as bicarbonate, lactate and gluconate), 1 g elemental calcium per day (Hansson 1987)
- calcium lactate-gluconate (5.24 g) and calcium carbonate (0.8 g), 1 g elemental calcium per day (Reid 1993).

Chevalley (1994) also randomised patients to calcium as an osseino-mineral complex: 800 mg/day, which is not a formulation in the BNF and results were reported for these patients combined with those of the calcium carbonate group.

Of these interventions, only Avenell (2004), Reid (1993) and RECORD (2005) corresponded to the type and dose in the BNF. One study had the licensed formulation and a dose between the licensed doses (Chevalley 1994); one had a similar formulation and the licensed dose (Hansson 1987); one had the licensed dose of elemental calcium but a different formulation (Reid 2006); two had slightly above the licensed dose (which corresponded to the recommended daily intake) (Prince 2007; Recker 1998); and one had a calcium dose significantly higher than the licensed dose (Riggs 1998). It was decided to include all the studies except Riggs (1998) and to treat Chevalley (1994), Prince (2007) and Recker (1998) with caution.

## ***Comparisons***

The following comparisons were carried out:

- calcium versus placebo: nine studies
  - 18 months intervention period (Chevalley 1994)
  - 2 years (Reid 1993)
  - 2 to 5 years (RECORD 2005)
  - 3 years (Hansson 1987)
  - 4 years (Recker 1996; Riggs 1998; Peacock 2000)
  - 5 years (Prince 2007; Reid 2006).
- Calcium versus no treatment for up to 46 months: one study (Avenell 2004)

None of the studies reported the use of additional calcium supplements and only one study (Chevalley 1994) gave the patients vitamin D. In this study, all the participants were given a single oral dose of vitamin D 300,000 IU at the beginning of the study.

One study (Reid 2006) reported that some patients (it was unclear how many) were eliminated in the per protocol analysis because they took osteoporosis medication during the study. No further details were given, but there may have been confounding of results. The other studies did not report whether medications were allowed during the study, although two studies stated that bisphosphonate use was an exclusion criterion (Peacock 2000; Riggs 1998).

## ***Methodological quality***

The method of sequence generation was adequate in four studies in which a computer-generated sequence was used (Avenell 2004; Peacock 2000; RECORD 2005; Reid 2006); Peacock (2000) stratified participants into 16 strata based on age, gender, serum vitamin D concentrations, and dietary calcium before randomisation. None of the other studies clearly reported a method of sequence generation.

Four studies reported on the method of allocation concealment: this was adequate in Prince (2007), in which the pharmacy had the randomisation list and assigned the medications; and adequate in two (Avenell 2004; RECORD

2005), in which a centralised facility was used; allocation concealment was partially adequate in Riggs (1998), which stated that the study medications were dispensed by a research pharmacist.

Seven studies described the use of double-blind procedures (Chevalley 1994; Peacock 2000; Prince 2007; Recker 1996; RECORD 2005; Reid 1993; Riggs 1998). The patients were not blinded in Avenell (2004). Five studies (Avenell 2004; Recker 1996; RECORD 2005; Reid 1993; Riggs 1998) further stated that the assessors were blind to treatment allocation. The other two studies (Hansson 1987; Reid 2006) did not state double blinding, but it is likely that patient blinding occurred because of the use of placebos.

Sample size calculations were performed by three studies: Prince (2007) reported that 737 people per group were required in order to detect differences in fracture rate with an 80% power. RECORD (2005) reported that 4,200 participants were required to have 80% power ( $p < 0.05$ ) to detect a decrease in incidence of fractures from 15% in controls to 12%; 5,292 people were recruited. The study by Reid (2006) was powered to detect a 40% decrease in fracture rate. None of the other studies reported the use of power calculations.

All studies but one reported comparability of baseline characteristics: Recker (1996) reported that the patients were not comparable on the proportion of fractures, but were comparable on age, bone mineral content and calcium intake. Riggs (1998), Chevalley (1994) and Hansson (1987) only had limited baseline data. Hansson (1987) was comparable on age and bone mineral content. Participants in Riggs (1998) were comparable on age, number of years since menopause and dietary calcium intake. Chevalley 1994 reported that participants were comparable on age, BMI and the number of years since menopause. Peacock (2000) reported that the men and women were comparable on age, weight, height and total BMD. Prince (2007) reported comparability on age, weight, height, years since menopause, T-scores, smoking and activity levels. The groups in RECORD (2005) were comparable on age, gender, race, type of fracture, time since fracture, weight, smoking, physical activity, thyroxine, steroids and thiazide diuretics. Reid (1993)

reported comparability on age, years since menopause, weight, height, dietary calcium intake and alcohol intake. Reid (2006) reported baseline comparability for age, weight, height and years since menopause.

One study reported more than 50% missing data: Peacock (2000) reported that 71 participants (56%) dropped out from the calcium group and 61 (45%) from the placebo group, that is, differential missing data between groups. Two studies (Hansson 1987; Riggs 1998) reported more than 20% missing data: Hansson (1987) reported that three participants (12%) dropped out of the calcium group and six (24%) dropped out of placebo. In Riggs (1998), 31 participants (26%) dropped out of the calcium group and 28 (24%) dropped out of placebo. Five studies (Avenell 2004; Chevalley 1994; Prince 2007; RECORD 2005; Reid 1993; 2006) reported less than 20% missing data. Recker (1996) did not provide details of attrition.

An ITT analysis was carried out for six studies, but it was unclear how this was achieved in four studies (Prince 2007; Recker 1996; Reid 2006; Riggs 1998); Avenell (2004) and RECORD (2005) reported that fracture incidence results from the last time of measurement were used in the analysis. In four studies (Chevalley 1994; Hansson 1987; Peacock 2000; Reid 1993) it was unclear whether an ITT analysis was used.

Fractures were primary outcomes in five studies (Chevalley 1994; Prince 2007; Recker 1996; RECORD 2005; Reid 2006); they were secondary outcomes in Reid (1993) and Riggs (1998) and adverse events in the other studies. All studies but four (Avenell 2004; Hansson 1987; RECORD 2005; Riggs 1998) stated that vertebral fractures were assessed radiographically using a decrease in height of at least 20%. Hansson (1987) did not report on the methods used to assess vertebral and nonvertebral fractures and Riggs (1998) reported measurements based on a decrease in height of 15%. Avenell (2004) and RECORD (2005) only reported clinical vertebral fractures.

Most studies that recorded nonvertebral fractures did not report whether they were osteoporotic or if they included trauma fractures. The exception was Prince (2007), which reported osteoporotic fractures only, and RECORD

(2005) which excluded trauma fractures. Recker (1996) and Reid (2006) did not record nonvertebral fractures.

Overall, the study by Peacock (2000) was considered to have high potential for bias because of high levels of missing data, together with a difference in missing data between groups, therefore this study was not included in the analysis. Five other studies were considered to have some potential for bias: Recker (1996) because of baseline differences; Avenell (2004) because the patients were not blinded; and Hansson (1987) and Riggs (1998) because there was more than 20% missing data; Hansson (1987) also had differential dropout rates across groups. Reid (2006) was treated with caution because some patients (unclear how many) were eliminated from the per protocol analysis because they took osteoporosis medication during the trial. In addition, the study by Riggs (1998) used a calcium dose that was considerably higher than the licensed dose and overall it was decided to exclude this study from the analysis. The other four studies (Avenell 2004; Hansson 1987; Recker 1996; Reid 2006) were considered in sensitivity analyses.

Finally, we note that the Chevalley (1994), Prince (2007) and Recker (1998) studies all used doses of calcium a little above the licensed dose and that the results for Chevalley (1994) are reported together with those of an unlicensed formulation.

## **Results**

### **Calcium versus placebo**

#### *Vertebral fractures*

Seven studies reported the number of people with vertebral fractures; one of these (Reid 1993) only reported symptomatic fractures; denominators were not given for this study and the randomised numbers were assumed. Avenell (2004) and RECORD (2005) recorded clinical fractures only. A meta-analysis of seven studies in 6013 participants found no significant differences in rates of vertebral fractures in the calcium group compared with placebo (figure 91); RR 0.84 (95% CI 0.66 to 1.08). There was no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.80$ ). A funnel plot was fairly symmetrical (figure 92)

A sensitivity analysis in the absence of Reid (2006), which was possibly confounded (figure 93), gave a RR of 0.90 (95% CI 0.67 to 1.19). The funnel plot was fairly symmetrical (figure 94). In the absence of all studies with an increased potential for bias, a meta-analysis of the two remaining studies, Prince (2007) and RECORD (2005), showed no significant difference between interventions; RR 1.03 (95% CI 0.67 to 1.57).

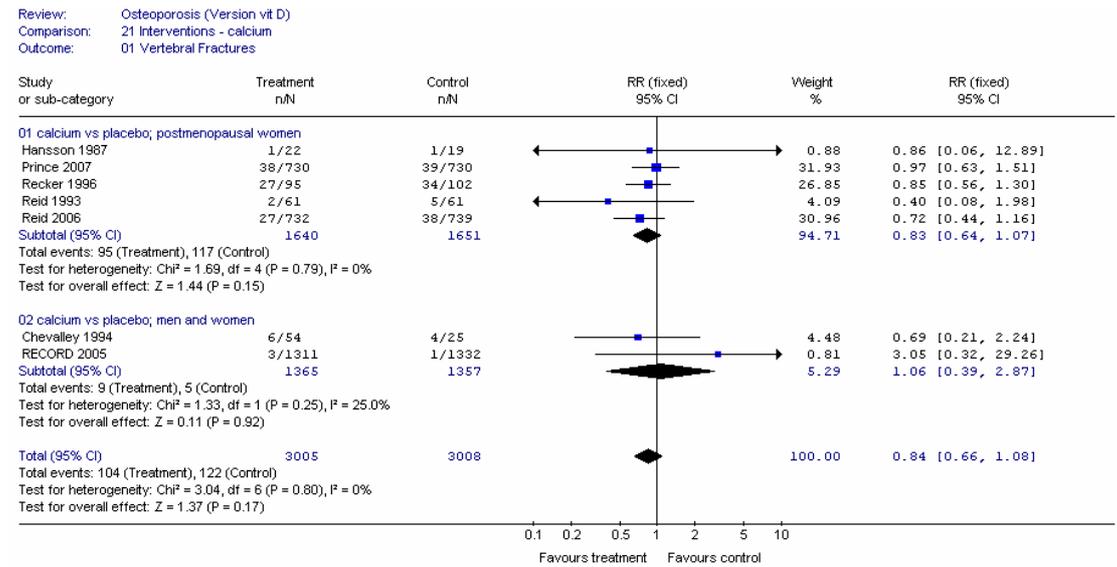


Figure 91. Number of patients with vertebral fractures

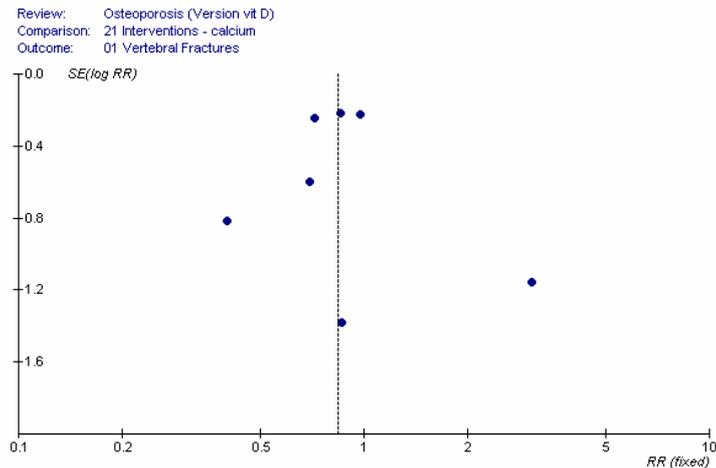
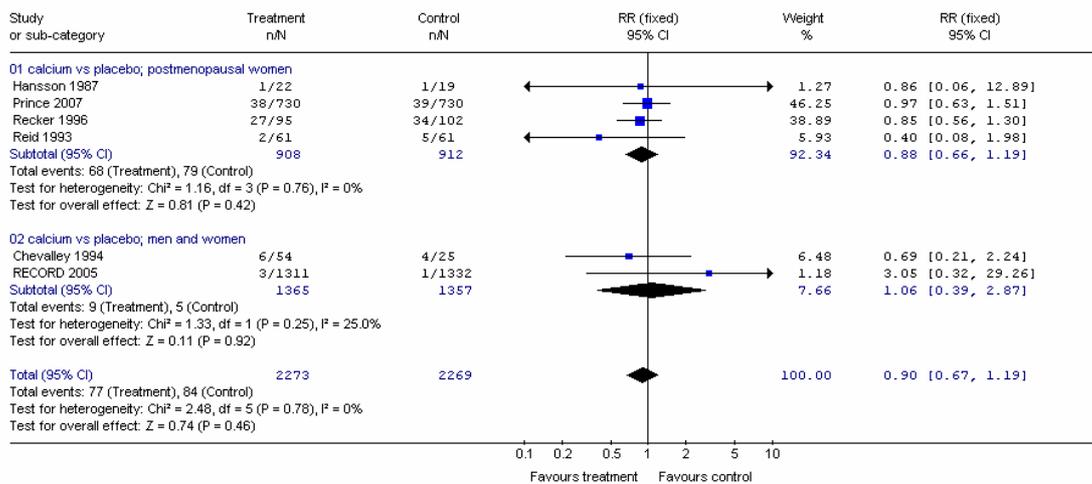


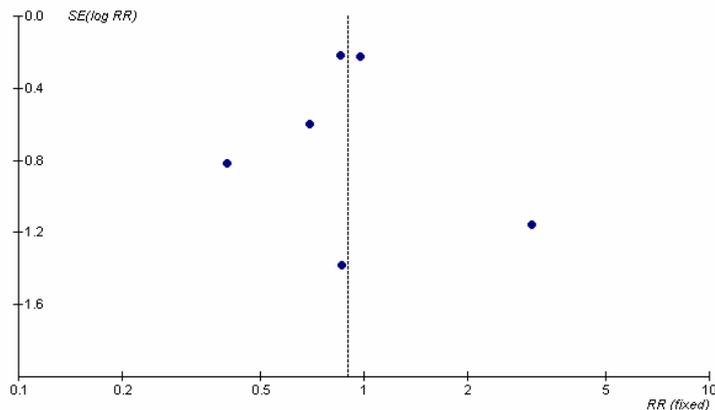
Figure 92. Number of patients with vertebral fractures: funnel plot

Review: Osteoporosis (Version vit D)  
 Comparison: 21 Interventions - calcium  
 Outcome: 01 Vertebral Fractures



**Figure 93. Number of patients with vertebral fractures; sensitivity analysis without Reid (2006)**

Review: Osteoporosis (Version vit D)  
 Comparison: 21 Interventions - calcium  
 Outcome: 01 Vertebral Fractures

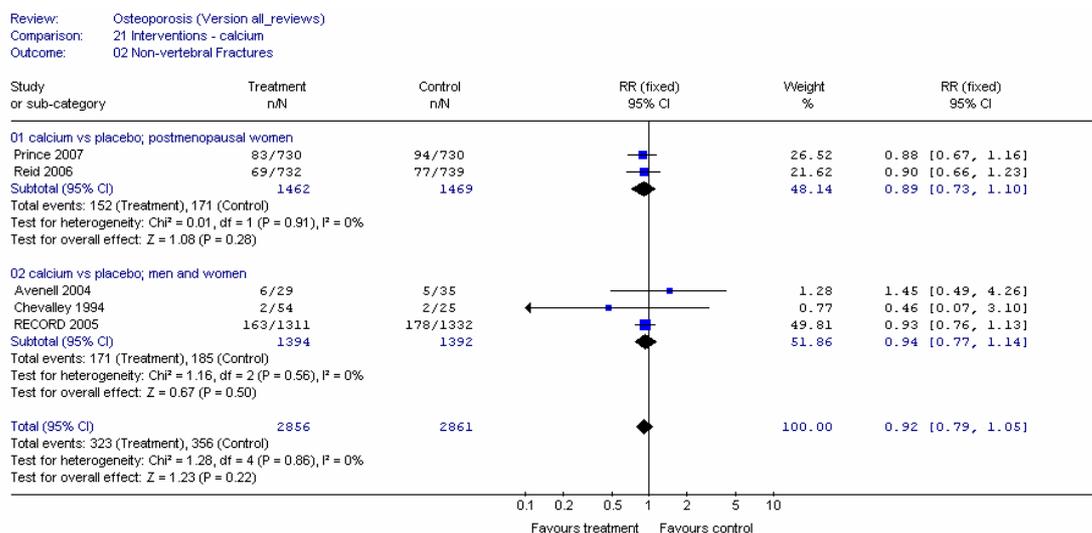


**Figure 94. Number of patients with vertebral fractures; sensitivity analysis without Reid (2006): funnel plot**

*Nonvertebral fractures*

Five studies reported the number of people with nonvertebral fractures. In Reid (2006) and RECORD (2005) the number of people with nonvertebral fractures was calculated by subtracting the number with vertebral fractures from the number with any osteoporotic fracture. In Prince (2007) the number of people with fractures at any appendicular site was used. A meta-analysis of five studies in 5717 patients found no significant difference between interventions (figure 95); RR 0.92 (95% CI 0.79 to 1.05). There was no

heterogeneity ( $I^2 = 0\%$ ;  $p = 0.86$ ) and a funnel plot was fairly symmetrical (not shown). In the absence of the studies with more risk of bias, the studies by Prince (2007) and RECORD (2005) gave a RR of 0.91 (95% CI 0.78 to 1.05).

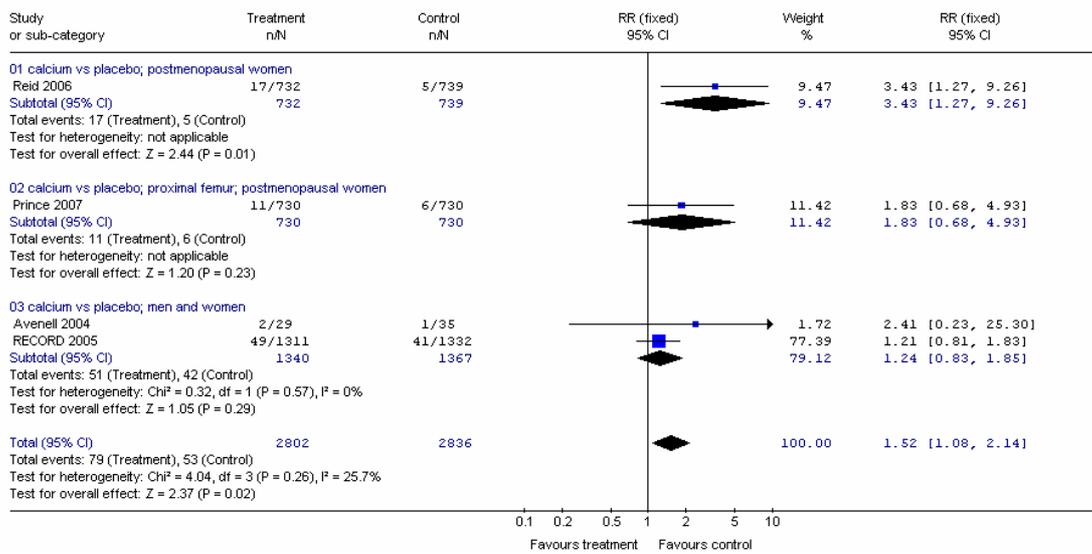


**Figure 95. Number of patients with nonvertebral fractures**

### *Hip fractures*

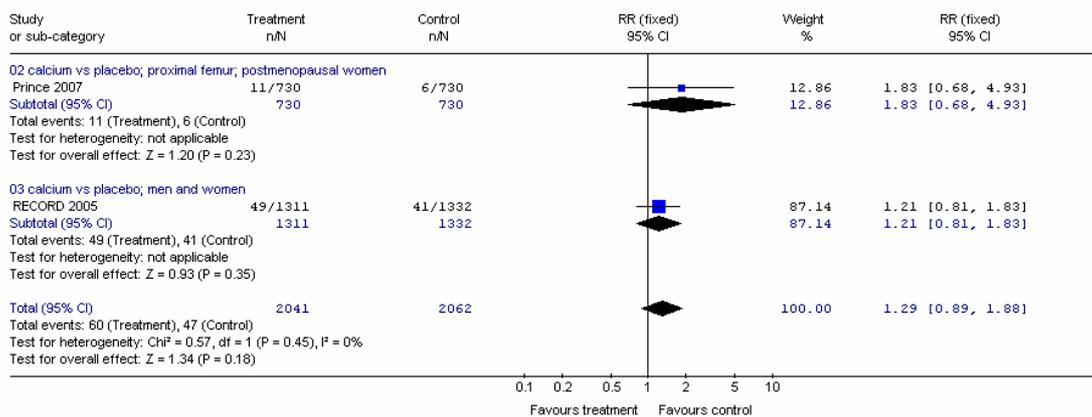
Four studies in 5638 participants (Avenell 2004; Prince 2007; RECORD 2005; Reid 2006) reported the number of people with hip fractures. Meta-analysis showed there were significantly more people with hip fractures in the calcium group compared with placebo (figure 96); RR 1.52 (95% CI 1.08 to 2.14); this corresponds to a number needed to harm (NNH) of 100 (95% CI 50 to  $\infty$ ) for a placebo group rate of 0.7–3%. There was no significant heterogeneity between studies ( $I^2 = 26\%$ ;  $p = 0.26$ ). A sensitivity analysis in the absence of the potentially confounded Reid (2006) and the higher risk Avenell (2004) gave a RR of 1.29 (95% CI 0.89 to 1.88), that is, not a significant effect and different from the full meta-analysis (figure 96b). The sensitivity analysis was preferred.

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 21 Interventions - calcium  
 Outcome: 03 Hip fractures



**Figure 96. Number of patients with hip fractures**

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 21 Interventions - calcium  
 Outcome: 08 Hip fractures: sensitivity analysis

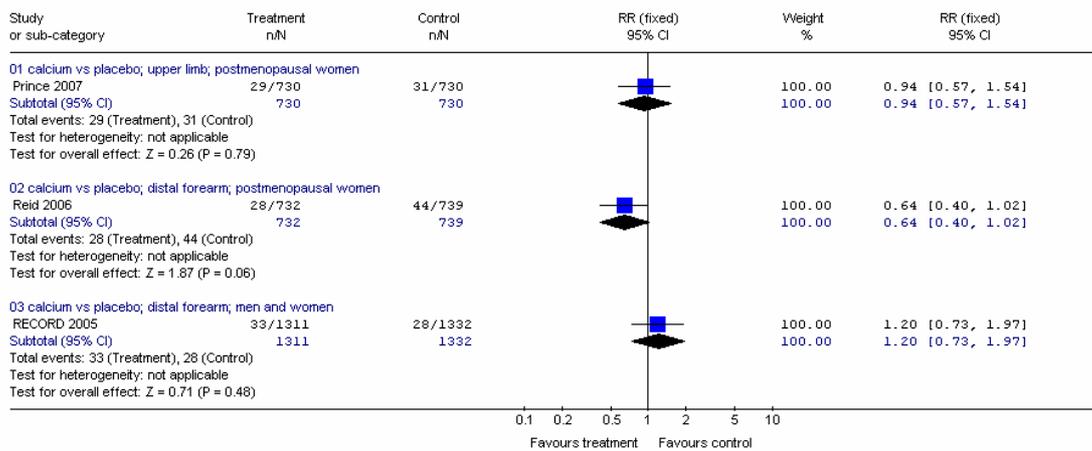


**Figure 96b. Sensitivity analysis: number of patients with hip fractures**

### Arm fractures

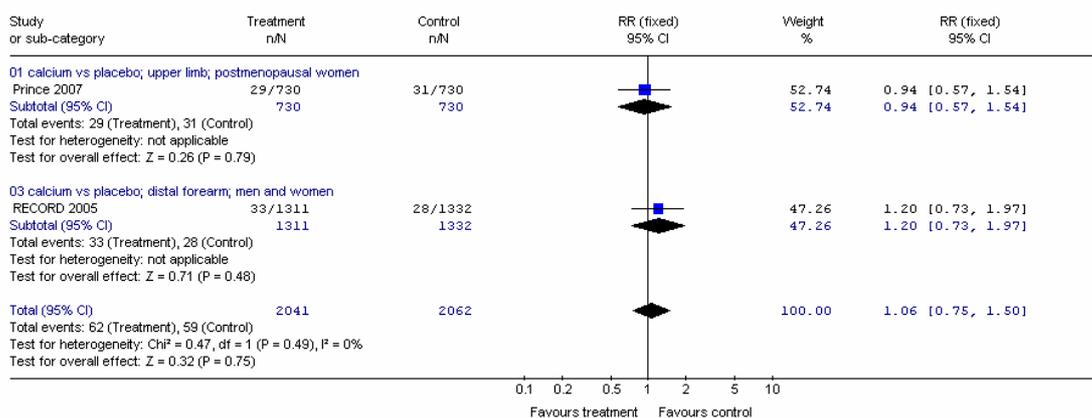
Three studies reported fractures in the upper limb and distal forearm. There was no significant difference in any of the studies (figure 97). In a meta-analysis of Prince (2007) and RECORD (2005) in 4103 patients, which had less potential for bias, the RR was 1.06 (95% CI 0.75 to 1.50) for fractures of the 'upper limb' and distal forearm (figure 97b); there was no heterogeneity ( $I^2 = 0\%$ ;  $p = 0.49$ ).

Review: Osteoporosis (Version vit D)  
 Comparison: 21 Interventions - calcium  
 Outcome: 04 Arm Fractures



**Figure 97. Number of patients with arm fractures**

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 21 Interventions - calcium  
 Outcome: 10 Arm Fractures - sensitivity analysis

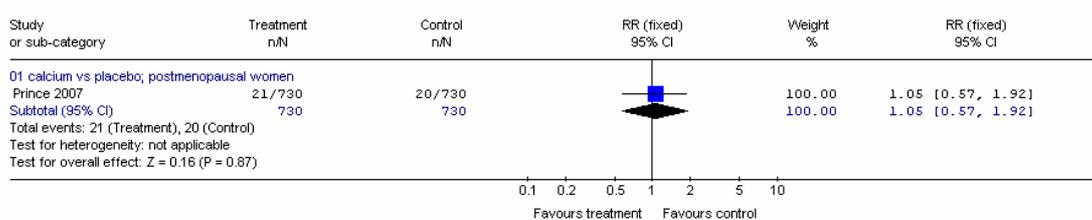


**Figure 98. Sensitivity analysis: number of patients with arm fractures**

### Wrist fractures

One study in 1460 patients found no significant differences between interventions in the number of patients with wrist fractures (figure 99); RR 1.05 (95% CI 0.57 to 1.92).

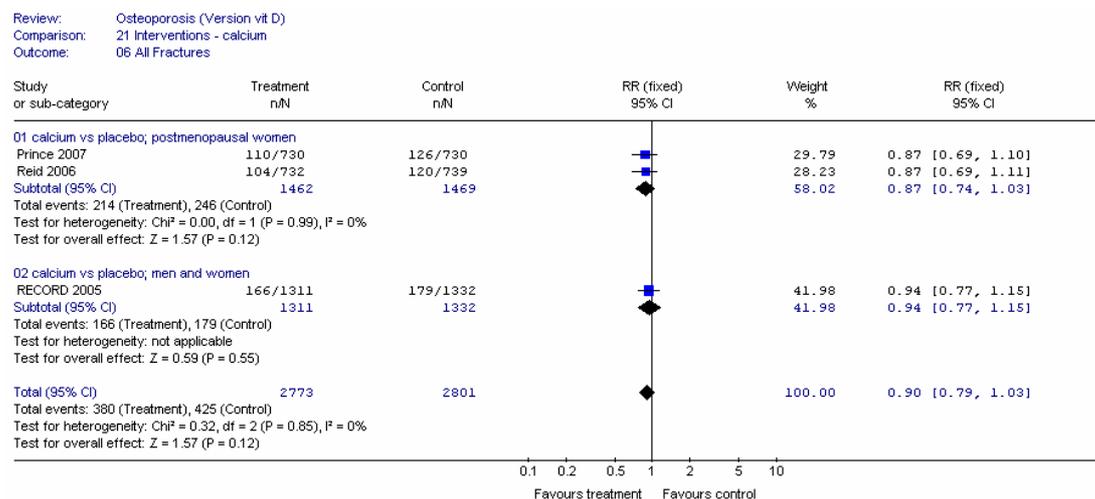
Review: Osteoporosis (Version all populations)  
 Comparison: 21 Interventions - calcium  
 Outcome: 05 Wrist fractures



**Figure 99. Number of patients with wrist fractures**

## All fractures

A meta-analysis of three studies in 5574 participants found there was no significant difference in the number of patients with any fracture (figure 100); RR 0.90 (95% CI 0.79 to 1.03). There was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.85$ ).



**Figure 100. Number of patients with any fracture**

## 2.6 Cross review comparisons

### Description of studies

A total of 37 papers were evaluated for inclusion. Thirteen of these only reported BMD as an outcome and were excluded from the review (Adami 1995; Bergstrom 2005; Deng 2004; Diamond 1997; Finkelstein 2003; Garcia-Delgado 1997; Henderson 2001; Hwang 2006; Kung 1999; 2006; Lydeking-Olsen 2004; Rhee 2006; Rozhinskaya 1999). Two studies (Gallagher 2001; Greenspan 2003) were excluded from this review because 59% and 65% respectively of the patients (those with an intact uterus) had an unlicensed dose of HRT. Both of these studies were included, however, in other reviews in this document. Thirteen other studies were excluded for reasons given in appendix IV. Eight reports were included (Black 2003; Eviö 2004; Luckey 2004; McClung 2005; Michalská 2006; Muscoso 2004; Palomba 2005; Sambrook 2004). One additional study was considered, Recker (2007), the Evista Alendronic acid Comparison (EVA) trial, which was terminated after a mean duration of 312 days because the study had not recruited sufficient

patients within the timelines. This study was included because it was a large study that was specifically designed to compare two interventions head-to-head, with fracture as the primary outcome.

### ***Study design***

No studies were conducted exclusively in the UK. Two were conducted in the USA (Black 2003; Luckey 2004); two in Italy (Muscoso 2004; Palomba 2005); one in Finland (Eviö 2004); one in the Czech Republic (Michalská 2006); one multisite study in Europe, South America and Asia-Pacific (Sambrook 2004); one multisite study in North America (Recker 2007); and one other multinational (McClung 2005).

Michalská (2006) enrolled only patients who had been treated with alendronic acid for at least 3 years (mean of 43 months) prior to randomisation (to double-blind raloxifene or placebo or open-label alendronic acid). Although patients were not randomised to alendronic acid followed by raloxifene or alendronic acid followed by alendronic acid, this is, in effect, what the trial was comparing, and this trial can be considered as a continuation trial and will be treated separately from those in which raloxifene and alendronic acid are compared.

The trialists in three studies received industry support from Eli Lilly (McClung 2005; Michalská 2006; Recker 2007); four from Merck (Black 2003; Eviö 2004; Luckey 2004; Sambrook 2004); one additionally from SmithKline Beecham and NPS (Black 2003); one study received a research grant from nonindustry funding (Palomba 2005). The other trials did not state the funding.

Muscoso (2004) was the largest study, with 2000 participants. Recker (2007) and Palomba (2005) included 1423 and 1100 women, respectively. The remaining trials were smaller (Sambrook 2004, n = 487; Luckey 2004, n = 456; Black 2003, n = 238; McClung 2005, n = 203; Michalská 2006, n = 99; Eviö 2004, n = 90). Two studies had fewer than 100 patients (Eviö 2004; Michalská 2006). Two studies had fewer than 50 patients in the intervention arm (Eviö 2004, n = 30; Michalská 2006, n = 33), but no studies had fewer than 20 patients in the intervention arm. One study (Muscoso 2004)

included 2000 patients, randomised to alendronic acid (1000), clodronate (800), risedronate (100) and raloxifene (100).

All the studies took place in secondary care.

### ***Population***

All the trials randomised postmenopausal women only. The mean time since the menopause, when reported, ranged from around 12 to 23 years.

Mean baseline T-scores were reported (or calculated) at the femoral neck or lumbar spine for all studies. On this basis, four were assessed as being in patients with osteoporosis only (Eviö 2004; Muscoso 2004; Palomba 2005; Recker 2007). Five (Black 2003; Luckey 2004; McClung 2005; Michalska 2006; Sambrook 2004) included those with osteoporosis or osteopenia,

Some participants included in three studies (Black 2003; Michalská 2006; Sambrook 2004) had baseline fractures (19–47%). Two studies included participants with no prior fractures (Palomba 2005; Recker 2007). The remaining five studies did not state whether participants had fractures.

The mean age of participants across studies ranged from 62 to 72 years, with individual ages ranging from 50 to 90 years.

Use of cigarettes was reported by three studies (Eviö 2004; Luckey 2004; Palomba 2005), which all reported that some participants had a history of smoking. Ethnicity data were limited; most participants were Caucasian.

### ***Interventions***

The following interventions were used.

#### **Bisphosphonates**

- Alendronic acid:
  - 10 mg/day alendronic acid (Black 2003; Eviö 2004; McClung 2005; Muscoso 2004; Palomba 2005; Recker 2007)
  - 10 mg/day alendronic acid following at least 3 years of alendronic acid (Michalská 2006)
  - 70 mg/week alendronic acid (Luckey 2004; Sambrook 2004).

- Risedronate:
  - 5 mg/day risedronate sodium (Muscoso 2004).

### **PTH**

- Teriparatide:
  - 20 micrograms/day teriparatide (McClung 2005).
- PTH (1-84):
  - 100 micrograms/day PTH (1-84) (Black 2003).

### **HRT**

- 2 mg oestradiol plus 1 mg norethisterone acetate (Eviö 2004).

### **Raloxifene**

- 60 mg/day raloxifene (Luckey 2004; Muscoso 2004; Palomba 2005; Recker 2007; Sambrook 2004).
- 60 mg/day raloxifene following at least 3 years of alendronic acid (Michalská 2006).

### ***Comparisons***

The included studies had 10 comparisons that are eligible for this review. Two studies also had cross comparisons with unlicensed interventions (Muscoso 2004: clodronate; Palomba 2005: HRT); these comparisons were not considered in this review. The head-to-head comparison of alendronic acid and risedronate sodium in Muscoso (2004) is reported in the cross bisphosphonates review (section 3.1.7). The following comparisons were carried out:

### **Bisphosphonates versus raloxifene**

- Alendronic acid versus raloxifene – six studies:
  - intervention duration mean 312 days, that is, less than 1 year (Recker 2007)
  - 1 year (Luckey 2004; Palomba 2005; Sambrook 2004)
  - 2 years (Muscoso 2004)
  - 2 years following at least 3 years of alendronic acid (Michalská 2006).
- Risedronate sodium versus raloxifene: one study:

- 2 years' duration (Muscoso 2004).

### **Bisphosphonates versus PTH**

- Alendronic acid versus teriparatide:
  - 18 months' duration (McClung 2005).
- Alendronic acid versus PTH (1-84):
  - 1-year duration (Black 2004).

### **Bisphosphonate versus HRT**

- Alendronic acid versus HRT – two studies:
  - intervention duration 2 years (Eviö 2004)

The studies gave calcium and vitamin D supplements as follows. In Eviö (2004), participants had their dietary calcium assessed at intake, then they were advised to take calcium supplements (0.5–1 g daily) to increase their intake to 1 g daily; they were also advised to take vitamin D from October to April (but this was not provided by the study). In Luckey (2004) participants were also given calcium and vitamin D concomitantly, adjusted according to intake. Palomba (2005) gave 0.5 g/day calcium to participants with a calcium intake below 1 g/day; no vitamin D supplement was used. In Sambrook (2004), calcium and vitamin D supplements were given in accordance with the standard care in the local community. In Black (2003) and Recker (2007), all participants were also given calcium (0.5 g) and vitamin D (400 IU) concomitantly. McClung (2005) and Muscoso (2004) gave all participants calcium (1 g) and vitamin D (800 IU) concomitantly, and Michalská (2006) gave 0.5 g calcium and 800 IU vitamin D.

### ***Methodological quality***

Four studies reported an adequate method of sequence generation: three (Luckey 2004; Palomba 2005; Sambrook 2004) used computer-generated randomisation and Recker (2007) used random number tables. In the remaining studies no details of the techniques used for randomisation were reported.

Allocation concealment was reported in two studies (Lucky 2004; Recker 2007). One of these had adequate methods: Recker (2007) used a computerised telephone system. In Lucky (2004), the allocation was described as concealed but no details were given. Allocation concealment was not stated in the other studies.

All the trials were stated to be double blind except for three (Michalská 2006; Muscoso 2004; Palomba 2005). In these trials the patients were not blinded: Michalská (2006) and Muscoso (2004) reported an open-label study was carried out and Palomba (2005) used clearly different interventions. It was unclear if the outcome assessors were blinded in Muscoso (2004) and Michalská (2006).

Details recorded at baseline in all the studies found participants comparable on age, BMI, years since menopause (when reported) and other characteristics, apart from four studies. In one study, one of the groups was found to have a higher rate of prior fracture. In the study by Michalská (2006) the two groups were comparable on age, weight, height, years since menopause and duration of prior alendronic acid treatment, but the alendronic acid group had fewer prior fractures: 27.3 versus 48.5% on raloxifene). In Muscoso (2004), the groups were comparable on age but no other parameters were given. In Palomba (2005), no baseline comparability data were presented.

A priori sample size calculations were carried out in four studies (Lucky 2004; Michalská 2006; Palomba 2005; Recker 2007). Recker (2007) assumed a sample size of 1750 per arm, giving 90% power to establish equivalence between alendronate and raloxifene for any osteoporotic fracture (vertebral or nonvertebral). However, the study was stopped early and this number was not reached. Lucky (2004) reported that 150 patients per group had 90% power to detect a treatment difference in change from baseline in lumbar BMD of 1.5% between groups. Michalská (2006) reported that 33 patients per group were needed to detect a 2.4% difference in lumbar spine BMD between two groups with 80% power. Palomba (2005) reported that 40 patients per genotype group were required to detect a 2% difference in mean percentage

change from baseline in BMD within groups and between groups at a power of 80%. One further study (Black 2003) reported an apparently post-hoc sample size/power calculation: 'given the standard deviations in this trial, with a power of 90%, we could detect a difference in the areal BMD of about 2.8% for the spine and 2.2% for the hip'. The other studies did not report power calculations.

One study reported that there were no missing data (Michalská 2006). Six studies reported missing data of less than 20% (Black 2003; Luckey 2004; McClung 2005; Palomba 2005; Recker 2007; Sambrook 2004). However, Recker (2007) only reported vertebral fracture data for 255/716 (36%) of the randomised patients because radiographs were not available. Eviö (2004) reported missing data of more than 20%: this was similar between the groups: 8/30 (27%) discontinued alendronic acid; 7/30 (23%) discontinued HRT. Muscoso (2004) did not clearly report study attrition.

Four studies used ITT analyses (Luckey 2004; Muscoso 2004; Recker 2007; Sambrook 2004). Four studies used available case analyses (Black 2003; Eviö 2004; McClung 2005; Palomba 2005) and Michalská (2006) used a per protocol analysis.

Recker (2007) reported assessing vertebral fractures radiographically, using a 20% reduction in height measure. Palomba (2005) used radiographic vertebral fractures but no details were given. In two studies the method was not stated (Black 2003; Muscoso 2004) and in five studies vertebral fractures were not recorded (Eviö 2004; Luckey 2004; McClung 2005; Michalská 2006; Sambrook 2004).

Fractures were reported as primary outcomes in one study (Recker 2007), as secondary outcomes in two studies (Muscoso 2004; Palomba 2005) and as safety/adverse event data in six studies (Black 2003; Eviö 2004; Luckey 2004; McClung 2005; Michalská 2006; Sambrook 2004). All the adverse event data were clinical fractures.

Fractures were defined as being osteoporotic in one study (Recker 2007). Seven studies did not define the types of fracture (Black 2003; Eviö 2004;

McClung 2005; Michalská 2006; Muscoso 2004; Palomba 2005; Sambrook 2004). One study (Lucky 2004) reported that all clinical fractures were included, whether or not associated with trauma.

Overall, five studies had potential for bias (Eviö 2004; Michalská 2006; Muscoso 2004; Palomba 2005; Recker 2007). Eviö (2004) had more than 20% missing data, and Recker (2007) had only 35% of patients providing results for the vertebral fracture outcome, which is likely to lead to unreliable conclusions. In addition, Recker (2007) was terminated early, with a mean duration of about 10 months. Three studies did not blind the patients (Michalská 2006; Muscoso 2004; Palomba 2005), and Michalská (2006) had baseline differences in the proportion of women with a prior fracture. Therefore, these five studies were considered in sensitivity analyses.

## **Results**

### **Bisphosphonates versus selective estrogen (oestrogen) receptor modulators**

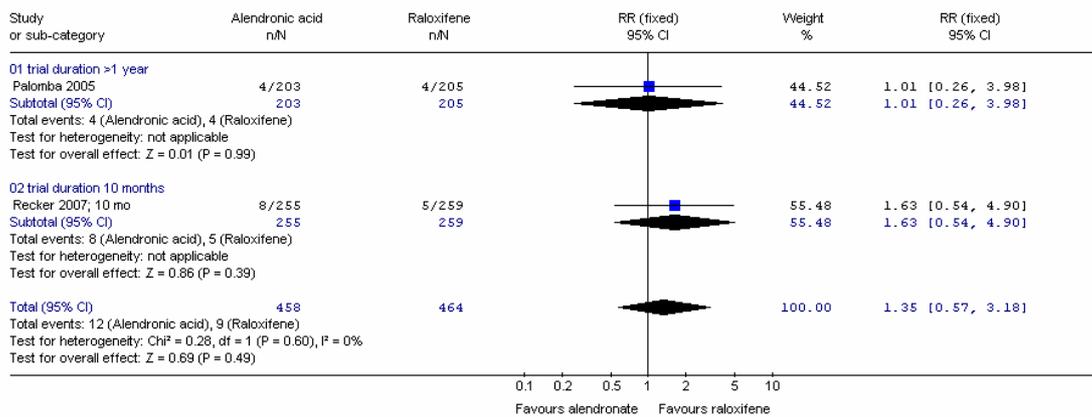
#### **Alendronic acid versus raloxifene**

Five studies compared alendronic acid versus raloxifene (Lucky 2004; Muscoso 2004; Michalská 2006; Recker 2007; Sambrook 2004); one of these (Muscoso 2004) only reported the number of fractures and so this study was not considered further.

#### *Vertebral fractures*

Two studies (Palomba 2005; Recker 2007) reported vertebral fractures in 922 patients on alendronic acid or raloxifene. The study by Recker (2007) was included even though the trial was terminated early and the mean duration was less than 12 months; this study only reported vertebral fracture data for 255/716 (36%) of the randomised patients because radiographs were not available, and must be considered at high risk of bias. A meta-analysis of the two studies in 922 patients showed there was no significant difference between groups, although the CI was fairly wide (figure 101); there was no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.60$ ). In the absence of the potentially flawed Recker (2007), the CI was too wide to determine if there was a difference between groups.

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 23 Interventions - cross review  
 Outcome: 01 Alendronic acid versus raloxifene; vertebral fractures

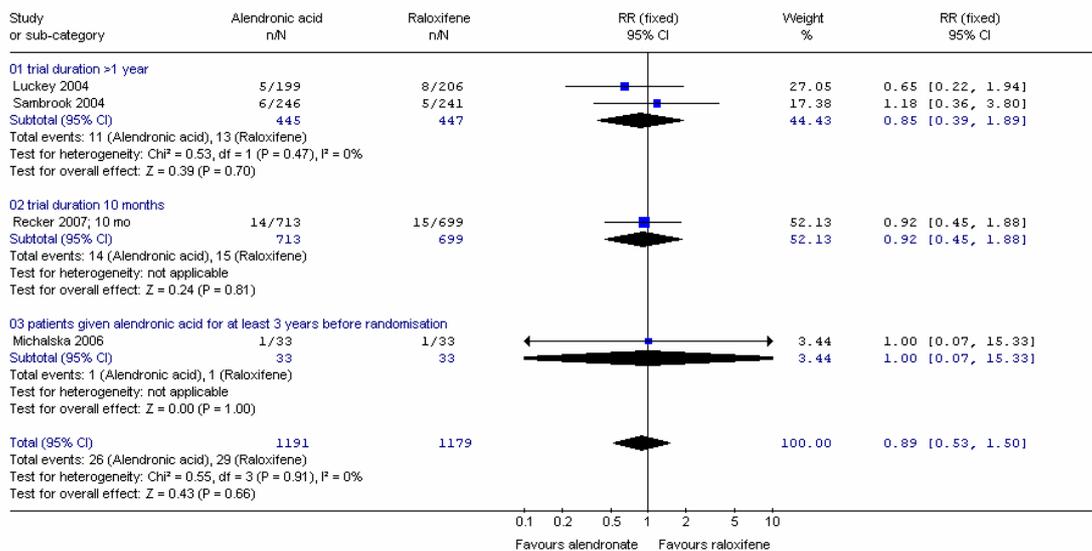


**Figure 101. Number of patients with vertebral fractures**

### *Nonvertebral fractures*

Four studies (Luckey 2004; Michalská 2006; Recker 2007; Sambrook 2004) reported nonvertebral fractures in patients on alendronic acid or raloxifene; all studies except Recker (2007) (in which it was unclear) reported clinical fractures only. Two of these studies were treated as separate subgroups: Michalská (2006) because it was akin to a continuation study for patients who had been receiving alendronic acid for at least 3 years, and Recker (2007) because the duration of the trial was less than 1 year (however, there were insignificant missing data for this outcome). A meta-analysis showed there was no significant difference between the groups (figure 102); RR 0.89 (95% CI 0.53 to 1.50), and no heterogeneity between studies ( $I^2 = 0\%$ ;  $p = 0.91$ ). Two of these studies were considered at risk of bias (Michalská 2006; Recker 2007); a sensitivity analysis in their absence made little difference to the summary statistics; RR 0.85 (95% CI 0.39 to 1.89) ( $I^2 = 0\%$ ;  $p = 0.47$ ).

Review: Osteoporosis (Version all\_reviews)  
 Comparison: 23 Interventions - cross review  
 Outcome: 02 Alendronic acid versus raloxifene; non-vertebral fractures

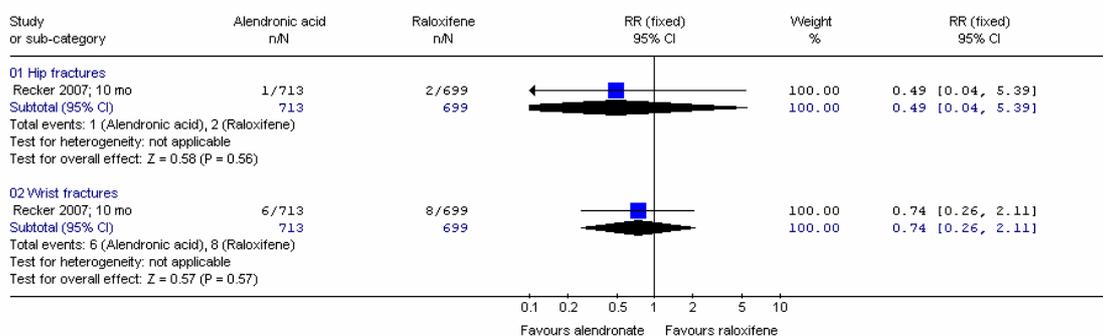


**Figure 102. Number of patients with nonvertebral fractures**

### Hip fractures and wrist fractures

One study (Recker 2007) in 1412 patients reported hip fractures and wrist fractures in a study that lasted only 10 months. The CIs were too wide to determine if there was a difference between interventions (figure 103).

Review: Osteoporosis (Version all populations)  
 Comparison: 23 Interventions - cross review  
 Outcome: 03 Alendronic acid versus raloxifene; trial duration ~10 months only



**Figure 103. Number of patients with hip fractures and number with wrist fractures**

### All fractures

Three studies (Luckey 2004; Sambrook 2004; Recker 2007) reported the number of patients with any fracture in 3404 patients on alendronic acid or raloxifene; all except Recker (2004) recorded clinical fractures. The study by Recker (2007) reported all patients with vertebral and nonvertebral fractures, but vertebral fractures were only assessed in one-third of the patients. A meta-analysis showed there was no significant difference between

interventions (figure 104); RR 0.99 (95% CI 0.62 to 1.60), with no heterogeneity between groups ( $I^2 = 0\%$ ;  $p = 0.69$ ). All studies except Luckey (2004) and Sambrook (2004) were considered at risk of bias; a sensitivity analysis including only these studies gave a RR of 0.85 (95%CI 0.39 to 1.89), which is different from the full meta-analysis, but does not change the overall conclusions.

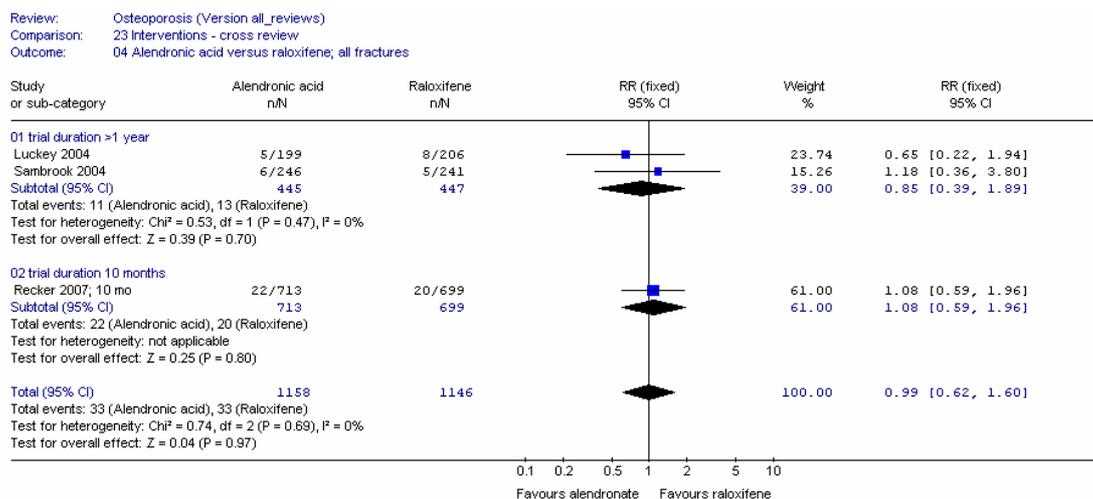


Figure 104. Number of patients with any fracture

## Bisphosphonate versus parathyroid hormone

### Alendronic acid versus teriparatide

#### All clinical fractures

One study (McClung 2005) reported all clinical fractures in 203 patients on alendronic acid or teriparatide. There was no significant difference between groups, although the confidence interval was fairly wide; RR 0.90 (95% CI 0.36 to 2.23) (figure 105).

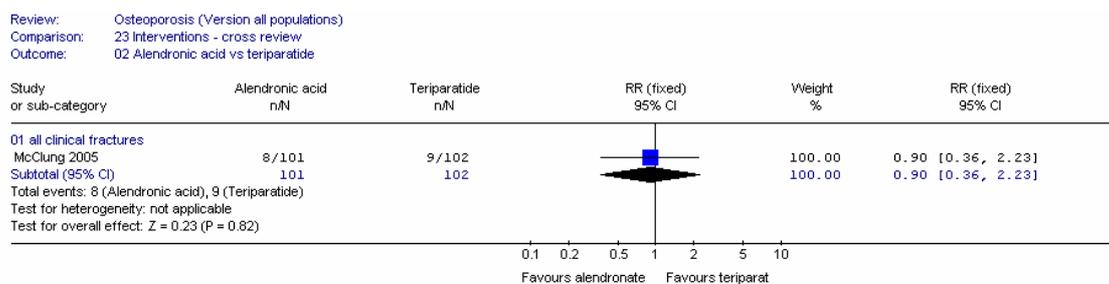


Figure 105. Number of patients with any clinical fracture

## **Bisphosphonate versus parathyroid hormone**

### **Alendronic acid versus PTH (1-84)**

#### *All clinical fractures*

One study (Black 2003) reported all clinical fractures in patients on alendronic acid or PTH (1-84). The authors reported eight fractures overall and stated that the incidence was similar in all three treatment groups (about 3%).

## **Bisphosphonate versus HRT**

### **Alendronic acid versus HRT**

#### *Hip fractures*

One study (Eviö 2004) in 60 patients reported the number of patients with a clinical hip fracture. One person in the alendronic acid group had a fracture. The CI was too wide to determine if there was a difference between groups (figure 106). The denominators were assumed to be the numbers randomised.

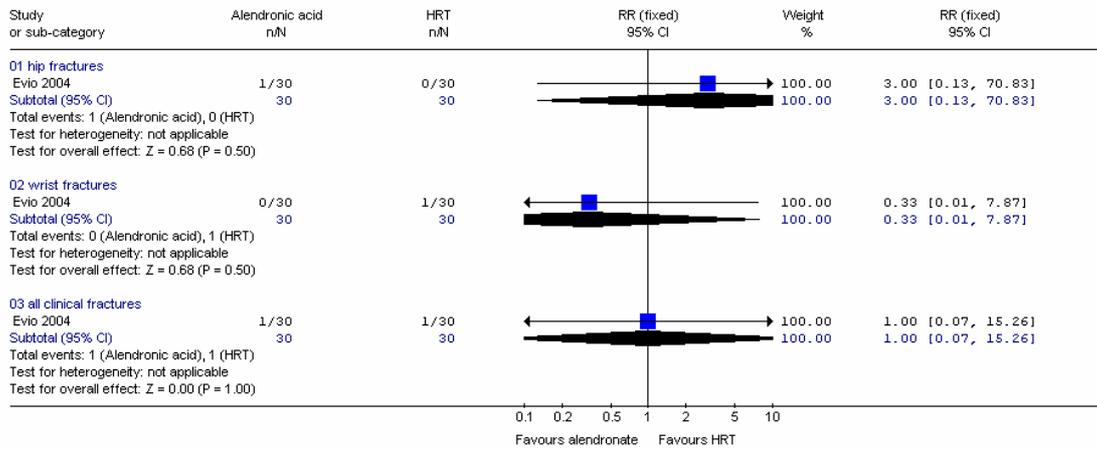
#### *Wrist fractures*

One study (Eviö 2004) in 60 patients reported the number of patients with a clinical wrist fracture. One person in the HRT group had a fracture. The CI was too wide to determine if there was a difference between groups (figure 106). The denominators were assumed to be the numbers randomised.

#### *All clinical fractures*

One study (Eviö 2004) in 60 patients showed the CI was too wide to determine if there was a difference between groups (figure 106). The denominators were assumed to be the numbers randomised.

Review: Osteoporosis (Version all populations)  
 Comparison: 23 Interventions - cross review  
 Outcome: 13 Alendronic acid versus HRT



**Figure 106. Number of patients with clinical fractures**