Screening for Postmenopausal Osteoporosis: A Summary of the Evidence

Heidi D. Nelson, MD, MPH; Mark Helfand, MD, MPH; Steven H. Woolf, MD, MPH; Janet D. Allan, PhD, RN

**Epidemiology**

Half of all postmenopausal women will have an osteoporosis-related fracture during their lives, including one-quarter who will develop a vertebral deformity and 15% who will suffer a hip fracture. Hip fractures are associated with high mortality rates and loss of independence. Although many vertebral fractures are detected only incidentally on radiography, some cause severe pain, leading to 150,000 hospital admissions per year in persons older than 65, 161,000 physician office visits, and more than 5 million days of restricted activity in those 45 years of age or older.

Low bone density has been used to predict risk for fractures as well as to diagnose osteoporosis. Osteoporosis has been defined as "a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, leading to enhanced bone fragility and a consequent increase in fracture risk." This definition emphasizes that, in addition to bone mass, the structure of bone is also an important factor in the pathogenesis of fractures.

A World Health Organization working group proposed that osteoporosis should be diagnosed in epidemiologic studies when bone mineral density is 2.5 standard deviations (SDs) or more below the mean for healthy young adult women at the spine, hip, or wrist (corresponding to a T-score of ≤-2.5) or when patients have a history of an atraumatic fracture. By the World Health Organization definition, up to 70% of women older than 80 have osteoporosis. Age is also an important factor in the relationship between bone density and the absolute risk for fracture. Older women have a much higher fracture rate than younger women with the same bone density because of increasing risk from other factors, such as bone quality and tendency to fall.

Despite the high prevalence of osteoporosis and the effect of fractures on mortality, independence, and quality of life, it is not clear whether it is appropriate to screen asymptomatic postmenopausal...
women. Recent systematic reviews and guidelines disagree about which women should be screened and when.12-20 This disagreement reflects, in part, gaps in the evidence. For example, most guidelines recommend using risk factors to select patients for bone density testing, but because of inadequate data there is no consensus on which risk factors to use.

As part of the U.S. Preventive Services Task Force (USPSTF) update of its 1996 recommendation,21 we examined evidence on the benefits and harms of screening postmenopausal women for osteoporosis. Specifically, we addressed the role of risk factors in identifying high-risk women, techniques of bone density testing to identify fracture risks, effectiveness of treatment in reducing fracture risk, and harms of screening and treatment.

Methods

Additional methods used for this review, including determination of the quality of studies,22 are detailed in Appendix Tables 1, 2, and 3 and in a separate report.23 The analytic framework and key questions are detailed in Appendix Figure 1. Relevant studies were identified from multiple searches of MEDLINE (1966 to May 2001), HealthSTAR (1975 to May 2001), and Cochrane databases; reference lists of systematic reviews; and experts. We also sent letters to manufacturers of bone density devices requesting additional information about the performance of their instruments, but we received no new data.

Two reviewers read each abstract to determine its eligibility. We included English-language abstracts that included original data about postmenopausal women and osteoporosis and addressed screening, or the effectiveness of risk factor assessment, bone density testing, or treatment. We considered screening to be the process of assessing postmenopausal women without known osteoporosis for risk of osteoporotic fractures by identification of risk factors, including low bone density. Postmenopausal women were those who had experienced surgical or natural menopause, regardless of age. Women with pre-existing atraumatic fractures were not considered in the screening population because they had confirmed osteoporosis according to the World Health Organization definition.

For studies of prediction, we selected articles that reported the relationship between risk factor assessment methods or bone density tests and bone density, bone loss, or fractures. We reviewed studies of medications used for treatment, and present results for bisphosphonates. We focused on randomized controlled trials of therapies reporting radiographically verified, nontraumatic fracture outcomes, because fractures are a stronger measure of effectiveness than bone density. We excluded studies of primary prevention of osteoporosis, such as the role of nutrition, calcium consumption, and physical activity. We also excluded secondary causes of osteoporosis, such as corticosteroid use and certain chronic diseases, and studies that did not provide sufficient information to determine the method for selecting patients and for analyzing data. Investigators read the full-text versions of the retrieved papers and re-applied the initial eligibility criteria. To assess the internal validity of individual studies, we applied a set of criteria developed by the USPSTF (Appendix Table 1).22

In this paper, we highlight studies that are applicable to current practice standards, have high quality internal validity ratings, and are most generalizable to the U.S. population of postmenopausal women under consideration for screening. We created an outcomes table to summarize the number of hip and vertebral fractures prevented based on age-specific prevalence rates and treatment effects obtained from results of the reviewed studies. We conducted a sensitivity analysis to determine the influence of risk factors on the number needed to screen.

This research was funded by the Agency for Healthcare Research and Quality (AHRQ) under a contract to support the work of the USPSTF. Agency staff and Task Force members participated in the initial design of the study and reviewed interim analyses and the final manuscript. Additional reports were distributed for review to content experts and revised accordingly before preparation of this manuscript.23,24
Results

Studies of Screening

We identified no studies about the effectiveness of screening in reducing osteoporotic fractures. Without direct evidence from screening studies, recommendations about screening need to rely on evidence that risk factor assessment or bone density testing can adequately identify women who could ultimately benefit from treatment.

Risk Factor Assessment

Hundreds of studies report associations between risk factors and low bone density and fractures in postmenopausal women. The most comprehensive study of risk factors in a U.S. population is the Study of Osteoporotic Fractures (SOF), a good-quality prospective study of 9,516 women 65 years of age and older. In this study, 14 clinical risk factors were identified as significant predictors of hip fracture in multivariable models (age; maternal hip fracture; no weight gain; height; poor self-rated health; hyperthyroidism; current use of benzodiazepines, anticonvulsants, or caffeine; not walking for exercise; lack of ambulation; inability to rise from a chair; poor scores on 2 measures of vision; high resting pulse; and any fracture since 50 years of age). The relative risk for hip fracture per decrease of 1 SD in calcaneal bone density was 1.6 (95% confidence interval [CI], 1.3-1.9). This was comparable to the magnitude of the relative risks of most of the other significant predictors in the model, which ranged from 1.2 to 2.0. Women with at least 5 of the 14 risk factors had increased rates of hip fractures compared with those who had 0 to 2 risk factors at all levels of calcaneal bone density.

To determine which risk factors could be important in women younger than 65, we reviewed 8 observational studies of risk factors and fractures of various types conducted in populations in which at least 50% of participants were younger than 65. Table 1 lists risk factors that were statistically significant predictors for fractures in multivariable models. These results could not be quantitatively combined because risk factors were defined differently in each study.

Bone Density Tests

Several technologies are available to measure bone density, although correlations among different bone density devices are low (0.35 to 0.60). Dual-energy x-ray absorptiometry (DXA) is considered the gold standard because it is the most extensively validated test against fracture outcomes.
<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Relative Risk (95% confidence interval)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age: per 2 years</td>
<td>1.11 (1.01-1.21)</td>
<td>26</td>
</tr>
<tr>
<td>per 5 years</td>
<td>1.94 (1.55-2.42)</td>
<td>27</td>
</tr>
<tr>
<td>Body Mass Index</td>
<td></td>
<td></td>
</tr>
<tr>
<td>per increase of 10 kg/m²</td>
<td>0.58 (0.36-0.92)</td>
<td>27</td>
</tr>
<tr>
<td>≥25.6</td>
<td>0.7 (0.5-0.9) wrist; 1.6 (1.0-2.4) ankle</td>
<td>26</td>
</tr>
<tr>
<td>≥28.6</td>
<td>0.5 (0.4-0.7) wrist; 2.0 (1.3-3.1) ankle</td>
<td>26</td>
</tr>
<tr>
<td>low</td>
<td>1.1 (1.0-1.2)</td>
<td>29</td>
</tr>
<tr>
<td>Height: per 0.1 m</td>
<td>1.58 (1.18-2.12)</td>
<td>27</td>
</tr>
<tr>
<td>Mother with fracture</td>
<td>1.27 (1.16-1.40)</td>
<td>30</td>
</tr>
<tr>
<td>Grandmother with hip fracture</td>
<td>3.70 (1.55-8.85)</td>
<td>31</td>
</tr>
<tr>
<td>Hormone replacement therapy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current use</td>
<td>0.82 (0.74-0.91)</td>
<td>30</td>
</tr>
<tr>
<td>per 5 years use</td>
<td>0.5 (0.2-0.9)</td>
<td>28</td>
</tr>
<tr>
<td>&gt;2 years use</td>
<td>0.44 (0.22-0.89)</td>
<td>32</td>
</tr>
<tr>
<td>Long history of use</td>
<td>0.70 (0.50-0.96)</td>
<td>33</td>
</tr>
<tr>
<td>African American</td>
<td>0.54 (0.41-0.72)</td>
<td>30</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>9.17 (3.38-24.92)</td>
<td>27</td>
</tr>
<tr>
<td>Chronic conditions</td>
<td>1.3 (1.1-1.5)</td>
<td>26</td>
</tr>
<tr>
<td>Disability pension</td>
<td>3.79 (2.15-6.68)</td>
<td>27</td>
</tr>
<tr>
<td>Long-term work disability</td>
<td>1.3 (1.1-1.6)</td>
<td>26</td>
</tr>
<tr>
<td>Self-rated health (fair/poor)</td>
<td>1.79 (1.52-2.11)</td>
<td>30</td>
</tr>
<tr>
<td>Moderate daily physical activity</td>
<td>0.61 (0.37-0.99)</td>
<td>32</td>
</tr>
<tr>
<td>Alcohol: ≥100 g/wk</td>
<td>1.70 (1.08-2.67)</td>
<td>33</td>
</tr>
<tr>
<td>Regular use</td>
<td>1.4 (1.3-4.4)</td>
<td>29</td>
</tr>
<tr>
<td>1 to 6 drinks/wk</td>
<td>0.85 (0.75-0.96)</td>
<td>30</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>1.5 (1.3-1.5); 1.14 (1.00-1.30)</td>
<td>26, 30</td>
</tr>
<tr>
<td>Former</td>
<td>1.09 (1.00-1.19)</td>
<td>30</td>
</tr>
<tr>
<td>≥11 cigarettes/day</td>
<td>3.0 (1.9-4.6)</td>
<td>26</td>
</tr>
<tr>
<td>Unmarried</td>
<td>2.16 (1.28-3.64)</td>
<td>27</td>
</tr>
<tr>
<td>College education or higher</td>
<td>1.26 (1.16-1.38)</td>
<td>30</td>
</tr>
<tr>
<td>Age at menopause</td>
<td>0.94 (0.88-0.99)</td>
<td>32</td>
</tr>
<tr>
<td>Years since menopause</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 to 19</td>
<td>1.18 (1.01-1.38)</td>
<td>30</td>
</tr>
<tr>
<td>20 to 29</td>
<td>1.31 (1.12-1.54)</td>
<td>30</td>
</tr>
<tr>
<td>30</td>
<td>1.51 (1.26-1.81)</td>
<td>30</td>
</tr>
<tr>
<td>Oophorectomy &lt;45 years</td>
<td>3.64 (1.01-13.04)</td>
<td>33</td>
</tr>
<tr>
<td>5 or more children</td>
<td>2.5 (1.1-6.7)</td>
<td>29</td>
</tr>
</tbody>
</table>
When used in the same patients, DXA machines from different manufacturers differ in the proportion of patients diagnosed to have osteoporosis by 6% to 15%.\textsuperscript{80-85} Published studies consistently show that the probability of receiving a diagnosis of osteoporosis depends on the choice of test and site.\textsuperscript{86-90} One analytical study, for example, found that 6% of women older than 60 would receive a diagnosis of osteoporosis if DXA of the total hip were used as the only test, compared with 14% for DXA of the lumbar spine, 3% with quantitative ultrasonography, and 50% with quantitative computed tomography.\textsuperscript{87}

The likelihood of receiving a diagnosis of osteoporosis also depends on the number of sites tested. Testing in the forearm, hip, spine, or heel generally identifies different groups of patients. For example, a physician cannot definitively say that a patient does not have osteoporosis on the basis of a forearm test alone. Conversely, although test results at any site are associated to some degree with fractures at other sites, a physician may not be able to assess whether a patient with a low T-score on a hand or forearm test has substantial bone loss at other sites.

A meta-analysis assessed 23 publications from 11 separate prospective cohort studies published before 1996.\textsuperscript{91} Nearly all of the data were from women in their late 60s or older. No studies of ultrasonography were included. The meta-analysis indicated that DXA at the femoral neck predicted hip fracture better than measurements at other sites, and was comparable to forearm measurements for predicting fractures at other sites.\textsuperscript{92-94} For bone density measurements at the femoral neck, the pooled relative risk per decrease of one SD in bone density was 2.6 (CI, 2.0-3.5). In direct comparisons, heel ultrasonography was slightly worse than but comparable to DXA of the hip in women older than 65 (Table 3).\textsuperscript{92-94,100} For both tests, a result in the osteoporotic range is associated with an increased short-term probability of hip fracture. No data compare DXA and ultrasonography for prediction of fracture in women younger than 65.

The National Osteoporosis Risk Assessment study\textsuperscript{85} recently evaluated the performance of peripheral densitometry in predicting fractures. This prospective study of ambulatory postmenopausal women 50 years of age or older with no previous osteoporosis diagnoses recruited 200,160 participants from 4,236 primary care practices in 34 U.S. states. Women received baseline T-scores by measuring bone density at the heel (single-energy x-ray absorptiometry or quantitative ultrasonography), forearm (peripheral DXA), or finger (peripheral DXA). After 12 months of follow-up, women with T-scores less than or equal to -2.5 had an adjusted risk for all types of fractures that was 2.74 (CI, 2.40-3.13) times higher than women with normal baseline bone density. Results varied by type of test and site; those identified as osteoporotic by DXA had higher fracture rates. Tests were not compared with DXA of the femoral neck, and the study did not describe how tests performed by age group or risk category.

### Treatment

The U.S. Food and Drug Administration has approved hormone replacement therapy, bisphosphonates, raloxifene, and calcitonin for osteoporosis prevention and/or treatment. Our review of estrogen and selective estrogen receptor modulators is presented elsewhere.\textsuperscript{101}

A recent meta-analysis\textsuperscript{102} of 11 randomized trials\textsuperscript{103-113} enrolling 12,855 women found that at least 5 mg of alendronate per day reduced vertebral fractures in 8 trials (relative risk, 0.52; CI, 0.43-0.65). Alendronate also reduced forearm fractures in 6 trials involving 3,723 participants (dosage \(\geq 10\) mg/d; weighted relative risk, 0.48; CI, 0.29-0.78), hip fractures in 11 trials involving 11,808 participants (dosage \(\geq 5\) mg/d; weighted relative risk, 0.63; CI, 0.43-0.92), and other nonvertebral fractures in 6 trials involving 3,723 participants (dosage 10 to 40 mg/d; weighted relative risk, 0.51; CI, 0.38-0.69). These trials included follow-up data ranging from 1 to 4 years; effect sizes for longer periods of use are not known. We evaluated data from these trials to determine whether women who have a similar overall risk for fracture but different bone densities derive a similar benefit from treatment. This question is clinically important because accepted criteria for initiating treatment are lacking.
<table>
<thead>
<tr>
<th>Author</th>
<th>Design</th>
<th>N</th>
<th>Validated</th>
<th>Risk Factors Included</th>
<th>Outcome</th>
<th>Performance</th>
<th>Quality Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slemenda, 1990</td>
<td>Cross-sectional</td>
<td>124</td>
<td>No</td>
<td>Age, height, weight, calcium intake, caffeine intake, alcohol and tobacco use, urinary</td>
<td>Correct classification of high or low BMD (lowest third of subjects)</td>
<td>Midshaft radius: 68% low, 77% high. Lumbar spine: 61% low, 45% high. Femoral neck: 66% low, 93% high.</td>
<td>Poor</td>
</tr>
<tr>
<td>Falch, 1992</td>
<td>Cross-sectional</td>
<td>73</td>
<td>Yes</td>
<td>Low body weight, reduced renal phosphate reabsorption, smoking.</td>
<td>Bone loss</td>
<td>Sensitivity 36%, specificity 89%, PPV 74%.</td>
<td>Poor</td>
</tr>
<tr>
<td>Ribot, 1992</td>
<td>Cross-sectional</td>
<td>1565</td>
<td>No</td>
<td>Weight, menopause, duration of menopause.</td>
<td>Vertebral BMD &lt; -2 SD</td>
<td>Sensitivity 73%, specificity 66%.</td>
<td>Fair</td>
</tr>
<tr>
<td>Michaelsson, 1996</td>
<td>Cross-sectional</td>
<td>175</td>
<td>No</td>
<td>Weight &gt; 70kg.</td>
<td>Femoral neck BMD &lt; -2.5 SD</td>
<td>Sensitivity 94%, specificity 36%, PPV 21%, NPV 97%.</td>
<td>Fair</td>
</tr>
<tr>
<td>Verhaar, 1998</td>
<td>Cross-sectional</td>
<td>61</td>
<td>No</td>
<td>1. Arm span-height difference of at least 3 cm.</td>
<td>BMD ≤ -2.5 SD and vertebral fracture</td>
<td>Arm span only: sensitivity 58%, specificity 56%. Arm span, age, arm span length: sensitivity 81%, specificity 64%.</td>
<td>Poor</td>
</tr>
<tr>
<td>Ballard, 1998</td>
<td>Cross-sectional</td>
<td>1158</td>
<td>No</td>
<td>Age, age at menopause, height, weight, gravidity, parity, current use of steroids, current</td>
<td>Osteoporosis of femoral neck and/or spine</td>
<td>ROC area 0.73.</td>
<td>Fair</td>
</tr>
<tr>
<td>Lydick, 1998</td>
<td>Cross-sectional</td>
<td>1279</td>
<td>Yes</td>
<td>SCORE = age (3 times first digit of age in years), weight (-1 times weight in pounds divided by 10 and truncated to integer), race (5 if not black), estrogen use (1 if never used), rheumatoid arthritis (4 if present), history of fractures (4 for each fracture after age 45 of wrist, hip, or rib, to a maximum of 12).</td>
<td>Femoral neck BMD ≤ -2 SD</td>
<td>Sensitivity 89%, specificity 50%; ROC area 0.81 using a score of 6, or greater.</td>
<td>Good</td>
</tr>
</tbody>
</table>

Table 2. Studies of risk factors assessment

Continued on page 225
<table>
<thead>
<tr>
<th>Author</th>
<th>Design</th>
<th>N</th>
<th>Validated</th>
<th>Risk Factors Included</th>
<th>Outcome</th>
<th>Performance</th>
<th>Quality Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goemaere, 1999</td>
<td>Cross-sectional</td>
<td>300</td>
<td>No</td>
<td>18-item questionnaire of risk factors for osteoporosis (race, height loss, age, weight, smoking, coffee, alcohol, dairy product use, activity, family history, existence of comorbidities, history of wrist fracture, menopause before 45 years, corticosteroid use)</td>
<td>Lumbar spine, femoral neck, and hip BMD</td>
<td>Lumbar spine: ROC area 0.66; Femoral neck: ROC area 0.69; Hip: ROC area 0.76.</td>
<td>Fair</td>
</tr>
<tr>
<td>Cadarette, 2000</td>
<td>Cross-sectional</td>
<td>926</td>
<td>Yes</td>
<td>ORAI = age (15 points if ≥75, 9 if 65-74.5 if 55-64), weight (9 if &lt;60kg, 3 if 60-69) kg), current use of HRT (2 if not currently using).</td>
<td>Hip or lumbar spine BMD ≤ -2.5</td>
<td>Sensitivity 95%, specificity 41% using a score of ≥9.</td>
<td>Good</td>
</tr>
<tr>
<td>Kleerekoper, 1989</td>
<td>Case-control</td>
<td>663</td>
<td>No</td>
<td>Model 1: total months of lactation, family history of osteoporosis, years post menopause, weight. Model 2: breast fed, surgical menopause, age at menarche, age, smoking status.</td>
<td>Vertebral fractures</td>
<td>Model 1: ROC area (SE) 0.55 (0.07); sensitivity 56%; specificity 54%. Model 2: ROC 0.51 (0.042); sensitivity 63% specificity 39%.</td>
<td>Fair</td>
</tr>
<tr>
<td>van Hemert, 1990</td>
<td>Cohort</td>
<td>1014</td>
<td>No</td>
<td>Age, metacarpal cortical area, relative cortical area, BMI, height, diameter of forearm, diameter of knee, age at menarche, age at menopause, smoking, number of children, period of lactation.</td>
<td>Osteoporotic fractures sensitivity 82%.</td>
<td>Sensitivity 48%, specificity 50%.</td>
<td>Fair</td>
</tr>
<tr>
<td>Cooper, 1991</td>
<td>Case-control</td>
<td>1012</td>
<td>No</td>
<td>Age, height, vertebral fracture after age 45, age of last menstrual period, number of children, ever use oral corticosteroid.</td>
<td>Vertebral fractures</td>
<td>Sensitivity 51%, specificity 69%.</td>
<td>Fair</td>
</tr>
<tr>
<td>Wolinsky and Fitzgerald, 1994</td>
<td>Cohort</td>
<td>368</td>
<td>No</td>
<td>White race, female gender, living in southern U.S., age, having been hospitalized in the previous year, previous fall, body mass.</td>
<td>Hip fractures</td>
<td>ROC 0.71; sensitivity 64.7%, specificity 65.7%.</td>
<td>Fair</td>
</tr>
</tbody>
</table>

Continued on page 226
<table>
<thead>
<tr>
<th>Author</th>
<th>Design</th>
<th>N</th>
<th>Validated</th>
<th>Risk Factors Included</th>
<th>Outcome</th>
<th>Performance</th>
<th>Quality Rating*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnell, 1995&lt;sup&gt;16&lt;/sup&gt;</td>
<td>Case-control</td>
<td>5618</td>
<td>No</td>
<td>Late menarche, poor mental score, low BMI, low physical activity, low exposure to sunlight, and low consumption of calcium and tea.</td>
<td>Hip fractures</td>
<td>Sensitivity 55%, specificity 65%</td>
<td>Fair</td>
</tr>
<tr>
<td>Ranstam, 1996&lt;sup&gt;19&lt;/sup&gt;</td>
<td>Case-control</td>
<td>7474</td>
<td>No</td>
<td>Mental-functional risk score: knowledge of the day of week, knowledge of age, ability to wash, ability to dress.</td>
<td>Hip fractures</td>
<td>A less than perfect score had a sensitivity 46%, specificity 79%.</td>
<td>Fair</td>
</tr>
<tr>
<td>Tromp, 1998&lt;sup&gt;50&lt;/sup&gt;</td>
<td>Cohort</td>
<td>1469</td>
<td>No</td>
<td>Female gender, living alone, past fractures, inactivity, height, use of analgesics.</td>
<td>Probability of fractures</td>
<td>No predictors = 0%; 4 predictors = 12.9%</td>
<td>Fair</td>
</tr>
<tr>
<td>Burger, 1999&lt;sup&gt;51&lt;/sup&gt;</td>
<td>Cohort</td>
<td>5208</td>
<td>No</td>
<td>Model with BMD: age, gender, height, use of a walking aid, current smoking, BMD of femoral neck. Model without BMD: age, gender, height, use of a walking aid, current smoking, weight.</td>
<td>Hip fractures</td>
<td>Model with BMD: ROC area 0.88; sensitivity 70%, specificity 84%. Model without BMD: ROC area 0.83; sensitivity 70%, specificity 83%.</td>
<td>Fair</td>
</tr>
</tbody>
</table>

*Based on criteria developed by the U.S. Preventive Services Task Force (22).

**Note:** BMD indicates bone mineral density; BMI, body mass index; CI, confidence interval; HRT, hormone replacement therapy; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating characteristic (values ≥0.80 are usually required to consider a text to be effective); RR, relative risk; SE, standard error.
The Fracture Intervention Trial (FIT) of alendronate was conducted with 2 groups of participants and provides some information about levels of risk. One group (FIT-I) included a higher-risk sample of 2,027 women who had T-scores of -1.6 or lower and pre-existing vertebral fractures. The 3-year risk for hip fracture was 2.2% in the placebo group and 1.1% in the alendronate group (relative hazard, 0.49; CI, 0.23-0.99), and the 3-year risk for any clinical fracture was 18.2% in the placebo group and 13.6% in the alendronate group (relative hazard, 0.72; CI, 0.58-0.90). A second study from FIT (FIT-II) included a lower-risk sample of 4,432 women who also had T-scores of -1.6 or lower, but did not have pre-existing vertebral fractures. The 4-year incidences of hip fracture (1.1%) and any clinical fractures (14.1%) in the placebo group were lower than those observed in the FIT-I placebo group. In FIT-II, only the subgroup of treated patients who had T-scores lower than -2.5 (n = 1,627) had a significant risk reduction for all clinical fractures, from 19.6% to 13.1% (relative risk, 0.64; CI, 0.50-0.82). No reduction in risk for fractures was seen in patients who had T-scores between -1.6 and -2.5.

The results from FIT suggest that women with more risk factors for fracture relating to bone structure and integrity, such as age, very low bone density, or pre-existing vertebral fractures, derive the greatest absolute benefit from treatment. However, FIT did not examine other nonskeletal risk factors, such as psychomotor impairment, poor gait, and other factors that increase the risk for falling. The effect of some of these risk factors on the benefit of treatment was examined in a randomized trial of another bisphosphonate, risedronate. Risedronate had no effect on hip fracture rates among women 80 years of age or older who had one or more risk factors for falls but who did not necessarily have low bone density. In the same report, in women 70 to 79 years of age with severe osteoporosis (T score < -3), risedronate reduced hip fractures by 40% (relative risk, 0.6; CI, 0.4-0.9; number needed to treat for benefit, 77).

Trial results are not applicable to a screening program unless the trials included patients who would be identified by screening the general population. We examined recruitment and eligibility characteristics of 10 of the 11 published randomized trials of alendronate to assess whether selection biases or other biases might have affected their generalizability (Table 4). Overall, the trials included relatively healthy women with low bone density who were not using estrogen. Except for participants in 2 trials involving women who had recently gone through menopause and were not osteoporotic, most participants were older than 65.

The FIT-II is the largest study and provided the most detailed description of recruitment and results. In FIT-II, researchers recruited the sample of 4,432 women by mailing a query to more than 1 million women selected from the general population in 11 cities. Women who had medical problems (for example, dyspepsia) or who used estrogen were excluded. Fifty-four thousand women (approximately 5.4%) responded by telephone; of these, 26,137 (52%) had a screening visit. A higher than expected proportion of these (65%) had sufficiently low bone density to enroll in the study. Of this 65%, 57% were classified as “ineligible, did not wish to continue, or screened after recruitment to this arm.” It is not clear from this description how many patients did not meet the eligibility criteria. In addition, an unspecified number of patients (up to 28,000) were found to be ineligible at the initial stage of recruitment. The demographic characteristics of eligible and screened but excluded participants were not reported. None of the other randomized trials disclosed any details of how their samples were recruited or how many respondents were found to be ineligible before randomization.

In other clinical areas, the results of industry-sponsored trials were significantly more favorable to newer therapies than trials funded by nonprofit organizations. Because all 11 trials of alendronate were funded wholly or in part by the manufacturer, we were unable to assess the influence of sponsorship on effect size. If effectiveness of treatments is less than estimated in these trials, the efficiency of screening to identify candidates for treatment will be reduced and the number needed to screen for benefit will increase.
Harms

Several potential harms are associated with screening and treatment. A test result indicating osteoporosis could produce anxiety and perceived vulnerability that may be unwarranted. On a quality-of-life questionnaire, women with osteoporosis voiced significantly more fears than women who had normal bone density. Some women may be falsely reassured if abnormal results from last year’s DXA appear “improved” on this year’s normal calcaneal ultrasonogram. The potential time, effort, expense, and radiation exposure of repeated scans over a lifetime have not yet been determined.

Potential harms may also arise from inaccuracies and misinterpretations of bone density tests. The variation among techniques, along with the lack of methods to integrate bone density results with clinical predictors, makes it difficult for clinicians to provide accurate information to patients about test results. In 1 study, physicians found densitometry reports confusing, and were not confident that their interpretations of T-scores were accurate. False-positive results could lead to inappropriate treatment, and false-negative results could lead to missed treatment opportunities.

Harms of treatment depend on the medication used. Overall, gastrointestinal side effects occur in approximately 25% of patients taking alendronate, but in controlled trials these rates were usually not significantly higher than those for placebo. High rates of serious gastrointestinal side effects have been observed among Medicare enrollees taking alendronate. The long-term adverse effects of alendronate are unknown.

Costs of screening vary with technique, and average 2000 Medicare reimbursement rates were $133 for DXA and $34 for ultrasonography. Abnormal results on ultrasonography may require a confirmatory DXA before treatment because clinical trials are based on entry criteria using DXA. Most women would require repeated tests over several years.

---

### Table 3. Prospective studies of DXA and ultrasound reporting hip fractures

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Population</th>
<th>Age range</th>
<th>Followup (years)</th>
<th>N</th>
<th>Probability of low risk DXA of the hip</th>
<th>Probability of high risk DXA of the hip</th>
<th>Probability of low risk QUS of the heel</th>
<th>Probability of high risk QUS of the heel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study of Osteoporotic fractures</td>
<td>Community-dwelling white women from 4 areas in the U.S. recruited from lists</td>
<td>65-69</td>
<td>1.8-2.9</td>
<td>5236</td>
<td>0.009</td>
<td>0.005</td>
<td>0.006</td>
<td>0.018</td>
</tr>
<tr>
<td>(SOP)18,24-26</td>
<td></td>
<td>65-79</td>
<td>2.9</td>
<td>2371</td>
<td>0.003</td>
<td>0.0028</td>
<td>0.006</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥80</td>
<td>2.9</td>
<td>3013</td>
<td>0.0007</td>
<td>0.0005</td>
<td>0.003</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65-69</td>
<td>1.8</td>
<td>1728</td>
<td>0.018</td>
<td>0.0076</td>
<td>0.0007</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td></td>
<td>70-74</td>
<td>1.8</td>
<td>731</td>
<td>0.024</td>
<td>0.0024</td>
<td>0.014</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75-79</td>
<td>1.8</td>
<td>291</td>
<td>0.028</td>
<td>0.012</td>
<td>0.012</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥85</td>
<td>1.8</td>
<td>5656</td>
<td>0.002</td>
<td>0.0033</td>
<td>0.008</td>
<td>0.012</td>
</tr>
<tr>
<td>Epidemiologie de L’Osteroporose</td>
<td>Women from 5 cities in France recruited from voting lists and health insurance companies.</td>
<td>≥75</td>
<td>2</td>
<td>3982</td>
<td>0.013</td>
<td>0.002</td>
<td>0.002</td>
<td>0.025</td>
</tr>
<tr>
<td>(EPIDOS)97-100</td>
<td></td>
<td>&lt;80</td>
<td>2</td>
<td>3616</td>
<td>0.0028</td>
<td>0.006</td>
<td>0.006</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥80</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Probability of hip fracture if bone density was classified as high or low risk.

**Note:** DXA indicates Dual energy x-ray absorptiometry; QUS, quantitative ultrasound.
Table 4. Randomized controlled trials of alendronate with fracture outcomes

<table>
<thead>
<tr>
<th>Author</th>
<th>Duration (years)</th>
<th>Age (years)</th>
<th>Population</th>
<th>Exclusion Criteria*</th>
<th>Lost to Followup</th>
<th>Rating†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adami, 1995103</td>
<td>2</td>
<td>48-76</td>
<td>9 Italian Centers, T-score &lt; -2 (0.67g/cm²); 5% vertebrae fractures.</td>
<td>Narrow</td>
<td>32/211(15.2%)</td>
<td>Fair to Good</td>
</tr>
<tr>
<td>Black, 1996104</td>
<td>3</td>
<td>55-81</td>
<td>11 U.S. cities, BMD &lt;0.68 g/cm²; no previous vertebral fractures.</td>
<td>Broad (medical illness, dyspepsia, etc)</td>
<td>81/2,027 (4%)</td>
<td>Good</td>
</tr>
<tr>
<td>Bone, 1997105</td>
<td>2</td>
<td>&gt;60</td>
<td>15 U.S. sites, BMD &lt;0.84 g/cm²; average 20 yrs since menopause; 30.7% vertebral fractures.</td>
<td>Broad (medical illness, NSAIDs, GI drugs)</td>
<td>19/359 (5.3%)</td>
<td>Fair to Good</td>
</tr>
<tr>
<td>Chesnut, 1995106</td>
<td>2</td>
<td>42-75</td>
<td>7 centers, spine BMD &lt;0.88, average hip BMD 0.7; ≥ 5 years since menopause.</td>
<td>Broad</td>
<td>26/157 (16.6%)</td>
<td>Fair</td>
</tr>
<tr>
<td>Cummings, 1998107</td>
<td>4</td>
<td>55-81</td>
<td>11 U.S. cities, BMD &lt;0.68 g/cm² (ave. 0.59); no previous vertebral fractures.</td>
<td>Broad (medical illness, dyspepsia)</td>
<td>179/4,432 (4%)</td>
<td>Good</td>
</tr>
<tr>
<td>Greenspan, 1998108</td>
<td>2.5</td>
<td>&gt;65</td>
<td>1 Boston center, no BMD entry criteria.</td>
<td>Narrow (“good health”)</td>
<td>33/120 (27.5%)</td>
<td>Fair</td>
</tr>
<tr>
<td>Hosking, 1998109</td>
<td>4</td>
<td>45-59</td>
<td>4 centers, BMD &gt; 0.8 g/cm²; &lt;10% prevalent vertebral fractures.</td>
<td>Narrow (“good health”)</td>
<td>287/1,499 (19.1%)</td>
<td>Fair</td>
</tr>
<tr>
<td>Liberman, 1995110</td>
<td>3</td>
<td>45-80</td>
<td>2 multicenter trials, T-score &lt; -2.5; 21% prevalent vertebral fractures.</td>
<td>Narrow (“good health”)</td>
<td>170/994 (17.1%)</td>
<td>Good</td>
</tr>
<tr>
<td>McClung, 1998111</td>
<td>3</td>
<td>40-59</td>
<td>15 centers, T-score &lt; -2; 6-36 months since menopause; no previous vertebral fractures.</td>
<td>Narrow (“good health”, estrogen use)</td>
<td>31% at 3 years</td>
<td>Fair</td>
</tr>
<tr>
<td>Pols, 1999112</td>
<td>1</td>
<td>40-82</td>
<td>153 centers, T-score &lt; -2.8.</td>
<td>Narrow (“good health”)</td>
<td>211/1,908 (11.1%)</td>
<td>Fair</td>
</tr>
</tbody>
</table>

*In general, “narrow” criteria excluded estrogen users and patients with illnesses affecting bone metabolism.
†Based on criteria developed by the U.S. Preventive Services Task Force.22

years before receiving a diagnosis of osteoporosis and leaving the screening pool. Treatment costs also vary; alendronate currently costs approximately $3 per daily dose.

### Screening Strategies

To estimate the effect of screening 10,000 postmenopausal women for osteoporosis on reducing hip and vertebral fractures, we created an outcomes table based on assumptions from the reviewed studies (Table 5). These estimates include age-specific prevalence rates expressed in 5-year age intervals123 and treatment effects based on trial results (risk reduction, 37% for hip fracture and 50% for vertebral fracture).102,104,115,124 We estimated an adherence rate of 70% based on reports of adherence and side effects from treatment trials, assuming less optimal compliance in the general population.

When the assumptions in Table 5 are used, if 10,000 women 65 to 69 years of age underwent
bone densitometry (DXA of the femoral neck), 1,200 would be identified as high-risk (T-score ≤ -2.5). If these women were offered treatment that resulted in a 37% reduction in hip fracture risk, and a 50% reduction in vertebral fracture risk, and 70% adhered to therapy, then 14 hip fractures and 40 vertebral fractures would be prevented over a 5-year period. The number of women in this age group needed to screen to prevent 1 hip fracture in 5 years would be 731, and the number of women with low bone density needed to treat for benefit would be 88. The number needed to screen to prevent 1 vertebral fracture would be 248, and the number needed to treat for benefit would be 30. Treatment has significant costs and potential harms; when the number-needed-to-screen for benefit is high, the balance of benefits and harms may become unfavorable. These numbers become more favorable in older persons because the prevalence of osteoporosis increases steadily with age.

There is interest in whether risk assessment can be used to select patients for bone densitometry, which is costly. Our literature review indicated that the prevalence of osteoporosis, the predictability of densitometry, and the effectiveness of treatment might be lower for younger than for older postmenopausal women. To determine whether it is useful to consider clinical risk factors when screening younger postmenopausal women, we also included risk estimates for clinical risk factors in a sensitivity analysis. Our review of observational studies with younger postmenopausal women indicated that 3 consistent predictors of fracture are increasing age, low weight or body mass index, and nonuse of hormone replacement therapy (defined by current use, ever use, or certain durations of use). These 3 variables are also used in ORAI to identify women with low bone density and were the variables most strongly associated with low bone density in a study enrolling mostly younger postmenopausal women in the United States. On the basis of these studies, we estimated that 1 of these risk factors increases the probability of having osteoporosis by up to 100% and increases the risk for fracture by 70% (relative risk, 1.7).

For the younger age groups, the presence of clinical risk factors influences the outcomes. For example, only 5 hip fractures are prevented over 5 years when all women 60 to 64 years of age are screened; however, 9 hip fractures are prevented if women have a factor that increases fracture risk by 70%. For women 60 to 64 years of age who have such a risk factor, the number needed to screen is 1,092 and the number needed to treat for benefit is 72 to prevent 1 hip fracture. These numbers approach those of women 65 to 69 years of age (Figure 1).

**Discussion**

Although many studies have been published about osteoporosis in postmenopausal women, no trials have evaluated the effectiveness of screening; therefore, no direct evidence that screening improves outcomes is available. Instruments developed to assess clinical risk factors for low bone density or fractures generally have moderate to high sensitivity and low specificity. Many have not been validated, and none have been widely tested in a practice setting. Among different bone density tests measured at various sites, bone density measured by DXA at the femoral neck is the best predictor of hip fracture and is comparable to forearm measurements for predicting fractures at other sites. Heel ultrasonography and other peripheral bone density tests, however, can also predict short-term fracture risk. Bisphosphonates decrease fracture risk by approximately 40% to 50% in women with low bone density.

Support for population screening would be based on evidence that the prevalences of osteoporosis and fractures increase with age, that the short-term risk for fracture can be estimated by bone density tests and risk factors, and that the fracture risk among women with low bone density can be significantly reduced with treatment. We applied these data to generate an outcomes table of screening strategies that provide estimates of the numbers of women needed to screen and treat to prevent fractures. Age-based screening is supported by prevalence data, that is, the number needed to screen to prevent fractures decreases sharply as age and prevalence increase. Use of risk factors, particularly increasing age, low weight, and nonuse of estrogen replacement, to
screen younger women may identify additional high-risk women and provide absolute benefit similar to that yielded by screening older women without risk factors. These findings relate to screening asymptomatic women only and do not apply to women considered for testing because of pre-existing or new fractures or the presence of secondary causes of osteoporosis.

Our approach has several limitations, however, and results from a well-designed trial of screening strategies should supersede our estimations which are based on indirect evidence. The estimates in the outcomes table are limited by assumptions that are arguable or highly variable by patient and setting. Our assumptions of treatment effect and adherence are especially optimistic and reflect results of clinical trials, not clinical practice. We chose a 5-year time horizon based on the short-term predictability of bone density tests as well as on results of short-term treatment trials. Long-term outcomes may provide a more accurate estimate of benefits. Also, we cannot exclude the possibility that harms outweigh benefits, particularly since the long-term effects of bisphosphonates are not yet known.

The evidence on which we based our conclusions is also limited. Overall, evidence is stronger for women older than 65 than for younger women because more research has been done in older age groups. Bone loss in the perimenopausal and early postmenopausal years is important to long-term bone health, but few published studies address screening and treatment for younger postmenopausal women. Fracture risk is determined not only by bone density, but also by bone characteristics that are difficult to measure in a clinical setting, such as bone structure and morphologic characteristics. No bone density studies or treatment trials include large numbers of non-white women, and it may be

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age Group</th>
<th>50-54 y</th>
<th>55-59 y</th>
<th>60-64 y</th>
<th>65-69 y</th>
<th>70-74 y</th>
<th>75-79 y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base Case Assumptions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prevalence of osteoporosis</td>
<td></td>
<td>0.0305</td>
<td>0.0445</td>
<td>0.065</td>
<td>0.120</td>
<td>0.2025</td>
<td>0.285</td>
</tr>
<tr>
<td>Relative risk for hip fracture with treatment</td>
<td></td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>Relative risk for vertebral fracture with treatment</td>
<td></td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
<td>0.52</td>
</tr>
<tr>
<td>Adherence to treatment</td>
<td></td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identified as high-risk (osteoporotic)</td>
<td></td>
<td>305</td>
<td>445</td>
<td>650</td>
<td>1,200</td>
<td>2,025</td>
<td>2,850</td>
</tr>
<tr>
<td>Hip fractures prevented</td>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>14</td>
<td>39</td>
<td>70</td>
</tr>
<tr>
<td>NNS to prevent 1 hip fracture</td>
<td></td>
<td>7,446</td>
<td>4,338</td>
<td>1,856</td>
<td>731</td>
<td>254</td>
<td>143</td>
</tr>
<tr>
<td>NNT to prevent 1 hip fracture</td>
<td></td>
<td>227</td>
<td>193</td>
<td>121</td>
<td>88</td>
<td>51</td>
<td>41</td>
</tr>
<tr>
<td>Vertebral fractures prevented</td>
<td></td>
<td>5</td>
<td>7</td>
<td>22</td>
<td>40</td>
<td>95</td>
<td>134</td>
</tr>
<tr>
<td>NNS to prevent 1 vertebral fracture</td>
<td></td>
<td>1,952</td>
<td>1,338</td>
<td>458</td>
<td>248</td>
<td>105</td>
<td>75</td>
</tr>
<tr>
<td>NNT to prevent 1 vertebral fracture</td>
<td></td>
<td>60</td>
<td>60</td>
<td>30</td>
<td>30</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

*Estimates for assumptions include age-specific prevalence rates for osteoporosis and probabilities of fractures; relative risk of 0.63 for hip fractures and 0.52 for vertebral fractures with treatment; treatment adherence of 0.7 (see text).

**Note:** Formulas for calculations are described in Appendix Table 2.

NNS indicates number needed to screen for benefit; NNT, number needed to treat.
difficult to provide ethnicity-specific screening recommendations in the absence of more evidence.

The role of clinical risk factors is still unclear. Although many risk factors are associated with osteoporosis and fractures, how to use them to select women to test or treat is uncertain. The risk factors identified by our literature review and used in the outcomes table are only best estimates. Other risk factors may prove to be equally predictive when used for screening purposes. Further validation of existing risk assessment instruments or development of new ones would be useful. Few studies have evaluated the effect of altering modifiable risk factors, such as smoking cessation, strength and balance training, and visual correction. These interventions may prove to be as effective as drug therapy in preventing fractures, and may also be important effect modifiers that would alter the effectiveness of treatments.

Peripheral bone density tests have not been extensively studied for screening. Results from National Osteoporosis Risk Assessment Study\textsuperscript{30} indicate that peripheral tests can predict short-term fracture rates in a primary care population that would be targeted for screening. Most treatment trials use DXA of the hip as an entry criterion, and results may not apply to women whose diagnosis is determined by other tests. A sequential approach, in which women with low values on a peripheral test are subsequently tested by DXA of the hip to determine treatment needs, may be useful, although this approach has not been evaluated. Further research is needed to define the appropriate use of these technologies.

How frequently to screen has also not been specifically studied, but data are needed to determine optimal screening intervals. Estimations can be made based on the age-specific prevalence of osteoporosis and the precision of bone density tests. Less frequent testing for younger postmenopausal women when prevalence is lower (for example, 5-year intervals) and more frequent testing for older women (for example, 2-year intervals) might be reasonable, but further research is needed. Screening intervals of less than 2 years seem unwarranted because the precision error of densitometry would likely exceed the estimated bone loss in such a brief
period. After a woman is screened and determined to have osteoporosis, future screening with bone density testing would be unnecessary.

Osteoporotic fractures present an enormous health burden on an expanding elderly population. Further research to more accurately determine the benefits and harms of screening is of paramount importance.

**Acknowledgements**

The authors thank Peggy Nygren, MA; Nancy Carney, PhD; Kathryn Pyle Krages, AMLS, MA; Benjamin Chan, MS; and the reviewers of the full evidence report for their contributions to this project.

**References**


Screening for Postmenopausal Osteoporosis


89. Kroger H, Lunt M, Reeve J, et al. Bone density reduction in various measurement sites in men and women with osteoporotic fractures of spine and hip:


Appendix

Appendix Figure 1. Analytic framework

Key Questions
Arrow 1: Does screening using risk factor assessment and/or bone density testing reduce fractures?
Arrow 2: Does risk factor assessment accurately identify women who may benefit from bone density testing?
Arrow 3: Do bone density measurements accurately identify women who may benefit from treatment?
Arrow 4: What are the harms of screening?
Arrow 5: Does treatment reduce the risk of fractures in women identify by screening?
Arrow 6: What are the harms of treatment?

The analytic framework is a schematic outline used to define the population, preventive service, diagnostic or therapeutic interventions, and intermediate and health outcomes considered in the review. The arrows represent key questions that the evidence must answer, and demonstrate the chain of logic that evidence must support, to link the preventive service to improved health outcomes.
### Appendix Table 1. Summary of evidence quality

<table>
<thead>
<tr>
<th>Evidence Linkage (Fig 1)</th>
<th>Key Questions</th>
<th>Evidence Code*</th>
<th>External Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrow 1:</td>
<td>Does screening using risk factor assessment or bone density testing reduce fractures?</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Arrow 2:</td>
<td>Does risk factor assessment accurately identify women who may benefit from bone density testing?</td>
<td>II-2</td>
<td>Poor-fair: no instruments used widely for screening purposes although some were developed from community-based studies.</td>
</tr>
<tr>
<td>Arrow 3:</td>
<td>Do bone density measurements accurately identify women who may benefit from treatment?</td>
<td>II-2</td>
<td>Fair: not known how well results of studies translate to practice.</td>
</tr>
<tr>
<td>Arrow 4:</td>
<td>What are the harms of screening?</td>
<td>II-2, III</td>
<td>Poor: small studies, selected participants.</td>
</tr>
<tr>
<td>Arrow 5:</td>
<td>Does treatment reduce the risk of fractures in women identified by screening?</td>
<td>II</td>
<td>Poor-fair: participants of trials may be different than primary care patients.</td>
</tr>
<tr>
<td>Arrow 6:</td>
<td>What are the harms of treatment?</td>
<td>I, II-2</td>
<td>Poor-fair: difficult to know how risks impact individual patients.</td>
</tr>
</tbody>
</table>

*Evidence codes based on study design categories.22

I = randomized, controlled trials
II-1 = controlled trials without randomization
II-2 = cohort or case-control analytic studies,
II-3 = multiple time series, dramatic uncontrolled experiments
III = opinions of respected authorities, descriptive studies

†Based on criteria developed by the U.S. Preventive Services Task Force.22
### Appendix Table 2. Formulas for calculations in outcomes table

**Number of hip fractures in untreated women with osteoporosis**

*No risk factors:*

\[(5\text{-year probability of hip fracture in women with osteoporosis}) \times (\text{prevalence of osteoporosis}) \times N\]

*At least one risk factor:*

\[1.7 \times (5\text{-year probability of hip fracture in women with osteoporosis}) \times (\text{prevalence of osteoporosis}) \times N\]

**Number of hip fractures in treated women with osteoporosis**

*No risk factors:*

\[(\text{RR for hip fracture from treatment trials}) \times (0.7 \text{ adherence}) \times (\text{number of hip fractures in untreated women with osteoporosis}) + (1 - 0.7 \text{ adherence}) \times (\text{number of hip fractures in untreated women with osteoporosis})\]

*At least one risk factor:*

\[(\text{RR for hip fracture from treatment trials}) \times (0.7 \text{ adherence}) \times (\text{number of hip fractures in untreated women with osteoporosis with at least one risk factor}) + (1 - 0.7 \text{ adherence}) \times (\text{number of hip fractures in untreated women with osteoporosis with at least one risk factor})\]

**Number-needed-to-screen for benefit**

\[N \div (\text{number of hip fractures without treatment} - \text{number with treatment})\]

**Number-needed-to-treat**

\[\text{Number of women with osteoporosis} \div (\text{number of hip fractures without treatment} - \text{number with treatment})\]
The Methods Work Group for the U.S. Preventive Services Task Force developed a set of criteria to determine how well individual studies were conducted (internal validity). The Task Force defined a 3-category rating of “good,” “fair,” and “poor,” based on these criteria. In general, a good study is one that meets all criteria well. A fair study is one that does not meet, or it is not clear that it meets, at least one criterion but has no known important limitation that could invalidate its results. A poor study has important limitations. These specifications are not meant to be rigid rules but rather are intended to be general guidelines, and individual exceptions, when explicitly explained and justified, can be made.

**Randomized Controlled Trials**
- Adequate randomization, including concealment and equal distribution of potential confounders among groups.
- Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination).
- Important differential loss to follow-up or overall high loss to follow-up.
- Equal, reliable, and valid measurements (includes masking of outcome assessment).
- Clear definition of interventions.
- Important outcomes considered.
- Intention-to-treat analysis.

**Case-Control Studies**
- Accurate ascertainment of cases.
- Nonbiased selection of case patients and controls with exclusion criteria applied equally to both.
- High response rate.
- Diagnostic testing procedures applied equally to each group.
- Measurement of exposure accurate and applied equally to each group.
- Appropriate attention to potential confounding variables.

**Cohort Studies**
- Consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts.
- Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination).
- Important differential loss to follow-up or overall high loss to follow-up.
- Equal, reliable, and valid measurements (includes masking of outcome assessment).
- Clear definition of interventions.
- Important outcomes considered.
- Adjustment for potential confounders in analysis.

**Diagnostic Accuracy Studies**
- Screening test relevant, available for primary care, adequately described.
- Study uses a credible reference standard, performed regardless of test results.
- Reference standard interpreted independently of screening test.
- Handles indeterminate results in a reasonable manner.
- Spectrum of patients included in study.
- Adequate sample size
- Administration of reliable screening test.