I. Background and Objectives for the Systematic Review

Osteoporosis is a systemic skeletal disease characterized by decreasing bone mass and microarchitectural deterioration of bone tissue, with consequent increases in bone fragility and susceptibility to fracture. In addition to fractures, the clinical complications of osteoporosis include disability and chronic pain. Approximately 52 million people in the United States are estimated to have osteoporosis or low bone density. It is especially common in postmenopausal women. The economic burden of osteoporosis is large and growing: the most recent estimate of U.S. annual costs due to fractures alone have been nearly $20 billion. A recent projection of the burden and costs of incident osteoporosis-related fractures in the U.S. from 2005 to 2025 estimates more than 2 million fractures in 2010 with direct medical costs of more than $18 billion (more than 25 percent attributable to men). Although the bulk of these costs are incurred by individuals 65 years of age and older, direct costs and productivity loss among working women under age 65 are considerable.

Osteoporosis is diagnosed by assessing bone mineral density (BMD). A number of techniques have been developed to measure BMD. Perhaps the most commonly used method is dual energy x-ray absorptiometry (DXA). In DXA, results are presented as a T-score which compares a person’s BMD to that of healthy 20 to 29 year old adults of the same sex and represents the number of standard deviations above or below the mean. Osteoporosis is defined as a T-score of -2.5 or less. Low bone mass or density is defined as a T-score between -2.5 and -1. A T-score of -1 or greater is considered normal. A second measurement called the Z-score expresses the number of standard deviations a BMD is above or below the mean for individuals of the same age, sex, and body size, and it is not used for persons over the age of 50.

Risk factors for osteoporotic fracture include increasing age, female sex, decreasing estrogen levels in women, low testosterone levels in men, low body weight, family history of hip fracture, race/ethnicity, previous clinical vertebral fracture, previous fracture due to minimal trauma (i.e., previous osteoporotic fracture), any fracture sustained over the age of 50, rheumatoid arthritis, current smoking, alcohol intake (three or more drinks/day), and low bone mineral density. Several algorithms have been devised and validated for the prediction of 10-year osteoporotic fracture risk. The FRAX, which takes into account age, year of birth, height, weight, secondary osteoporosis, femoral neck bone mineral density, family history of fracture, previous fracture history, use of glucocorticoids, and smoking and alcohol use history to predict the 10-year probability of fracture, has been endorsed to be used by current osteoporosis guidelines (such as the National Osteoporosis Foundation guidelines) to select treatment candidates. A question of considerable interest is whether response to treatment is affected by or predicted by FRAX score.

A number of agents are currently available to treat osteoporosis, including the bisphosphonate class of drugs, several peptide hormones (parathyroid hormone and calcitonin),
estrogen, and selective estrogen receptor modulators, as well as testosterone, calcium, vitamin D, and exercise, which have been theorized to play a role in preserving bone mass. The increasing prevalence and cost of osteoporosis has heightened interest in the effectiveness and safety of these interventions.

In December, 2007, the RAND Evidence-based Practice Center (EPC) completed the first Comparative Effectiveness Review (CER) on the efficacy/effectiveness of these interventions in preventing osteoporosis-related fracture and their safety, entitled Comparative Effectiveness of Treatments to Prevent Fractures in Men and Women with Low Bone Density or Osteoporosis.6

This review noted that there was good evidence that alendronate, etidronate, ibandronate, risedronate, zoledronic acid, estrogen, parathyroid hormone (1-34, also called teriparatide), and raloxifene all prevent vertebral fractures when compared with placebo; the evidence for calcitonin was fair. The effects of vitamin D varied with dose, analogue, and study population. For hip fracture, there was good evidence that alendronate, risedronate, and estrogen prevent hip fractures more than placebo; the evidence for zoledronic acid was fair. Again, the effects of vitamin D varied with dose, analogue, and study population. No evidence was available for calcium, testosterone, or physical exercise regarding fracture prevention.

Further, the evidence was insufficient to determine the relative superiority of any agent. However, treatment efficacy depends upon adherence, and although fracture trials that reported data on adherence/compliance tended to find relatively good adherence to medication use, observational studies tended to report poor adherence with bisphosphonates and calcium. Finally, in terms of harms, raloxifene, estrogen, and estrogen–progestin increased the risk for thromboembolic events, and etidronate increased the risk for esophageal ulcerations and gastrointestinal perforations, ulcerations, and bleeding. Some of these adverse effects - and the dosing instructions implemented to avoid them - appear to reduce compliance with and persistence of treatment.

In 2008, the RAND EPC was asked to conduct an assessment of the need to update the original report. As part of this assessment, the RAND EPC conducted an “abbreviated” search and review of the literature addressing the topics of the first review. The abbreviated search consisted of a survey of experts in the field and a Medline search (using the same search terms as the original report) of five of the leading medical journals and four leading specialty journals dating from 2006 to mid-2008. The studies identified in this search that addressed the key questions were then reviewed and abstracted, and their findings qualitatively assessed using a process devised by the EPC to determine whether they confirmed, contradicted, or augmented the conclusions of the original report.

<table>
<thead>
<tr>
<th>Journals Included in Search</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generalist Journals Included in Search</strong></td>
</tr>
<tr>
<td>Annals of Internal Medicine</td>
</tr>
<tr>
<td>British Medical Journal</td>
</tr>
<tr>
<td>JAMA (Journal of the American Medical Association)</td>
</tr>
<tr>
<td>The Lancet</td>
</tr>
<tr>
<td>New England Journal of Medicine</td>
</tr>
<tr>
<td><strong>Specialty Journals Included in Search</strong></td>
</tr>
<tr>
<td>Bone</td>
</tr>
<tr>
<td>Journal of Bone and Mineral Research</td>
</tr>
<tr>
<td>Osteoporosis International</td>
</tr>
<tr>
<td>Endocrine Reviews</td>
</tr>
</tbody>
</table>
The assessment concluded that at least some of the findings of the first report may need to be updated. Since the publication of the original report, several new agents (e.g., ibandronate) have been approved for use to treat or prevent osteoporosis, new dosing forms of several existing agents have been approved, and additional data on the efficacy of existing agents (zoledronic acid; calcium and vitamin D) have also been identified. In addition, the assessment found new evidence on adherence and on the safety of some agents, including the risk of atrial fibrillation with the use of some bisphosphonates and the risk of osteosarcoma with the use of selective estrogen receptor modulators (SERMS) and teriparatide. Also, the U.S. Food and Drug Administration (FDA) issued a warning in January 2008 linking the use of bisphosphonates with musculoskeletal pain and issued a labeling revision in December 2008 regarding the possible association of the use of pamidronate with deterioration of renal function. Based on these findings, AHRQ commissioned a full update of the original CER. In addition to an update of the analyses for each of the original key questions (modified to consider the newer agents), at the suggestion of the technical expert panel (TEP), the new report will include an additional question to address the role of bone density monitoring in predicting response to treatment in terms of fracture prevention. This question has become a topic of considerable controversy and may influence the prescription and use of preventive measures.

II. The Key Questions

Although these are the same questions that were addressed in the first report, they have been refined. In addition, KQ5 was added at the suggestion of the TEP and local subject matter expert. These questions were posted for public comment. In response to these comments, we have provided some minor qualifications to the questions. We will address the comments in greater detail in the report.

KQ1: (a) What are the comparative benefits in fracture reduction among and also within the following treatments for low bone density:

- Bisphosphonates: specifically alendronate, risedronate, etidronate, ibandronate, pamidronate, and zoledronic acid
- Calcium
- Denosumab (awaiting FDA approval)
- Menopausal hormone therapy
- Parathyroid hormone (PTH, both 1-34 and 1-84)
- Selective estrogen receptor modulators (SERMs), specifically raloxifene and lasofoxifene (awaiting FDA approval)
- Testosterone for men
- Vitamin D
- Combinations or sequential use of the above agents
- Exercise compared with the above agents
b) How does the anti-fracture benefit vary with continued use of pharmacotherapy, and what are the comparative anti-fracture efficacies of continued long-term therapy with the various pharmacotherapies?

- Population(s):
  - For KQ1, all patients with, or at risk for, osteoporotic fractures (or who meet the current guidelines for initiating treatment) are included.
  - Subgroups of patients are considered in KQ2.
- Interventions:
  - The interventions are listed above as part of the question
  - The trade names, chemical names, indications, dosages, frequencies, and routes of administration are listed in Table 1 of the original report
- Comparators:
  - The comparators include the other agents listed as well as placebo and usual care.
- Outcomes
  - The primary outcome is change in fracture incidence (vertebral, non-vertebral, wrist, hip, spine)
  - No secondary outcomes will be considered
  - Adverse events will be analyzed for KQ4.
- Timing:
  - Minimum duration of follow-up will be 12 months. Assessment of long-term effects will be based on the consensus definition of “long-term” in studies and on TEP input.
- Settings:
  - For all interventions, the setting is considered to be primary/outpatient or, if relevant, residential care facilities.

KQ2: How does fracture reduction resulting from treatments vary between individuals with different risks for fracture as determined by the following factors:

- Bone mineral density (borderline/low/severe)\(^a\)
- FRAX or other risk assessment score
- Prior fractures (prevention vs. treatment)
- Age
- Gender
- Race/ethnicity
- Glucocorticoid use

\(^a\) Bone mineral density is usually reported as a T-score, the number of standard deviations above or below the mean for healthy 20-29-year old adults of the same sex, as determined by dual energy X-ray absorptiometry (DXA). Osteoporosis is defined as a T-score of -2.5 or less. Osteopenia is defined as a T-score between -2.5 and -1. A T-score of -1 or greater is considered normal. (ISCD Guidelines)
• Other factors (e.g., community dwelling vs. residents of long-term care facilities, vitamin D status)

All details are the same as for KQ1

KQ3: a) What are the adherence and persistence to medications for the treatment and prevention of osteoporosis?

b) What factors affect adherence and persistence?

c) What are the effects of adherence and persistence on the risk of fractures?

All details are the same as for KQ1 and 2 with the possible exception of the following:

• Outcomes
  – One primary outcome is adherence/persistence (a and b).\(^b\)
  – Another primary outcome is fracture risk (likelihood of sustaining a fracture during the follow-up period) (c)

KQ4: What are the short- and long-term harms (adverse effects) of the above therapies, \(\text{when used specifically to treat or prevent low bone density/osteoporotic fracture}\), and do these harms vary by any specific subpopulations (e.g., the subpopulations specified in KQ2 or any other subpopulations)?

All details are the same as for KQ1 and 2, with the exception of the outcomes.

• Outcomes:
  – Any reported harms/adverse events (AEs) that can be directly associated with use of an agent for the prevention or treatment of osteoporosis.
  – Studies that report harms but that do not report fracture risk as an outcome will also be considered.

KQ 5: With regard to treatment for preventing osteoporotic fracture:

a) How often should patients be monitored (via measurement of bone mineral density) during therapy?

b) How does bone density monitoring predict anti-fracture benefits during pharmacotherapy? And

c) Does efficacy of monitoring vary among the pharmacotherapies?

• Interventions: The same interventions will be considered as for KQ1 and KQ2.

• Comparators

\(^b\) The term adherence is defined as the extent to which a person's health behavior is consistent with medical advice; adherence is synonymous with compliance. Persistence is defined as adhering to a treatment for the recommended duration, while nonpersistence is defined as treatment discontinuation without medical recommendation (Gass 2006).
- Studies that use an untreated group as comparators will be the focus of the analysis.

- **Outcomes**
  - We propose using fracture incidence as the outcome.

- **Timing**
  - Timing and settings will be the same as for KQ1 and KQ2.
III. Analytic Framework

Clinical Populations
All adults with OP or at risk for OP fractures

Factors that affect adherence/persistence with treatment

Interventions
Bisphosphonates
SERMs
Anabolic Steroids
Peptide hormones (calcitonin, teriparatide)
Calcium
Vitamin D
Exercise

2nd Outcome: Adverse effects of treatment

Monitoring via DXA

Health Outcomes:
Change in fracture risk (KQ1,2)
Change in BMD (KQ[?], 6)

Risk assessment

Factors affecting outcomes
Postmenopausal women
Men treated for prostate cancer
Transplant recipients
Nursing home residents, etc.

Figure Notes: BMD bone mineral density; DXA dual energy x-ray absorptiometry; OP osteoporosis; SERMs Selective estrogen receptor modulators. *T connotes the timing of outcome measurement for studies that will be included, which will vary by key question.
According to this analytical framework (in which the circled numbers refer to the key questions), KQ1 examines the evidence for the effectiveness of the interventions of interest to treat or prevent osteoporosis (OP), i.e., on the risk for first or subsequent osteoporotic fracture among all adults with increased risk for osteoporosis. KQ2 assesses the evidence for factors that affect the effectiveness of treatment (e.g., comorbidity, gender, age, place of residence). KQ3 assesses the evidence that adherence to or persistence with treatment affects treatment effectiveness. KQ4 assesses the evidence for adverse effects of treatment. KQ5 assesses the ability of risk predictors to identify people who are at greater risk of OP fracture or who are more likely to benefit from treatment. KQ6 assesses the evidence regarding the effect of length of treatment and monitoring frequency on treatment outcomes (as well as the definition of treatment success).

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

We will adhere to four sets of inclusion/exclusion criteria; the first three are as described in the first report. Searches for KQ1-KQ4 will commence from 2006; searches for KQ5 may commence with 1966 if it is determined that the literature identified for the original report and this report do not adequately address the questions. We will include non-English articles.

KQ1 and KQ2. Inclusion criteria for effectiveness/efficacy questions: For titles obtained through the searches for studies of drugs and osteoporosis, we will accept any title that suggests the manuscript might include information on the treatment/prevention of osteoporotic fracture. Controlled clinical trials that report fracture outcomes for one or more of the drugs of interest will be accepted for the efficacy analysis and will go on to data extraction. For agents for which no clinical trial data are available, we will consider large observational studies. We will also include reports of post hoc analyses and open-label extensions of trials.

KQ3. Inclusion criteria for adherence question: for the titles identified from the search for adherence and compliance, titles will be accepted if they suggest that the manuscript might include information on adherence and compliance. Articles of any study design that report on adherence/persistence for any of the drugs of interest will also be included for further evaluation.

KQ4. For titles obtained from the search for AEs by drug of interest, titles will be accepted if they suggest that the manuscript includes information on the relationship between the AE and the drug. Controlled clinical trials and large case control or cohort studies (n > 1000) that report fracture or BMD or markers of bone turnover for one or more of the drugs of interest and that report one or more AE will be included in AE analyses. Also included in AE analysis will be articles identified through the search for specific AE, if any AE for any of the drugs of interest are reported. For rare AEs of particular current interest (e.g., osteonecrosis of the jaw, atrial fibrillation, low stress tibial fracture), we will include studies of any design.

For questions pertaining to efficacy and AEs, we will include grey literature (including scientific information packets). Although we will not systematically contact researchers for additional data, we might, for newly released drugs, contact FDA and the manufacturers for unpublished data.

KQ5. Inclusion criteria for KQ5 will be the same as for KQ1 and KQ2.

Source: www.effectivehealthcare.ahrq.gov
Published Online: May 14, 2010
B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions.

As described for the first report,6 we will use a two-pronged approach to searching for relevant literature. First, we will identify any relevant systematic reviews that have appeared since the original searches were conducted. We will also search the Web of Science to identify any articles (systematic reviews and original research) that cite the original report or the article based on that report.

For the second prong of our approach, we will conduct three main searches. Our basic search strategy will use the National Library of Medicine’s Medical Subject Headings (MeSH) key word nomenclature developed for MEDLINE® and adapted for use in the other databases. We will search MEDLINE® for the period from January 2006 to the present. We will also search EMBASE, the American College of Physicians Journal Club database, the Cochrane controlled trials register, and relevant pharmacological databases. The search will not be limited by publication type (e.g., reports of randomized controlled trials, systematic reviews). To identify additional systematic reviews and meta-analyses not captured in our primary search strategy, we also will search MEDLINE®, the Cochrane Database of Systematic Reviews, and the websites of the National Institute for Clinical Excellence and the NHA Health Technology Assessment Programme. To identify relevant meetings proceedings, we will use the Web of Science to search proceedings of the meetings of the American Society of Bone and Mineral Research, the National Osteoporosis Foundation, and the International Osteoporosis Foundation. We will also manually search the reference lists of review articles obtained as part of our search (“reference mining”). Any materials obtained from the Scientific Resource Center (such as the scientific information packets) or directly from the FDA or manufacturers will be added to the materials obtained from these searches.

Each title list will be screened separately by two reviewers with clinical training and experience in systematic review. Full articles will be obtained for all selected titles. The reviewers will then conduct a second round of screening to ascertain which articles meet the inclusion criteria and will go on to data abstraction. Selections at this stage will be reconciled, and disagreements will be settled by consensus.

Update searches will be conducted 6 months after the start date of the project and during the review period. Any new articles identified by these searches or identified by the TEP or reviewers will be subjected to the same screening process as the articles from the original searches.

In our search to identify clinical studies of drugs of interest to this review, we will use terms for osteoporosis, osteopenia, low bone density, and the drugs listed in the key questions (as well as any new agents identified by the TEP). In our search for the key AE, we will use terms for the AE and each of the drugs of interest for this report. In our search for adherence and persistence we will use terms for adherence and persistence and the drugs of interest for this report. In all cases both generic and trade names will be used. In our search for studies on the effects of monitoring, we will search on terms related to monitoring and DXA in combination with the drugs of interest.

Source: www.effectivehealthcare.ahrq.gov
Published Online: May 14, 2010
C. Data Abstraction and Data Management

During the second (article inclusion) screening phase, for articles that satisfy the inclusion criteria, the two reviewers will abstract all information needed to assess the relevance and applicability of the study (Participants [including loss to followup], Intervention, Comparator(s), Outcomes assessed, Timing [including run-ins, wash-out periods, and times to followup], and Setting [geographical and clinical]) using a preprinted form (“quality review form”). This information is reconciled and dual-entered into a database to allow articles to be tracked by design. In the report, evidence tables will be used to present this information, as well as outcome data.

Fracture outcomes data will be dually extracted from each article accepted for analysis by a reviewer and a biostatistician. Outcomes for which three or more publications have been identified will be pooled for analysis.

AE data will be dually extracted in two ways. For each article that reports AEs, a preprinted form will be completed that tabulates the AEs according to the Common Terminology Criteria for Adverse Events. Then, a verbatim account of the AE will be recorded on the form. We will also follow the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews.

D. Assessment of Methodological Quality of Individual Studies

For studies of efficacy, it is anticipated that only randomized controlled trials (RCTs) and existing meta-analyses will be included. The quality of RCTs will be assessed using the Jadad scale as well as an assessment of concealment of allocation. The quality of meta-analyses will be assessed using aspects of internal and external validity as suggested in the Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Items assessed include search strategy, inclusion criteria for individual studies, and method of synthesis, among others.

The need to include observational studies will also be carefully assessed according to the guidelines presented in the Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews. Assessment of the quality of these studies will depend on the types of studies included.

As was done for the original report, we will assess the applicability of each included study based on the similarity of the target populations to those for which this report is intended. This assessment will be separate from other quality assessments.

E. Data Synthesis

As described above, for comparisons for which we are able to identify three or more outcomes, we will conduct pooled analyses as described in the original report. For outcomes that were analyzed in that report for which new data are identified, we will attempt to pool the new and old data. For all other outcomes, we will provide narrative descriptions of the outcomes of each study. The data relevant to each outcome will be presented in individual tables.

For pooled analysis, we will assess heterogeneity and conduct sensitivity analysis. Wherever possible, data will be presented separately for men versus women; postmenopausal women versus younger women; individuals with prior fractures versus those with first fractures; and those with a relevant comorbidity (transplant patients, corticosteroid recipients) versus those without.

Source: www.effectivehealthcare.ahrq.gov
Published Online: May 14, 2010
F. Grading the Evidence for Each Key Question

For each outcome, heterogeneity will be assessed using the Q test and $I^2$ test. The evidence for publication bias will be assessed using the Egger’s test will be presented in funnel plots.

In addition, for each question, we will apply the criteria of the GRADE Working Group to assess the strength of the evidence.14
V. References

13. AHRQ Effective Health Care Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews; Chapter 4: Selecting Evidence: Observational Studies of Beneficial

Source: www.effectivehealthcare.ahrq.gov
Published Online: May 14, 2010

VI. Definition of Terms

The clinical definition of osteoporosis is based either on clinical evidence of fracture or on densitometric measurement of bone mineral density (expressed as grams [gms] per centimeter [cm]²).

The National Osteoporosis Foundation (NOF) defines osteoporosis as a disease characterized by increased bone fragility and risk for fracture; of special concern are fractures of the hip and spine.¹

The World Health Organization (WHO) defines osteoporotic bone as bone that is more than 2.5 standard deviations lower in BMD than that of the average 25-year-old adult (called the young-adult mean).¹⁵ Thus, the number of standard deviations below normal, also known as the T-score (e.g., a T-score of -1 indicates a BMD that is 1 SD below the young-adult mean) can be used to define osteoporosis. Osteopenia, considered a precursor to osteoporosis, is defined as a T-score between -1 and -2.5. (WHO). Severe osteoporosis can be defined as a BMD of -2.5 or below that is accompanied by at least one fragility fracture (NOF).

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

NOTE: The following protocol elements are standard procedures for all protocols.

VIII. Review of Key Questions

For Comparative Effectiveness reviews (CERs) the key questions were posted for public comment and finalized after review of the comments. For other systematic reviews, key questions submitted by partners are reviewed and refined as needed by the EPC and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.

IX. Technical Expert Panel (TEP)

A TEP panel is selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. The TEP provides information to the EPC to identify literature search strategies, review the draft report and recommend approaches to specific issues as requested by the EPC. The TEP does not do analysis of any kind nor contribute to the writing of the report.

X. Peer Review (Standard Language)
Approximately five experts in the field will be asked to peer review the draft report and provide comments. The peer reviewer may represent stakeholder groups such as professional or advocacy organizations with knowledge of the topic. On some specific reports such as reports requested by the Office of Medical Applications of Research, National Institutes of Health there may be other rules that apply regarding participation in the peer review process. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

It is our policy not to release the names of the Peer reviewers or TEP panel members until the report is published so that they can maintain their objectivity during the review process.