Long-term Drug Therapy and Drug Holidays for Osteoporosis Fracture Prevention: A Systematic Review

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

This information is distributed solely for the purposes of pre-dissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. The findings are subject to change based on the literature identified in the interim and peer-review/public comments and should not be referenced as definitive. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

Contract No. To be provided in the Final Report

Prepared by:
To be provided in the Final Report

Investigators:
To be provided in the Final Report

AHRQ Publication No. xx-EHCxxx
<Month Year>
Key Messages

Purpose of Review
To assess the current evidence about the efficacy and harms of long-term osteoporosis drug therapy (>3 years) to prevent fractures, of osteoporosis drug continuation versus discontinuation (placebo drug holiday), and whether osteoporosis drug effects differ by patient, bone, and drug characteristics.

Key Messages
- Evidence about efficacy and harms of long-term osteoporosis drug therapy is predominately from U.S.-based studies conducted in white, postmenopausal women.
- In postmenopausal women with osteoporosis, 4 years of alendronate compared with placebo reduces risk of incident clinical fractures and incident radiographic vertebral fractures; whereas, among women with osteopenia, it does not reduce risk of incident clinical fractures and may not reduce risk of incident radiographic vertebral fractures.
- In postmenopausal women with osteoporosis, 4 years of raloxifene compared with placebo reduces risk of incident radiographic vertebral fractures and incident clinical vertebral fractures, but does not lower risk of incident hip or other nonvertebral fractures.
- In postmenopausal women with osteoporosis, 7 years of estrogen versus placebo reduces risk of incident clinical fractures and incident hip fractures.
- In postmenopausal women with an intact uterus and past clinical fractures, but unknown bone mineral density (BMD), 5.6 years of estrogen/progestin versus placebo reduces risk of incident clinical fractures.
- Continued use of zoledronic acid for a total of 6 years versus discontinuation after 3 years lowers risk of incident radiographic vertebral fractures, but does not lower risk of incident nonvertebral fractures.
- Continued use of alendronate for a total of 10 years versus discontinuation after 5 years probably lowers risk of incident clinical vertebral fractures, but does not lower risk of incident nonvertebral fractures.
- Long-term bisphosphonate use may increase risk of atypical femoral fractures with confirmed radiological features (AFF), subtrochanteric or femoral shaft fractures without confirmed radiological AFF features, and osteonecrosis of the jaw (ONJ).
- Future trials should include diverse populations (including male sex, nonwhite race, and individuals with multiple comorbidities), include individuals with high fracture risk who do not meet criteria for osteoporosis, be large enough to compare between-treatment differences in risk of incident clinical fractures, and compare sequential treatments (e.g., anabolic followed by anti-resorptive) and drug holidays of various durations.
- Future observational studies on risk of AFF and ONJ should use a cohort design and define outcomes using standard radiographic criteria.
This report is based on research conducted by an Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. xxx-xxxx-xxxxx). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, policymakers, and others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies, may not be stated or implied.

This report may periodically be assessed for the currency of conclusions. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at www.effectivehealthcare.ahrq.gov. Search on the title of the report.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact epc@ahrq.hhs.gov.

**Suggested citation:** To be provided in the Final Report
Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an email list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officers named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to epc@ahrq.hhs.gov.

Gopal Khanna, M.B.A.         Arlene Bierman, M.D., M.S.
Director                   Director
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.        Laura Pincock, Pharm.D., M.P.H.
Director, EPC Program       Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Jill Huppert, M.D., M.P.H.
Task Order Officer
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality
Acknowledgments

To be provided in the Final Report

Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

The list of Technical Experts who provided input to this report follows:

To be provided in the Final Report
Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report do not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential nonfinancial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

The list of Peer Reviewers for this report follows:

To be provided in the Final Report
Long-term Drug Therapy and Drug Holidays for Osteoporosis Fracture Prevention: A Systematic Review

Structured Abstract

Objective. To summarize evidence on outcomes of long-term osteoporosis drug therapy to prevent fractures, on continuing versus discontinuing osteoporosis drug therapy (i.e., placebo drug holidays), and on whether osteoporosis drug intervention effects vary as a function of patient, bone, or drug characteristics.

Data sources. MEDLINE, Embase and Cochrane databases from 1995 to June 2018; ClinicalTrials.gov; bibliographies of relevant systematic reviews

Review methods. Long-term osteoporosis drug therapy was defined as >3 years and drug holiday as osteoporosis drug discontinuation for ≥1 year after ≥1 year of prior osteoporosis drug use. Two reviewers independently rated risk of bias (ROB) and strength of evidence (SOE), resolving discrepancies by consensus. Included studies were English-language trials for incident fractures and harms and controlled observational studies for additional harms outcomes. For low or medium ROB studies, one reviewer extracted data and a second verified accuracy.

Results. Of 56 eligible publications, 44 had low or medium ROB, including 32 publications of trials (7 unique studies) and 12 publications of observational studies (10 unique studies). Nearly all studies were comprised of postmenopausal women. Mean participant age was 72 years; all but 2 studies had a mean age <80 years. In postmenopausal women with osteoporosis, compared with placebo, 4 years of alendronate or raloxifene reduced risk of incident vertebral fractures (high SOE), 4 years of alendronate reduced risk of incident clinical fractures (moderate SOE). In postmenopausal women with past fractures, compared with placebo, both long-term estrogen and long-term estrogen/progestin reduced risk of incident clinical fractures and long-term estrogen reduced risk of incident hip fractures (all low SOE). Alendronate, denosumab and raloxifene for 4 years each significantly increased total hip and lumbar spine bone mineral density (BMD) versus placebo. Continuation versus discontinuation of alendronate after 5 years reduced risk of incident clinical vertebral fractures in one large trial (10 vs. 5 years, moderate SOE), but not in another smaller trial (7 vs. 5 years, low SOE). Continuation versus discontinuation of zoledronic acid (6 vs. 3 years) reduced risk of incident radiographic vertebral fractures (moderate SOE), but evidence was insufficient about risk of incident clinical vertebral fractures. Neither alendronate nor zoledronic acid continuation reduced risk of nonvertebral fractures (low SOE) versus discontinuation; for both, continuation was associated with generally stable hip BMD compared to small, but significant declines with discontinuation. Based primarily on observational studies, long-term bisphosphonates may increase risks of radiologically confirmed atypical femoral fractures (AFF), subtrochanteric or femoral shaft fractures without confirmed AFF features, and osteonecrosis of the jaw (ONJ).

Limitations. Minimal data for men or individuals with comorbidities. Low power to assess risks of incident clinical fractures. No data compared long-term effects of sequential treatments (e.g., anabolic followed by anti-resorptive) or different durations of drug holidays. Analyses of
possible treatment effect modifiers almost entirely post hoc. Observational studies used variable
drug treatment and control exposures and harms definitions.

Conclusions. For postmenopausal women with osteoporosis by BMD or past fractures, long-
term alendronate and raloxifene reduced risk of incident vertebral fractures; long-term
alendronate, estrogen, and estrogen/progestin reduced risk of clinical fractures; and long-term
estrogen reduces risk of incident hip fractures. Longer-term use of bisphosphonates versus
discontinuation may lower vertebral fracture risk and stabilize hip BMD, but doesn’t reduce
nonvertebral fracture risk and may increase risk of AFF and ONJ. Long-term estrogen and
estrogen/progestin increased risk of cardiovascular disease, and long-term estrogen increased
risk of dementia and breast cancer. To address remaining knowledge gaps, future trials and
observational studies should enroll diverse populations (sex, comorbidity), examine the effects of
sequential treatments and compare drug holidays of different durations, be powered for clinical
fractures, and use standard AFF and ONJ definitions. A priori analyses to examine whether
treatment outcomes vary by patient, bone and drug treatment characteristics may inform
individualized treatment decisions.
Contents

Evidence Summary ............................................................................................................................... ES-1

Chapter 1. Introduction ................................................................................................................... 1
  Background ....................................................................................................................................... 1
  Scope and Key Questions .................................................................................................................. 3
    Key Questions ............................................................................................................................... 3
    PICOTS ......................................................................................................................................... 5
    Analytic Framework .................................................................................................................... 9

Chapter 2. Methods .......................................................................................................................... 10
  Topic Refinement and Review Protocol .......................................................................................... 10
  Literature Search Strategy ............................................................................................................. 10
  Study Selection and Risk of Bias Assessment ................................................................................. 12
  Data Extraction ............................................................................................................................... 12
  Data Synthesis ............................................................................................................................... 13
  Strength of Evidence for Major Comparisons and Outcomes ...................................................... 13
  Applicability .................................................................................................................................. 14
  Peer Review and Public Commentary ............................................................................................. 14

Chapter 3. Search Results .............................................................................................................. 15

Chapter 4. Efficacy of Long-term Osteoporosis Drug Therapy .................................................... 18
  Alendronate .................................................................................................................................. 18
    Key Points .................................................................................................................................. 18
    Eligible Studies ........................................................................................................................... 19
    Outcomes .................................................................................................................................... 20
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
    Drug Characteristics .................................................................................................................. 21
  Zoledronic Acid (Zolendronate) ...................................................................................................... 23
    Key Points .................................................................................................................................. 23
    Eligible Studies ........................................................................................................................... 23
  Denosumab .................................................................................................................................... 24
    Key Points .................................................................................................................................. 24
    Eligible Studies ........................................................................................................................... 24
    Outcomes .................................................................................................................................... 25
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
    Drug Characteristics .................................................................................................................. 25
  Raloxifene ..................................................................................................................................... 26
    Key Points .................................................................................................................................. 26
    Eligible Studies ........................................................................................................................... 26
    Outcomes .................................................................................................................................... 27
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
    Drug Characteristics .................................................................................................................. 28
  Hormone Therapy .......................................................................................................................... 30
    Key Points .................................................................................................................................. 30
    Eligible Studies ........................................................................................................................... 30
    Outcomes .................................................................................................................................... 31
Appendices
Appendix A. Search Strategy
Appendix B. Risk of Bias Assessment Decision Aid
Appendix C. Included References Risk of Bias Assessment
Appendix D. Evidence Tables
Appendix E. Excluded Studies
Evidence Summary

Introduction

Osteoporosis is a skeletal disorder of low bone mass and microarchitectural deterioration of bone, leading to bone fragility and increased risk of fracture. Osteoporosis affects more than 10 million U.S. adults aged 50 years or older. About two million U.S. adults experience an osteoporotic or other low- or no-trauma fracture each year. These fractures are a frequent cause of pain, disability, and impaired quality of life; and hip and clinical vertebral fractures, specifically, are associated with increased mortality. Incident radiographic vertebral fractures, a common outcome of drug treatment trials, are defined by quantifying vertebral height losses between baseline and scheduled follow-up radiographs. These are clinically diagnosed in about 15-25% of cases, usually because of back pain, and, even in individuals whose new fracture is not diagnosed, new or worsened back pain is more common than in those without new radiographic vertebral fractures. Because risk of most fractures rises steeply with age, and the population is aging, fracture burden is projected to increase in coming decades.

Though several drugs are approved by the U.S. Food and Drug Administration (FDA) for treatment of osteoporosis to prevent fractures, the proportion of men and women with osteoporosis receiving these treatments is very low. In 2011 to 2013, approximately 10 to 20 percent of older U.S. adults hospitalized for hip fracture received or were prescribed osteoporosis medication in the next 6 to 12 months. Among patients who start treatment, long-term adherence often is low. Out of concern these patterns are resulting in many preventable fractures and associated morbidity and mortality, some have characterized this situation as “a crisis in the treatment of osteoporosis.” They have advocated work to disseminate the evidence on the benefits and harms of osteoporosis drug treatment, including of long-term treatment and of planned discontinuation of treatment (i.e., drug holidays), and on patient and clinician decision making about these treatments.

In short-term (18 to 36 months) randomized controlled trials (RCTs) in postmenopausal women with osteoporosis, bisphosphonates, denosumab, and teriparatide have lowered risk of nonvertebral fractures, clinical vertebral fractures, and radiographic vertebral fractures; and bisphosphonates and denosumab also have lowered risk of hip fractures. However, short-term drug trials have not shown reduced clinical fracture risk in postmenopausal women without osteoporosis, even in those with heightened fracture risk, such as due to low bone mass (i.e., osteopenia) combined with other risk factors. Short-term RCTs and observational studies have found that oral bisphosphonates increase upper gastrointestinal (GI) symptoms; bisphosphonates and denosumab are associated with rare atypical femoral fractures (AFF) and osteonecrosis of the jaw (ONJ); denosumab increases risk of infection; teriparatide increases risk of hypercalcemia; raloxifene increases risk of hot flashes; and both raloxifene and estrogen increase risk of venous thromboembolism and stroke. The benefits of long-term (>3 years) osteoporosis drug therapy on fracture prevention and the risk of harms are less clear, including among individuals with osteopenia, though heightened risk of AFF is a concern. Because of sparse data on fracture outcomes from long-term trials and the high cost of trials adequately powered to evaluate incident fracture outcomes, many investigators have sought to identify appropriate surrogate endpoints. Early treatment changes in bone mineral density (BMD), and, to a lesser extent, in bone turnover markers, may predict short-term risk of incident fractures. However, it is unclear whether changes in these measures predict a high enough proportion of the anti-fracture efficacy of osteoporosis drug treatments to be considered
adequate surrogate outcomes. Further, it needs to be established whether changes in these measures will predict long-term nonvertebral fracture risk with drug treatment, including within individuals.

The effect of long-term osteoporosis drug treatment versus control treatment may vary as a function of other factors, including patient, bone, and osteoporosis drug characteristics. Better understanding of these possible osteoporosis drug treatment effect modifiers may facilitate more informed decisions about treatment over time.

Uncertainty about the benefits of long-term bisphosphonate use coupled with concerns that long-term bisphosphonate persistence in bone might increase fracture risk by inhibiting normal repair of bone microdamage have led to the suggestion that bisphosphonate therapy be discontinued periodically. Though several groups advocate bisphosphonate “drug holidays,” as a strategy to preserve as much fracture benefit as possible while minimizing harms, there is no consensus about who should get them, when they should start, how long they should last, how they should be monitored, or the criteria for restarting osteoporosis drug therapy.

These uncertainties about the most appropriate use of long-term osteoporosis drug therapy and of osteoporosis drug holidays spurred the scheduling of a National Institutes of Health (NIH) Office of Disease Prevention (ODP) Pathways to Prevention (P2P) program workshop on this topic. The goal of the workshop is to present an evidence-based synthesis of the pertinent research base, identify research gaps in the area, and suggest future research needs to assist patients, clinicians, and other healthcare decisionmakers. To accomplish these aims, we conducted this systematic review to address the following questions: (1) What is the efficacy of long-term (>3 years) osteoporosis drug therapy versus control, primarily on risk of incident clinical fractures (recognized in clinical settings, confirmed by study review of nonstudy radiographic report), and secondarily on risk of incident radiographic vertebral fractures (usually not recognized in clinical settings, confirmed by comparing scheduled study radiographs), change in BMD, and harms; (2) Do risks of these fracture types and harms with long-term treatment vary as a function of patient, bone, or osteoporosis drug characteristics; (3) Among individuals receiving osteoporosis drug therapy to prevent fracture, what is the effect of continuing versus discontinuing therapy (i.e., osteoporosis drug holiday of ≥1 year after ≥1 year of prior osteoporosis drug use) on incident fractures and harms; and (4) Do these outcomes with drug holiday vary as a function of patient, bone, or osteoporosis drug characteristics?

**Methods**

The review was conducted following the Agency for Healthcare Research & Quality (AHRQ) methods guidance. The protocol is available at [https://effectivehealthcare.ahrq.gov/topics/osteoporosis-fracture-prevention/research-protocol](https://effectivehealthcare.ahrq.gov/topics/osteoporosis-fracture-prevention/research-protocol) and is registered in PROSPERO. We detail our literature search strategy, study selection criteria, and data extraction and synthesis methods in the full report.

**Results**

We identified 8251 unique publications through June 2018, of which 58 met eligibility criteria and were included in the review. Of 46 publications with low or medium risk of bias (ROB), there were 33 randomized or controlled clinical trials (8 unique studies) and 12 controlled observational studies (10 unique studies) (Appendix C). Most publications were based
on three RCTs of alendronate, zoledronic acid and raloxifene, respectively, and their extension studies. All trials enrolled only postmenopausal women, with most limited to women with osteoporosis as defined by BMD and vertebral fracture history, and a few also including women with osteopenia. The observational studies included between 84 to 100 percent women. Mean participant age was 72 years, with all but two studies reporting a mean age of <80 years. Most of the observational studies presumed that participants had osteoporosis because of past fracture or their use of osteoporosis medications, but none reported BMD status.

Efficacy of Long-term Osteoporosis Drug Therapy

Four eligible placebo-controlled RCTs with low or medium risk of bias (ROB) examined the effect of long-term treatment, one each for alendronate\textsuperscript{23}, raloxifene\textsuperscript{24, 25} denosumab\textsuperscript{26}, and estrogen/progestin.\textsuperscript{27} All participants in these trials were postmenopausal women. In a mix of women with osteoporosis and osteopenia by BMD, four years of alendronate versus placebo significantly reduced risk of incident radiographic vertebral fractures (high strength of evidence [SOE]) (Table A), while absolute risk reductions for incident hip and nonvertebral fracture were small and not statistically significant (low and moderate SOE, respectively).\textsuperscript{23} In women with osteoporosis by BMD or past fracture, four years of raloxifene versus placebo significantly reduced risk of incident radiographic vertebral fractures (high SOE) and incident clinical vertebral fractures (high SOE).\textsuperscript{24} However, raloxifene did not reduce risk of incident hip or nonvertebral fracture (moderate and high SOE, respectively). In women with a hysterectomy and prior clinical fractures, but unknown BMD, seven years of estrogen versus placebo reduced risk of incident clinical fractures and incident hip fractures (both low SOE).\textsuperscript{28} By comparison, in women with an intact uterus and prior clinical fractures, but unknown BMD, 5.6 years of estrogen/progestin versus placebo reduced risk of incident clinical fractures, but a possible reduction in risk of incident hip fractures was not statistically significant (both low SOE).\textsuperscript{29}

It was not possible to directly compare fracture risk between women on long-term denosumab versus placebo, because fracture results for long-term denosumab and denosumab discontinuation treatment arms were pooled.\textsuperscript{26} There were no usable data about the long-term fracture efficacy of other individual agents or of sequential osteoporosis drug therapy (e.g., anabolic followed by anti-resorptive).

Alendronate, denosumab and raloxifene for 4 years each increased hip and lumbar spine BMD compared to placebo.

Variation in Efficacy of Long-term Osteoporosis Drug Therapy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

Analyses from one RCT with low ROB found that long-term risk of incident fractures between alendronate and placebo varied as a function of baseline BMD.\textsuperscript{23} Relative risk of incident clinical fractures was significantly reduced among women with femoral neck (FN)-BMD $\leq -2.5$ (osteoporosis), but not in women with higher BMD (FN-BMD -1.6 to $>-2.5$ [osteopenia]) (moderate SOE) (Table A). In women with osteoporosis, relative risk of incident radiographic vertebral fracture with four years of alendronate versus placebo was significantly reduced by 50 percent (3% absolute reduction) (moderate SOE). Although women with FN-BMD -2.5 to -2 had a similar 44 percent relative reduction in risk of incident radiographic vertebral fractures versus placebo, they had fewer fracture events, and results were not
statistically significant. Women with higher FN-BMD had no reduction in risk of incident radiographic vertebral fractures with long-term alendronate. Authors did not report results for a test of interaction for the outcome of incident radiographic vertebral fractures. In a post hoc analysis, women with osteoporosis randomized to long-term alendronate versus placebo had a significant 55 percent relative reduction in hip fracture (1% absolute reduction), whereas women with osteopenia had no reduction in risk.23 No test of interaction was reported.

In additional post hoc analyses, some conducted in a subset of women with osteopenia30, 31, neither past nonvertebral fracture31, 10-year major osteoporotic fracture (MOF) probability calculated with FN-BMD32, nor pretreatment levels of bone turnover markers33 modified the effect of long-term alendronate versus placebo on risk of any incident fracture outcome. Neither age25, baseline BMD25, nor baseline radiographic vertebral fracture24, 25, 34, 35 modified the effect of long-term raloxifene versus placebo on risk of incident fractures. Two large trials that compared placebo versus hormone treatment reported inconsistent findings about whether treatment effect on risk of incident hip and clinical fractures in all women differed as a function of age or time since menopause.28, 36 However, they minimized their one significant interaction for age because of the many interactions examined28, and did not report whether treatment efficacy varied as a function of these factors within the subgroup of study participants with past fractures.

**Harms of Long-term Osteoporosis Drug Therapy**

One RCT with low ROB found no difference in risk of upper gastrointestinal tract adverse events between long-term alendronate and placebo.23 However, study participants were enrolled in part based on their low risk for upper GI adverse events. Too few events occurred in RCTs to draw conclusions from these studies (insufficient SOE) about risk between long-term alendronate therapy versus placebo for subtrochanteric or femoral shaft fracture without radiographic confirmation of AFF features (2 cases in all treatment groups combined), and we found no alendronate trial data about risk of AFF with confirmed radiologic features or about ONJ.37

Although we found no observational data from eligible studies about the risk of AFF with confirmed radiologic features with long-term alendronate, these studies suggested that long-term alendronate may increase risk of subtrochanteric and femoral shaft fractures without confirmed AFF features (insufficient for >6 years of treatment, low SOE for 3.8 years of treatment).38, 39 Results from several studies also suggested that long-term bisphosphonates as a class (low SOE) may increase risk of AFF with confirmed radiological features40, 41, and risk of subtrochanteric and femoral shaft fractures without confirmed AFF radiological features.42-44 Relative risks for AFF with confirmed radiological features with any long-term bisphosphonate were large, ranging from between 3 and 5 versus past use40 to between 40 to >100 versus no use.41 In the single cohort study reporting these data, absolute risk of AFF was increased by 11 per 10,000 person-years with ≥4 years of bisphosphonate use versus no bisphosphonate use. In addition to studies differing regarding whether or not case definitions required radiologic confirmation for AFF features, studies also differed in whether fractures were excluded for cancer and excess trauma, in their exposure control groups (no bisphosphonate use, limited past bisphosphonate use, raloxifene or calcitonin use), and fracture outcome control groups (no hip fracture, femoral shaft fractures with documented absence of radiologic AFF features).
Three observational studies that compared long-term alendronate versus control for risk of ONJ also used variable methods (e.g., case definitions, exposure control group) and reported inconsistent results. One reported a significant 3-fold increase in risk compared to no treatment (low SOE)\(^{45}\), a second reported a nonsignificant hazard ratio (HR) of 0.86 versus raloxifene or calcitonin (insufficient SOE)\(^{46}\), and the third reported a statistically significant HR of 7.42 (1.02 to 54.09) (insufficient SOE)\(^{47}\). A single cohort study reported no difference between long-term alendronate and no osteoporosis drug use on risk of subsequent atrial fibrillation or flutter.\(^{48}\)

Large magnitude, though mostly borderline statistically significant risk estimates suggested that compared with placebo, 4 years of raloxifene may increase risk of deep vein thrombosis (DVT) by approximately 3-fold (0.5% absolute risk increase)\(^{24,36,49}\) and pulmonary embolism (PE) by about 3 to 4-fold (0.2% absolute risk increase)\(^{24,35,36,49,50}\). In three publications from one observational dataset, risk between long-term raloxifene and no osteoporosis drug treatment were no different for subtrochanteric or femoral shaft fracture\(^{39}\), “any jaw event,”\(^{45}\) or atrial fibrillation.\(^{48}\) Two large trials that compared estrogen or estrogen/progestin versus placebo in postmenopausal women did not report harms results for their subgroups with prior fractures.\(^{28,29}\) However, in their larger populations, unselected for past fractures, both hormone interventions increased risk of cardiovascular disease, and estrogen/progestin increased risk of breast cancer\(^{51}\) and dementia.\(^{52}\)

Risks of harms with long-term denosumab use versus placebo or other control could not be determined because the only eligible study pooled harms results from the continuous long-term denosumab group with those from multiple denosumab discontinuation groups.\(^{26}\)

**Variation in Harms of Long-term Osteoporosis Drug Therapy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics**

Two post hoc analyses reported that risk of upper gastrointestinal tract adverse events was no different between long-term alendronate and placebo as a function of age, current nonsteroidal anti-inflammatory drug use, renal function, or history of upper gastrointestinal tract disease, although as noted earlier, study participants were enrolled in part based on their low risk for upper gastrointestinal tract adverse events.\(^{53,54}\) Three observational studies suggested that risk of AFF\(^{41}\) or subtrochanteric or femoral shaft fracture without radiologic confirmation of AFF features\(^{43,44}\) may increase with the duration of long-term bisphosphonate use. In each of these studies, compared to control exposure, >5 years of bisphosphonate use was associated with a larger relative increase in risk than was 3-5 years of bisphosphonate use. However, none of these studies reported tests for interaction by treatment duration, and the confidence intervals for their treatment duration strata overlapped widely. One study was inconclusive about whether relative risk for AFF associated with long-term bisphosphonate use versus no use increased with age.\(^{41}\) Three post hoc analyses evaluated whether risk of harms between long-term raloxifene and placebo varied as a function of patient, bone or treatment characteristics. The first of these studies reported that risk of DVT and PE between treatments did not vary as a function of baseline cardiovascular risk.\(^{55}\) The second study reported that risk of incident stroke was lower with raloxifene versus placebo in the subgroup of women at increased baseline cardiovascular risk.\(^{56}\) The third study reported that the relative risk of DVT (but not PE) was higher with raloxifene versus placebo in the first two years of treatment but not in the next two years.\(^{49}\)
<table>
<thead>
<tr>
<th>Comparison # of studies Treatment duration</th>
<th>Population Sample size</th>
<th>Outcome</th>
<th>Results Relative Difference (HR / RR / OR, 95% CI) Absolute Difference (ARR, 95% CI)</th>
<th>Strength of evidence* (justification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs placebo 1 RCT(^1) 4 yrs</td>
<td>PM women with osteopenia or osteoporosis (T-score &lt;-1.6) and no RVF n=4,432</td>
<td>Incident CF</td>
<td>No difference: HR=0.86 [0.73, 1.01] ARR=-2 [-4, 0]</td>
<td>Moderate (imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident NVF</td>
<td>No difference: HR=0.88 [0.74, 1.04] ARR=-1 [-3, 0]</td>
<td>Moderate (imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident hip fracture</td>
<td>No difference: HR=0.79 [0.43, 1.44] ARR=-0.2 [-0.8, 0.4]</td>
<td>Low (imprecise, sparse events, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident RVF</td>
<td>Lower risk with alendronate: HR=0.56 [0.39, 0.80] ARR=-2 [-3, -1]</td>
<td>High (unknown consistency)</td>
</tr>
<tr>
<td>Alendronate vs placebo 1 RCT(^7) 3 to 4.5 yrs</td>
<td>PM women with osteopenia or osteoporosis (T-score &lt;-1.6) with or without RVF n=6,459</td>
<td>Subtrochanteric or femoral shaft fracture DX with rare x-ray review for confirmation of AFF features</td>
<td>HR=1.03 [0.06, 16.46] ARR=0 [-0.09, 0.09]</td>
<td>Insufficient (highly imprecise, sparse data, unknown consistency)</td>
</tr>
<tr>
<td>Alendronate vs no osteoporosis drug treatment 2 observational studies(^3, 39, 45) 3.8 yrs (mean) and ≥6 yrs</td>
<td>National database of adults ≥60 yrs with nonhip fracture (90% women) n=534</td>
<td>Subtrochanteric or femoral shaft fracture DX codes without x-ray confirmation of AFF features</td>
<td>≥6 yrs; HR=1.37 [0.22, 8.62] ARR NA (data not reported)</td>
<td>Insufficient (medium ROB, highly imprecise, sparse data, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td>National database of all osteoporosis drug users and general population controls (85% women) n=220,360</td>
<td>Subtrochanteric or femoral shaft fracture DX codes without x-ray confirmation of AFF features</td>
<td>Higher risk with alendronate: 3.8 yrs; ST: HR=2.41 [1.78, 3.27] 3.8 yrs; FS: HR=2.90 [1.97, 4.26] ARRs 0% (very few events)</td>
<td>Low (medium ROB, sparse data, large effect, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td>ONJ DX codes without x-ray or pathology review</td>
<td>Higher risk with alendronate: 3.8 yrs; HR=3.15 [1.44, 6.87] ARR NA (data not reported)</td>
<td>Low (medium ROB, sparse data, large effect, unknown consistency)</td>
<td></td>
</tr>
<tr>
<td>Alendronate vs raloxifene 1 observational study(^7)</td>
<td>Women aged ≥50 yrs from database of 1 hospital n=8,354</td>
<td>ONJ DX codes with x-ray and pathology features</td>
<td>HR=7.42 [1.02, 54.09] ARR NA (data not reported)</td>
<td>Insufficient (medium ROB, highly imprecise, large</td>
</tr>
<tr>
<td>Comparison # of studies</td>
<td>Population</td>
<td>Outcome</td>
<td>Results</td>
<td>Strength of evidence*</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Treatment duration</td>
<td></td>
<td></td>
<td></td>
<td>(justification)</td>
</tr>
<tr>
<td>~4 yrs (mean)</td>
<td></td>
<td></td>
<td></td>
<td>effect, unknown consistency</td>
</tr>
<tr>
<td>Alendronate vs raloxifene or calcitonin 1 observational study</td>
<td>Adults aged ≥50 yrs (84% women) with recent hip or vertebral fracture now on osteoporosis drug treatment from national database n=43,645</td>
<td>ONJ DX codes without x-ray or pathology review</td>
<td>HR=0.86 [0.44, 1.69] ARR NA (data not reported)</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency)</td>
</tr>
<tr>
<td>up to 6 yrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bisphosphonate‡ vs no bisphosphonate 2 observational studies†</td>
<td>Adults aged ≥55 yrs from national database (87% women cases and 52% women controls in cohort analysis; 86% women in case-control analysis) n~2.8 million (cohort), n=1,124 (case-control)</td>
<td>AFF with radiologic features</td>
<td>Higher risk with bisphosphonate: Cohort ≥4 yrs: RR=126 [55, 238] ARR NA (data not reported) Case-control 3-4 yrs: OR=40 [17, 91] ARR=6 [3, 10] 4-5 yrs: OR=116 [58, 234] ARR=7 [3, 11] &gt;5 yrs: OR=93 [66, 132] ARR=6 [2, 10]</td>
<td>Low (medium ROB, imprecise, consistent, large effect)</td>
</tr>
<tr>
<td>&gt;3 yrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current vs past bisphosphonates‡ 2 observational studies†</td>
<td>Women aged ≥65 yrs from national primary practice database n=264</td>
<td>AFF with radiologic features</td>
<td>Higher risk with current bisphosphonate: &gt;3 yr: OR=9.46 [2.17, 41.3] ARR=10 [0, 20]</td>
<td>Low (medium ROB, imprecise, consistency unknown, large effect)</td>
</tr>
<tr>
<td>&gt;3 yrs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denosumab vs placebo 1 RCT‡</td>
<td>PM women with osteopenia or osteoporosis by BMD n=365</td>
<td>Incident CF</td>
<td>No difference: RR=0.97 [0.40, 2.35] ARR=0.4 [-10, 9]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td>Comparison # of studies</td>
<td>Population Sample size</td>
<td>Outcome</td>
<td>Results Relative Difference (HR / RR / OR, 95% CI)</td>
<td>Absolute Difference (ARR, 95% CI)</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------------------</td>
<td>---------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Treatment duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 yrs</td>
<td></td>
<td>SAE</td>
<td>No difference: RR=1.64 [0.69, 3.88] ARR=7 [-3, 17]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td>Raloxifene vs placebo</td>
<td>PM women with osteoporosis by BMD or RVF n=6,828</td>
<td>Incident NVF</td>
<td>No difference: 4 yrs: RR=0.93 [0.81, 1.06] * 8 yrs: HR=1.00 [0.82, 1.21] ‡ ARR NA (data not reported)</td>
<td>4 yrs: High (unknown consistency) 8 yrs: Moderate (medium ROB, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident hip fracture</td>
<td>No difference: 4 yrs: RR=0.97 [0.62, 1.52] † ARR=0 [-0.6, 0.5]</td>
<td>Moderate (imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident CVF</td>
<td>Lower risk with raloxifene: 4 yrs: RR=0.58 [0.43, 0.79] ** ARR=2 [-3, -1]</td>
<td>High (unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident RVF</td>
<td>Lower risk with raloxifene: 4 yrs: RR=0.64 [0.53, 0.76] ** ARR=5 [-6, -3]</td>
<td>High (unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAE</td>
<td>No difference: 8 yrs: RR=0.93 [0.86, 1.00] ** ARR=3 [-6, 0]</td>
<td>Moderate (medium ROB, unknown consistency, upper bound CI near significant)</td>
</tr>
<tr>
<td></td>
<td>National database of adults (85% women) exposed to osteoporosis drugs or controls n=19,324</td>
<td>Subtrochanteric or femoral shaft fracture DX codes without x-ray confirmation of AFF features</td>
<td>ST: HR=1.06 [0.34, 3.32] FS: HR=0.82 [0.21, 3.20] ARRs 0% (very few events)</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency, sparse data)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ONJ DX codes without x-ray or pathology review</td>
<td>2 cases, only in control group</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency, sparse data)</td>
</tr>
<tr>
<td></td>
<td>PM women with prior hysterectomy and past clinical fracture n=3,816</td>
<td>Incident CF</td>
<td>Lower risk with estrogen HR=0.73 [0.62, 0.86] ARR=5 [-7, -2]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident hip fracture</td>
<td>Lower risk with estrogen HR=0.55 [0.32, 0.94] ARR=1 [-2, 0]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td>PM women with past</td>
<td>Incident CF</td>
<td>HR=0.83 [0.17, 3.91] ARR NA (data not reported)</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

* ROB: Risk of bias, NA: Not applicable.
<table>
<thead>
<tr>
<th>Comparison</th>
<th>Population</th>
<th>Outcome</th>
<th>Results</th>
<th>Strength of evidence* (justification)</th>
</tr>
</thead>
<tbody>
<tr>
<td># of studies</td>
<td>Treatment duration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hysterectomy and osteoporosis by BMD n=53</td>
<td>Incident CF</td>
<td>HR=0.83 [0.49, 1.40] ARR NA (data not reported)</td>
<td>(medium ROB, highly imprecise, unknown consistency)</td>
<td></td>
</tr>
<tr>
<td>PM women with past hysterectomy and osteopenia by BMD n=363</td>
<td>Incident CF</td>
<td>HR=0.78 [0.68, 0.91] ARR=-3 [-5, -1]</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency)</td>
<td></td>
</tr>
<tr>
<td><strong>Estrogen/progestin vs. placebo</strong> 1 RCT23 5.6 yrs (mean)</td>
<td>PM women with intact uterus and past clinical fracture n=5,897</td>
<td>Incident CF</td>
<td>Lower risk with estrogen/progestin HR=0.77 [0.48, 1.22] ARR=-0.3 [-0.9, 0.2]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident hip fracture</td>
<td>No difference HR=0.77 [0.48, 1.22] ARR=-0.3 [-0.9, 0.2]</td>
<td>Low (medium ROB, imprecise, unknown consistency)</td>
</tr>
<tr>
<td></td>
<td>PM women with intact uterus and osteoporosis by BMD n~50</td>
<td>Incident CF</td>
<td>HR=0.53 [0.25, 1.10] ARR NA (data not reported)</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency)</td>
</tr>
<tr>
<td><strong>Estrogen/progestin vs. nonplacebo control</strong> 1 RCT27 4 yrs</td>
<td>PM women with osteoporosis by BMD (T-score ≤ -2) and RVF n=36</td>
<td>Incident CF</td>
<td>HR=0.93 [0.06, 13.5] ARR=-0.5 [-19, 18]</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency, sparse data)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident RVF</td>
<td>RR=0.37 [0.09, 1.62] ARR=-22 [-53, 8]</td>
<td>Insufficient (medium ROB, highly imprecise, unknown consistency, sparse data)</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; AL=alendronate; ARR=absolute risk reduction; BMD=bone mineral density; CCT=controlled clinical trial; CF=clinical fracture; CI=confidence intervals; CVF=clinical vertebral fracture; D=denosumab; DX=diagnosis; FS=femoral shaft; HR=hazard ratio; NA=not available; NR=not reported; NVF=nonvertebral fracture; ONJ=osteonecrosis of the jaw; OR=odds ratio; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; ROB=risk of bias; RR=risk ratio; RVF=radio graphic vertebral fracture; SAE=serious adverse event; ST=subtrochanteric; Z=zoledronic acid

*Definitions of terms for strength of evidence grades and domains ratings are detailed in the section of the main report titled, ‘Strength of Evidence for Major Comparisons and Outcomes.’
†Results reported for raloxifene 60 mg/d and 120 mg/day groups pooled together.
‡Included bisphosphonates varied by study. All studies included alendronate, risedronate, and one or more of the following: ibandronate, etidronate, and zoledronate.
**Results reported for raloxifene 60 mg/d dose group.
Effect of Osteoporosis Drug Holidays

In postmenopausal women who previously received 5 years of alendronate for osteopenia or osteoporosis, two trials compared continued alendronate for up to 5 more years versus discontinued alendronate (placebo). Neither found a reduction in risk of incident nonvertebral fractures. The trial that provided the best quality evidence enrolled a mix of women with osteoporosis and osteopenia, and reported that alendronate continuation was associated with an approximately 55 percent relative reduction (3% absolute reduction) in risk of incident clinical vertebral fracture, but not in risk of incident radiographic vertebral fracture, incident hip fracture, or other clinical fracture outcomes. The other trial, limited to women with osteoporosis, reported no difference in risk of incident clinical vertebral fracture, but did not report separate results for risk of hip fracture. In postmenopausal women who received three years of zoledronic acid for osteoporosis, those assigned to continue zoledronic acid for three more years versus placebo had an approximately 50 percent relative reduction (3% absolute reduction) in risk of radiographic vertebral fractures, but no difference in risk for any incident hip fracture or other clinical fractures. In a subsequent three year extension of this trial, too few incident fractures occurred to draw conclusions about differences in risk between treatment groups. Similarly, we could not draw conclusions from a small trial of denosumab continuation versus discontinuation because fracture results were pooled across treatment arms.

Compared to baseline after 5 years of alendronate, treatment with alendronate for an additional 5 years was associated with slightly decreased or stable hip BMD, whereas women changed to placebo for 5 years experienced significantly larger declines in hip BMD. Similarly, compared to baseline after 3 years of zoledronic acid, treatment with zoledronic acid for an additional 3 years was associated with slightly decreased or stable hip BMD, whereas women changed to placebo for 3 years experienced significantly larger declines in hip BMD. There was no difference in change in hip BMD between women who continued zoledronic acid for 9 years versus 6 years. Results of a small denosumab trial suggested that compared to baseline, hip and spine BMD were most increased in women assigned to denosumab for 4 years, less increased in women assigned denosumab for 2 years, placebo for 1 year, then denosumab for 1 year, and least increased in women assigned denosumab for 2 years followed by placebo for 2 years.

Variation in Effect of Osteoporosis Drug Holidays as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

In data limited to two post hoc analyses of alendronate continuation versus discontinuation, between-treatment risk of incident nonvertebral fracture or incident clinical vertebral fracture did not vary as a function of baseline FN-BMD or whether participants had a baseline radiographic vertebral fracture. We found no evidence about possible modifiers of the effect of continuing any other osteoporosis drug treatment versus discontinuation on risk of incident fracture. Further stratified analyses reported one statistically significant interaction (for effect of baseline FN-BMD on risk of alendronate continuation vs. discontinuation, for only 1 of 3 fracture outcomes, and only in subset of women without prevalent radiographic vertebral fracture). However, results were not adjusted for multiple testing.

Harms of Osteoporosis Drug Holidays

Trials of alendronate and zoledronic acid continuation versus discontinuation reported no difference between treatment groups in risk of serious adverse events.
AFF with confirmed radiological features, subtrochanteric or femoral shaft fracture without confirmed AFF features, or ONJ, occurred in these trials to draw conclusions about differences in their risk between treatment continuation and discontinuation groups.\textsuperscript{37, 65} Though atrial fibrillation appeared more frequently with zoledronic acid continuation versus discontinuation, the absolute number of events was low and possible differences between treatment groups were not statistically significant.\textsuperscript{65, 66} It was not possible to draw conclusions about differences in harms between the denosumab continuation and discontinuation arms in a single trial reporting because their harms results were pooled.\textsuperscript{26}

**Variation in Effect of Osteoporosis Drug Holidays as a Function of Patient, Bone, or Osteoporosis Drug Characteristics**

We found no evidence about whether the risk of harms between continuation of any osteoporosis drug treatment and discontinuation varied as a function of patient, bone or drug characteristics.

**Table B. Evidence for osteoporosis drug continuation versus discontinuation (discontinuation ≥1 year after prior treatment ≥1 year) (Key Questions 5-8)**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Population</th>
<th>Outcome</th>
<th>Results</th>
<th>Strength of evidence*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparisons</strong></td>
<td><strong>Number of studies</strong></td>
<td><strong>Number of subjects Follow-up time</strong></td>
<td><strong>Outcome</strong></td>
<td><strong>Results</strong></td>
</tr>
<tr>
<td>Alendronate continuation vs discontinuation (AL x 10 yrs vs. AL x 5 yrs followed by PBO x 5 yrs)</td>
<td>1 RCT\textsuperscript{52} n=1,099</td>
<td>PM women previously received alendronate 5 yrs for osteopenia or osteoporosis (T-score &lt;1.6)</td>
<td>Incident CF</td>
<td>No difference: RR=0.93 [0.71, 1.21] ARR=-1 [-6, 4]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incident NVF</td>
<td>No difference: RR=1.00 [0.76, 1.32] ARR=-0.1 [-5, 5]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incident hip fracture</td>
<td>No difference: RR=1.02 [0.51, 2.10] ARR=0 [-2, 2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incident CVF</td>
<td>Lower risk with alendronate continuation: RR=0.45 [0.24, 0.85] ARR=-3 [-5, 0]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incident RVF</td>
<td>No difference: RR=0.86 [0.60, 1.22] ARR=-1 [-5, 2]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SAE</td>
<td>Stated as no difference, but no data provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Subtrochanteric or femoral shaft fracture DX with rare x-ray review</td>
<td>No difference: HR=1.33 [0.12, 14.67] ARR=-0.1 [-0.5, 0.7]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AFF not defined</td>
<td>No information provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONJ not defined</td>
<td>No cases in either group</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Incident NVF</td>
<td>No difference: A7 v vs. A5/P2: Low</td>
</tr>
<tr>
<td>Comparison</td>
<td>Population</td>
<td>Outcome</td>
<td>Results</td>
<td>Strength of evidence*</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------</td>
<td>----------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Alendronate continuation vs</td>
<td>PM women previously received alendronate 5 yrs for osteoporosis (T-score ≤ 2.5)</td>
<td>Incident CVF</td>
<td>A7 vs. A5/P2: RR=0.87 [0.40, 1.91] ARR=-1 [-7, 5]</td>
<td>(imprecise) A10 vs. A7/P3: Insufficient (medium ROB, imprecise)</td>
</tr>
<tr>
<td>discontinuation (AL x 7 yrs [A7] vs. AL x 5 yrs followed by PBO x 2 yrs [A5/P2]; AL x 10 yrs [A10] vs. AL x 7 yrs + PBO x 3 yrs [A7/P3])</td>
<td></td>
<td>A10 vs. A7/P3: RR= 0.81 [0.38, 1.71] ARR=-2 [-11, 6]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 RCT 63, 64</td>
<td>n=350</td>
<td>Incident RVF</td>
<td>Did not differ significantly between groups at 10 yrs</td>
</tr>
<tr>
<td>Zoledronic acid continuation vs</td>
<td>PM women with osteopenia</td>
<td>Incident CF</td>
<td>No difference: RR=1.37 [0.39, 4.78] ARR=1 [-2, 4]</td>
<td>Insufficient (imprecise)</td>
</tr>
<tr>
<td>discontinuation (Z x 2 yrs vs. Z x 1 yr followed by PBO x 1 yr)</td>
<td></td>
<td>SAE</td>
<td>No difference: RR=0.91 [0.50, 1.67] ARR=-1 [-7, 5]</td>
<td>Moderate (imprecise)</td>
</tr>
<tr>
<td></td>
<td>1 RCT 65</td>
<td>n=379</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoledronic acid continuation vs</td>
<td>PM women previously received zoledronic acid 3 yrs for osteoporosis by BMD or RVF</td>
<td>Incident CF</td>
<td>No difference: HR=1.04 [0.71, 1.54] ARR NA (data not reported)</td>
<td>Low (imprecise)</td>
</tr>
<tr>
<td>discontinuation (Z x 6 yrs vs. Z x 3 yrs followed by PBO x 3 yrs)</td>
<td></td>
<td>Incident NVF</td>
<td>No difference: HR= 0.99 [0.7, 1.5] ARR=-0.3 [-3, 3]</td>
<td>Low (imprecise)</td>
</tr>
<tr>
<td></td>
<td>1 RCT 65</td>
<td>N=1,233</td>
<td>Incident hip fracture</td>
<td>No difference: HR= 0.90 [0.33, 2.49] ARR=-0.2 [-1, 1]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident CVF</td>
<td>No difference: HR=1.81 [0.53, 6.2] ARR NA (data not reported)</td>
<td>Insufficient (imprecise)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident RVF</td>
<td>Lower risk with zoledronic acid continuation: OR=0.51 [0.26, 0.95] ARR=-3 [-6, -1]</td>
<td>Moderate (imprecise)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAE</td>
<td>No difference: RR=1.14 [0.96, 1.36] ARR=4 [-1, 9]</td>
<td>Moderate (imprecise)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AFF with radiologic features</td>
<td>No cases occurred</td>
<td>Insufficient (sparse data)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ONJ (exposed jaw bone &gt;6 wks)</td>
<td>One case occurred (in continuation group)</td>
<td>Insufficient (sparse data)</td>
</tr>
<tr>
<td>Comparison</td>
<td>Population</td>
<td>Outcome</td>
<td>Results</td>
<td>Strength of evidence*</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Zoledronic acid continuation vs discontinuation</td>
<td>PM women previously received zoledronic acid 6 yrs for osteoporosis by BMD or RVF</td>
<td>Incident CF</td>
<td>No difference: HR=1.11 [0.45, 2.73] ARR=1 [-7, 10]</td>
<td>Insufficient (imprecise, sparse data)</td>
</tr>
<tr>
<td>(Z x 9 yrs vs. Z x 6 yrs followed by PBO x 3 yrs)</td>
<td>1 RCT(^{x}) n=190</td>
<td>Incident RVF</td>
<td>No difference: OR=0.58 [0.13, 2.55] ARR=-2 [-6, 4]</td>
<td>Insufficient (imprecise, sparse data)</td>
</tr>
<tr>
<td>Denosumab continuation vs discontinuation</td>
<td>PM women with osteopenia or osteoporosis by BMD</td>
<td>Incident CF</td>
<td>No numerical data</td>
<td>Insufficient (no data)</td>
</tr>
<tr>
<td>(D x 4 yrs vs. D x 2 yrs followed by PBO x 2 yrs)</td>
<td>1 RCT(^{x}) n=314</td>
<td>SAE</td>
<td>No numerical data</td>
<td>Insufficient (no data)</td>
</tr>
</tbody>
</table>

Abbreviations: AFF=atypical femoral fracture; AL=alendronate; ARR=absolute risk reduction; BMD=bone mineral density; CCT=controlled clinical trial; CF=clinical fracture; CI=confidence intervals; D=denosumab; DX=diagnosis; FS=femoral shaft; HR=hazard ratio; NA=not available; NR=not reported; NVF=nonvertebral fracture; ONJ=osteonecrosis of the jaw; OR=odds ratio; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; ROB=risk of bias; RR=risk ratio; RVF=radiographic vertebral fracture; SAE=serious adverse event; ST=subtrochanteric; Z=zoledronic acid

*Definitions of terms for strength of evidence grades and domains ratings are detailed in the section of the main report titled, ‘Strength of Evidence for Major Comparisons and Outcomes.’

**Discussion**

In postmenopausal women with osteoporosis by BMD but without baseline vertebral fractures, 4-year treatment with alendronate versus placebo lowered risk of incident clinical fractures (moderate SOE) and incident radiographic vertebral fractures (moderate SOE). In postmenopausal women with osteoporosis by BMD and/or baseline vertebral fractures, 4-year treatment with raloxifene versus placebo lowered risk of incident clinical vertebral fractures and incident radiographic vertebral fractures (both high SOE), but not risk of hip or nonvertebral fractures. In postmenopausal women with past fractures but unknown BMD, both 7-year treatment with estrogen (in women with hysterectomy) and approximately 6-year treatment with estrogen/progestin (in women with an intact uterus) lowered risk of incident clinical fractures and estrogen lowered risk of incident hip fractures. However, we found no usable data about the long-term efficacy of other agents for reducing fracture risk versus placebo. Four-year treatment with alendronate, denosumab or raloxifene versus placebo each significantly increased both hip and vertebral BMD.

Continuation of zoledronic acid versus discontinuation after 3 years of treatment lowered risk of incident radiographic vertebral fractures (moderate SOE) but evidence was insufficient to
draw conclusions about differences in risk of incident clinical vertebral fractures. By comparison, continuation of alendronate versus discontinuation after 5 years of treatment may lower risk of incident clinical vertebral fractures (low SOE for no difference between 7 vs. 5 years, and moderate SOE for lower risk between 10 vs. 5 years), but did not lower risk of incident radiographic vertebral fractures (low SOE). Neither zoledronic acid nor alendronate continuation versus discontinuation reduced risk of incident clinical or nonvertebral fractures (low to moderate SOE). While we found low SOE that alendronate continuation versus discontinuation did not lower risk of incident hip fracture, evidence for zoledronic acid continuation was insufficient to draw conclusions. We found no usable data about the efficacy for fracture prevention with continuation versus discontinuation of other osteoporosis drug therapies. Continuation of zoledronic acid and alendronate each were associated with generally stable hip BMD compared to small, but significant declines with discontinuation.

For alendronate, the effect of long-term treatment versus placebo on risk of incident fracture differed between women with and without osteoporosis. Four-year treatment with alendronate versus placebo only reduced risk of incident clinical fractures in women with osteoporosis and not in women with osteopenia. Post hoc analyses suggested a similar pattern for risk of incident hip fractures. Otherwise, evidence within women with osteoporosis, and separately within women with osteopenia, suggested that risk of fracture between long-term alendronate or long-term raloxifene and placebo did not vary by baseline age, history of fracture, World Health Organization (WHO) Fracture Risk Assessment tool (FRAX) score, or change in bone turnover markers.

Though substantial differences in study designs prevented pooling and limited confidence in results across studies, mostly retrospective cohort and case-control observational studies suggested that long-term bisphosphonates increased risk of AFF with confirmed radiological features, subtrochanteric or femoral shaft fractures without confirmed AFF features, and ONJ. Further, risk of these fractures may continue to rise with longer treatment.

We identified no eligible RCTs or CCTs of long-term osteoporosis drug therapy versus placebo or of osteoporosis drug continuation versus discontinuation published since a 2014 AHRQ review on the benefits and harms of osteoporosis drug therapies. However, the efficacy of osteoporosis drug therapies versus placebo on fracture prevention reported in the current review appear less favorable than those reported in the prior review. At least in part, this is because of the focus of the current review on long-term trials. For alendronate, the smaller fracture reduction benefit in long-term trials versus that in short-term trials may have been at least partly attributable to a lower baseline fracture risk. The long-term alendronate trial excluded women with past vertebral fractures, a major risk factor for future fractures. It had a lower incidence of clinical fracture, nonvertebral fracture, hip fracture, and radiographic vertebral fracture in its placebo group than occurred in the placebo group of an associated 3-year alendronate trial conducted in women with past vertebral fractures. Relative risks of all these incident fracture types with alendronate versus placebo were numerically higher in the 4-year trial than in the 3-year trial, suggesting possibly less fracture reduction benefit with longer follow-up. However, confidence intervals overlapped and the indirect nature of these comparisons limits what conclusions should be drawn.

Recently published studies included in the current review have modestly refined prior estimates of the risk of AFF, subtrochanteric or femoral shaft fracture without confirmed radiologic AFF features, and ONJ associated with long-term bisphosphonate use. A 2013 meta-analysis examined the risk of subtrochanteric or femoral shaft fracture associated with 5 or more
years of bisphosphonate use versus control. None of the studies it included defined cases of AFF using American Society of Bone and Mineral Research (ASBMR) radiologic criteria. The authors of that review reported a pooled adjusted risk ratio (RR) of 1.62 (95% CI 1.18, 2.22). Multiple recent observational studies have defined AFF using ASBMR radiologic criteria and risk estimates have tended to be higher. A 2014 systematic review on the association between bisphosphonate use and ONJ suggested an increased risk, but noted considerable heterogeneity between studies. Although the current review includes two observational studies published since those reviews, these observational studies defined ONJ differently from each other and reported conflicting findings.

Our findings have several clinical implications. First, considering the risks associated with long-term estrogen and estrogen/progestin treatment, and the failure to identify any subgroup in which their fracture benefits outweigh their harms, they do not appear to be not viable options for long-term osteoporosis treatment. Second, in postmenopausal osteoporotic women, compared to placebo, only alendronate among osteoporosis drug treatment options has demonstrated a reduction in risk of nonvertebral fractures with treatment for more than 3 years. Though results also suggested a possible reduction in risk of hip fracture with 4 years of alendronate, there were few events and analyses were post-hoc and susceptible to type 1 error. No currently available osteoporosis drug therapy demonstrated a reduction in risk of incident nonvertebral fractures with treatment longer than 4 years. In patients who have completed 3 to 5 years of bisphosphonate treatment, and in whom the clinical question is whether there is fracture reduction benefit with continued bisphosphonate treatment versus discontinuation, evidence suggests that continued treatment for another 3 to 5 years does not lower risk of any incident nonvertebral fracture outcome, but lowers some incident vertebral fracture outcomes. However, not all vertebral fracture outcomes are created equal. Continued alendronate lowered risk of incident clinical vertebral fractures, which nearly always come to clinical attention due to back pain, while continued zoledronic acid lowered risk of the less clinically important outcome of incident radiographic vertebral fractures, which sometimes, but much less often are symptomatic.

Beyond highlighting some of the uncertainty remaining about the magnitude of benefits and harms of long-term osteoporosis drug treatment and osteoporosis drug treatment holidays versus placebo, our findings provide some information to help weigh the potential benefits versus harms. For example, considering the clinical decision of whether to continue alendronate in a previously osteoporotic postmenopausal women who has completed 5 years of alendronate, one trial found that 5 additional years of alendronate was associated with a 3 percent absolute reduction in risk of incident clinical vertebral fracture compared to placebo. This translated to approximately 33 women needing to be treated with 5 additional years of alendronate instead of placebo to prevent one incident clinical vertebral fracture. Data to estimate absolute risks of severe harms were even more limited. We could not reliably estimate the absolute risks of AFF from any osteoporosis drug holiday trials because of scant events. Further, only one observational study of long-term bisphosphonate treatment used a cohort design and defined AFF using ASBMR radiographic criteria. In that study, absolute risk of AFF with long-term bisphosphonate use was increased by 11 per 10,000 person-years (95% CI, 7, 14) versus no use. Given that exposure in this study was defined as bisphosphonate use for ≥4 years, approximately 227 women would need to be treated with ≥4 additional years of alendronate to cause 1 additional case of AFF. Though such rough calculations have potential utility for framing the magnitude of benefits versus risks for those making clinical treatment decisions, great caution

ES-15
should be exercised before doing so given that these numbers needed to treat and harm each are derived from just a single study, and from studies of different designs.

Limitations

The available data limit this review in several ways. First, there were few unique trials of long-term osteoporosis drug therapy or of drug continuation versus discontinuation (drug holiday), often only one for a given treatment comparison, and only one of which was designed with incident fracture as the primary outcome. Consequently, many studies had few incident clinical fractures, especially for hip fractures, and low statistical power to precisely estimate differences in their risk between treatment interventions. Second, all trials were conducted in generally healthy postmenopausal women, with all but the WHI hormone trials conducted in women with osteopenia or osteoporosis by BMD or radiographic vertebral fracture criteria. Therefore, results may not be generalizable to men, individuals with comorbidities, or those with other reasons for heightened fracture risk. Third, trials reported limited data on treatment harms and reporting appeared inconsistent between studies, limiting confidence around risk estimates for these outcomes and raising concerns about possible reporting bias. Fourth, observational studies investigating the association of treatment with risk of AFF or ONJ had marked methodologic differences that likely affected the specificity of the outcome and the risk estimates. Major differences included the definitions of the cases (e.g., whether or not fractures were defined using ASBMR radiographic AFF features), drug therapy exposure and control groups, and noncase control groups. Studies also varied in the extent to which they accounted for potential selection bias between the drug treatment exposure and control groups. Since studies often differed in multiple ways, differences in their findings for these outcomes were difficult to interpret. Fifth, few studies reported results on possible effect modifiers of drug therapy outcomes. Where such analyses were reported, they were almost entirely post hoc, often without testing for interaction, raising the likelihood of type 1 errors. Sixth, there were no eligible trials of with usable data on long-term treatment efficacy for fracture prevention for agents other than alendronate, raloxifene, estrogen, and estrogen/progestin, and no eligible long-term trials of sequential osteoporosis drug therapy (e.g., anabolic followed by anti-resorptive). Finally, there were no eligible studies with usable data that compared different durations of osteoporosis drug holidays.

Research Needs

Future trials of long-term osteoporosis drug therapy and osteoporosis drug continuation versus discontinuation should be large enough to have adequate statistical power to assess risks of clinical fracture endpoints, including hip fractures, the fracture type with the greatest risk of subsequent morbidity and mortality. Future long-term trials should include evaluations of sequential osteoporosis drug therapy, including comparisons of anabolic therapy followed by antiresorptive therapy, versus continuous antiresorptive therapy. Future long-term trials also should include men, in whom there are no long-term data, nonwhite men and women, and individuals with comorbidity, to increase generalizability to typical clinical populations. Trials also should be considered in populations who meet neither BMD nor baseline radiographic vertebral fracture criteria for osteoporosis and have been excluded from past trials, but who still are considered at high fracture risk. Future studies should systematically collect and report harms data. Randomized trials likely will continue to be limited for evaluating the risk of rare treatment harms such as AFF and ONJ, so future observational studies should use cohort designs, standard
case definitions,\textsuperscript{16, 72} standard non-case and exposure controls, and perform adequate statistical adjustment for selection bias. Trials should compare different durations of osteoporosis drug holidays, with or without restarting osteoporosis drug therapy and including a comparison group of continuous drug therapy. Observational studies may be supplementary and provide insights about the benefits and harms of drug holidays of different durations and patient and treatment characteristics that predict which patients are likely to benefit or be harmed by treatment continuation versus discontinuation.\textsuperscript{73, 74} Studies should specify analysis plans a priori to investigate possible effect modifiers of long-term therapy and drug holiday outcomes (e.g., age, BMD, and bone markers before and during the drug holiday). Results of patient-level data from osteoporosis drug trials on the association of early treatment changes in BMD and bone turnover markers with risk of incident fractures may improve understanding of the potential and limitations of these measures as surrogates for incident fracture.\textsuperscript{17, 18, 75}

Conclusions

Few osteoporosis treatment drugs have been shown to reduce risk of clinical fractures with long-term treatment. In postmenopausal women with osteoporosis and no prior osteoporosis drug treatment, both alendronate and raloxifene reduce risk of incident vertebral fracture with up to 4 years of treatment, while 4 years of alendronate also lowers risk of incident clinical fractures. Though trial data on estrogen and estrogen/progestin are not available for women with osteoporosis defined using standard criteria (i.e., BMD or past hip or vertebral fractures), their reduction in risk of incident clinical fractures in postmenopausal women with past fractures, as a proxy for osteoporosis, suggests they’d be effective in women known to have osteoporosis. However, due to their counterbalancing increases in risk of cardiovascular disease, dementia and breast cancer, and the failure of secondary analyses to identify any subpopulation in whom fracture benefits outweigh these harms, there seems no indication for the long-term use of estrogen and estrogen/progestin for fracture prevention. In patients who already have completed a course of osteoporosis drug treatment (i.e., 3 to 5 years), the most favorable trial evidence for continued treatment is for an approximately 3 percent absolute reduction in risk of incident clinical vertebral fractures with 10 years of alendronate versus 5 years. Observational data suggests that the absolute increases in risk for AFF and ONJ with long-term bisphosphonate treatment are far smaller. Future research to help patients and clinicians make decisions about long-term osteoporosis drug treatment should help refine these absolute risk estimates. It should examine how these risks vary as a function of patient, bone, and drug treatment characteristics (e.g., age, sex, pre-drug holiday BMD, duration of prior osteoporosis drug treatment). Further, because patients don’t experience these events dichotomously, quantifying their morbidity in a common measure may help patients weigh their trade-offs more easily.


50. Barrett-Connor E, Cauley JA, Kulkarni PM, et al. Risk-benefit profile for raloxifene: 4-year data From the Multiple Outcomes of Raloxifene


75. Black D. Change in BMD as a Surrogate for Fracture Risk Reduction in Osteoporosis Trials: Results from Pooled, Individual-level Patient Data from the FNIH Bone Quality Project [abstract].
Chapter 1. Introduction

Background

Osteoporosis is a skeletal disorder of low bone mass and microarchitectural deterioration of bone, leading to bone fragility and increased risk of fracture. In 1994, a World Health Organization (WHO) Study Group operationally defined osteoporosis in women as femoral neck bone mineral density (BMD) equal to or lower than 2.5 standard deviations below the mean BMD of young white women. Considering BMD at either the femoral neck (FN) or lumbar spine (LS), and extrapolating this definition to men and nonwhite women, osteoporosis affects more than 10 million U.S. adults aged 50 years or older. About 2 million U.S. adults experience an osteoporotic or other low- or no-trauma fracture each year. These fractures are a frequent cause of pain, disability, and impaired quality of life; and hip and clinical vertebral fractures, specifically, are associated with an increased risk of mortality. Incident radiographic vertebral fractures, a common outcome of drug treatment trials, are defined by quantifying vertebral height losses between baseline and scheduled follow-up radiographs. These are clinically diagnosed in about 15-25% of cases, usually because of back pain, and, even in individuals whose new fracture is not diagnosed, new or worsened back pain is more common than in those without new radiographic vertebral fractures. Because risk of most fractures rises steeply with age, and the population is aging, fracture burden is projected to increase in coming decades.

Though several drugs are approved by the U.S. Food and Drug Administration (FDA) for treatment of osteoporosis to prevent fractures, the proportion of men and women with osteoporosis receiving these treatments is very low. In 2011 to 2013, approximately 10 to 20 percent of older U.S. adults hospitalized for hip fracture received or were prescribed osteoporosis medication in the next 6 to 12 months. Among patients who start treatment, long-term adherence often is low. Out of concern these patterns are resulting in many preventable fractures and associated morbidity and mortality, some have characterized this situation as “a crisis in the treatment of osteoporosis.” They have advocated work to disseminate the evidence on the benefits and harms of osteoporosis drug treatment, including of long-term treatment and of planned discontinuation of treatment (i.e., drug holidays), and on patient and clinician decision making about these treatments.

In short-term (18 to 36 months) randomized controlled trials (RCTs), U.S. Food & Drug Administration approved bisphosphonates, denosumab, and teriparatide have lowered risk of nonvertebral fractures, clinical vertebral fractures, and radiographic vertebral fractures; and bisphosphonates and denosumab also have lowered risk of hip fractures. However, evidence of fracture protection is predominately from studies of postmenopausal women with osteoporosis defined by low BMD or by the presence of vertebral fractures found on screening x-rays. In contrast, short-term drug trials have not shown reduced clinical fracture risk in postmenopausal women without osteoporosis, even in those who had heightened fracture risk because of low bone mass (i.e., osteopenia) combined with other risk factors (e.g., falls, high WHO Fracture Risk Assessment Tool [FRAX®] score). In short-term RCTs and observational studies, oral bisphosphonates increase upper gastrointestinal symptoms; bisphosphonates and denosumab are associated with rare atypical femoral fractures (AFF) and osteonecrosis of the jaw (ONJ); denosumab increases risk of infection; teriparatide increases risk of hypercalcemia; raloxifene increases risk of hot flashes; and both raloxifene and estrogen increase risk of venous thromboembolism (VTE) and stroke.
The benefits of long-term (>3 years) osteoporosis drug therapy on fracture prevention and the risk of harms are less clear, including among individuals with osteopenia, though heightened risk of AFF is a concern. Because of sparse data on fracture outcomes from long-term trials and the high cost of trials adequately powered to evaluate incident fracture outcomes, many investigators have sought to identify appropriate surrogate endpoints. Early treatment changes in bone mineral density (BMD) and, to a lesser extent, in bone turnover markers, may predict short-term risk of incident fractures. However, it is unclear whether changes in these measures predict a high enough proportion of the anti-fracture efficacy of osteoporosis drug treatments to be considered adequate surrogate outcomes. Further, it needs to be established whether changes in these measures will predict long-term nonvertebral fracture risk with drug treatment, including within individuals.

The effect of long-term osteoporosis drug treatment versus control treatment may vary as a function of patient, bone, and osteoporosis drug characteristics (i.e., possible effect modifiers). Better understanding of these factors may allow prescribers and patients to collectively make more informed treatment decisions over time to maximize benefit (e.g., reduced fracture risk) while minimizing harms. However, we are unaware of any systematic literature reviews showing whether the efficacy and harms of long-term osteoporosis drug therapy vary as a function of patient characteristics.

Uncertainty about the benefits of long-term bisphosphonate use coupled with concerns that long-term bisphosphonate persistence in bone might increase fracture risk by inhibiting normal repair of bone microdamage have led to the suggestion that bisphosphonate therapy be discontinued periodically. Though several groups advocate bisphosphonate “drug holidays,” as a strategy to preserve as much fracture benefit as possible while minimizing harms, there is no consensus about who should get them, when they should start, how long they should last, how they should be monitored, or the criteria for restarting therapy.

These uncertainties about the most appropriate use of long-term osteoporosis drug therapy and of osteoporosis drug holidays spurred the scheduling of a National Institutes of Health (NIH) Office of Disease Prevention (ODP) Pathways to Prevention (P2P) program workshop on this topic. The goal of the workshop is to present an evidence-based synthesis of the pertinent research base, identify research gaps in the area, and suggest future research needs to assist patients, clinicians, and other healthcare decisionmakers. To accomplish these aims, we conducted this systematic review to address the following questions: (1) What is the efficacy of long-term osteoporosis drug therapy (>3 years) versus control on risk of incident clinical fractures (nearly always symptomatic and recognized in clinical settings, radiographs confirmed by study), incident radiographic vertebral fractures (most not symptomatic or clinically recognized, defined by comparing study radiographs), change in BMD, and harms; (2) Do risks of these fracture types and harms with long-term treatment vary as a function of patient, bone, or osteoporosis drug characteristics?; (3) Among individuals receiving osteoporosis drug therapy to prevent fracture, what is the effect of continuing versus discontinuing therapy (i.e., osteoporosis drug holiday of ≥1 year after ≥1 year of prior osteoporosis drug use) on incident fractures and harms?; and (4) Do these outcomes with osteoporosis drug holidays vary as a function of patient, bone, or osteoporosis drug characteristics?

<table>
<thead>
<tr>
<th>Table 1. Drugs Used for Osteoporosis Treatment and Prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drug class</strong></td>
</tr>
<tr>
<td>Bisphosphonate</td>
</tr>
<tr>
<td>Treatment Type</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Bisphosphonate</td>
</tr>
<tr>
<td>Bisphosphonate</td>
</tr>
<tr>
<td>Bisphosphonate</td>
</tr>
<tr>
<td>Biologic</td>
</tr>
<tr>
<td>PTH related anabolic</td>
</tr>
<tr>
<td>PTH related anabolic</td>
</tr>
<tr>
<td>SERM</td>
</tr>
<tr>
<td>Estrogen and Estrogen/ Prog est combination products</td>
</tr>
<tr>
<td>Estrogen with SERM</td>
</tr>
<tr>
<td>Anti-sclerostin monoclonal antibody</td>
</tr>
</tbody>
</table>

**Abbreviations:** IV=intravenous; PTH=parathyroid hormone, SC=subcutaneous; SERM=selective estrogen receptor modulator
*FDA approved for osteoporosis prevention, but not for osteoporosis treatment.
†Not currently FDA approved for any indication. Will be included in this review if it receives FDA approval before the close of the draft report peer/public review comment period.

**Scope and Key Questions**

**Key Questions**

Key Question 1. Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, what is the efficacy of long-term (>3 years) osteoporosis drug therapy in reducing risk of incident fracture and on change in BMD?

Key Question 2. Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, does efficacy of long-term osteoporosis drug therapy for reducing risk of incident fracture vary as a function of patient, bone, or osteoporosis drug characteristics§?

Key Question 3. Among men and postmenopausal women aged >50 years with osteoporosis* or osteopenia/low bone mass†, what is the risk of harms associated with long-term (>3 years) osteoporosis drug therapy?

Key Question 4. Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, does the risk of harms associated with long-term (>3 years) osteoporosis drug therapy vary as a function of patient, bone, or osteoporosis drug characteristics§?
Key Question 5. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy (≥1 year) started for osteoporosis* or osteopenia/low bone mass† to prevent fracture, what is the effect of osteoporosis drug treatment holidays (≥1 year) on incident fracture risk and change in BMD?

Key Question 6. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy (≥1 year) started for osteoporosis* or osteopenia/low bone mass† to prevent fracture, does the effect of osteoporosis drug treatment holidays (≥1 year) on incident fracture risk vary as a function of patient, bone or osteoporosis drug characteristics§?

Key Question 7. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy (≥1 year) started for osteoporosis* or osteopenia/low bone mass† to prevent fracture, what is the risk of harms of osteoporosis drug treatment holidays (≥1 year)?

Key Question 8. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy started for osteoporosis* or osteopenia/low bone mass† to prevent fracture, does risk of harms associated with osteoporosis drug treatment holidays vary as a function of patient, bone, or osteoporosis drug characteristics?

*Osteoporosis defined by hip or lumbar spine DXA BMD T-score < -2.5, past clinical hip or vertebral fracture, or prevalent radiographic vertebral fracture.
†Osteopenia/low bone mass defined by hip or lumbar spine DXA BMD T-score < -1.0 and > -2.5.
§ Patient characteristics (age, sex, race, osteoporosis status*, fracture history [clinical fractures, radiographic vertebral fractures], calculated fracture risk [e.g. FRAX®], comorbid conditions); Bone characteristics (BMD, biomarkers); Osteoporosis drug characteristics (dose, frequency, treatment duration, delivery route).
**Table 2. PICOTS**

Table 2 outlines the populations, interventions, comparisons, outcomes, timing, and settings (PICOTS) eligible for the present review.

<table>
<thead>
<tr>
<th>KQ</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparator</th>
<th>Health Outcomes and Harms</th>
<th>Timing</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ 1: Long-term treatment efficacy</td>
<td>Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment.</td>
<td>Osteoporosis drug treatment (see Table 1)</td>
<td>Placebo, active control</td>
<td>Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF) Intermediate: Primary: Incident radiographic vertebral fracture Secondary: DXA BMD change</td>
<td>&gt;3 years</td>
<td>Any</td>
</tr>
<tr>
<td>KQ 2: Effect modifiers of long-term treatment efficacy</td>
<td>Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment.</td>
<td>Possible effect modifiers of incident fractures with long-term treatment: Patient characteristics: pretreatment age (and years since menopause for estrogen-related treatments), race, sex, comorbid conditions (DM, CKD, CVD), osteoporosis status (osteoporosis*, low bone mass, normal), fracture history (clinical fractures, radiographic vertebral fractures), calculated pre-treatment fracture risk (e.g., FRAX®) Bone characteristics: pretreatment and early treatment (e.g. 1 year) imaging (L-spine, total hip &amp; femoral neck DXA BMD) and biochemical markers (CTX, NTX, P1NP, BSAP) Osteoporosis drug characteristics: dose, frequency, treatment duration, delivery route</td>
<td>Placebo, active control</td>
<td>Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF) Intermediate: Primary: Incident radiographic vertebral fracture Secondary: DXA BMD change</td>
<td>&gt;3 years</td>
<td>Any</td>
</tr>
<tr>
<td>KQ</td>
<td>Population</td>
<td>Intervention</td>
<td>Comparator</td>
<td>Health Outcomes and Harms</td>
<td>Timing</td>
<td>Setting</td>
</tr>
<tr>
<td>----</td>
<td>------------</td>
<td>--------------</td>
<td>------------</td>
<td>--------------------------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>KQ 3: Long-term treatment harms</td>
<td>Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment. For rare harms only: Men and PM women aged ≥50 years being studied for fracture prevention treatment regardless of baseline BMD.</td>
<td>Osteoporosis drug treatment (see Table 1)</td>
<td>Placebo, no treatment, active control</td>
<td>See Table 3 below for class-specific harms</td>
<td>&gt;3 years</td>
<td>Any</td>
</tr>
<tr>
<td>KQ 4: Effect modifiers of long-term treatment harms</td>
<td>Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment. For rare harms only: Men and PM women aged ≥50 years being studied for fracture prevention treatment regardless of baseline BMD.</td>
<td>Possible effect modifiers of harms with long-term treatment will be the same as the possible effect modifiers of incident fractures with long-term treatment detailed above for KQ 2.</td>
<td>Placebo, no treatment, active control</td>
<td>See Table 3 below for class-specific harms</td>
<td>&gt;3 years</td>
<td>Any</td>
</tr>
<tr>
<td>KQ 5: Effect of drug treatment holidays</td>
<td>Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass currently receiving osteoporosis drug therapy for fracture prevention.</td>
<td>Osteoporosis drug treatment discontinuation (placebo drug holiday) for ≥1 year after ≥1 year prior osteoporosis drug treatment</td>
<td>Continued osteoporosis drug treatment after ≥1 year prior osteoporosis drug treatment</td>
<td>Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF) Intermediate: Primary: Incident radiographic vertebral fracture Secondary: DXA BMD change</td>
<td>≥1 year osteoporosis drug discontinuation after ≥1 year prior osteoporosis drug treatment</td>
<td>Any</td>
</tr>
</tbody>
</table>
### KQ 6: Effect modifiers of effect of drug treatment holidays

**Population**: Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass currently receiving osteoporosis drug therapy for fracture prevention.

**Intervention**: Possible effect modifiers of incident fractures with osteoporosis drug treatment holidays:

- **Patient characteristics**: age, sex, race, osteoporosis status*, fracture history [clinical fractures, radiographic vertebral fractures], calculated fracture risk [e.g. FRAX®], comorbid conditions
- **Bone characteristics**: BMD, biomarkers
- **Osteoporosis drug characteristics**: pre-drug holiday agent/class, time between drug initiation and start of drug holiday, duration of drug holiday, post-drug holiday agent/class

**Comparator**: Continued osteoporosis drug treatment after ≥1 year prior osteoporosis drug treatment

**Health Outcomes and Harms**: Final:
- Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF)
- Intermediate:
  - **Primary**: Incident radiographic vertebral fracture

**Timing**: >1 year osteoporosis drug discontinuation after ≥1 year prior osteoporosis drug treatment

**Setting**: Any

### KQ 7: Harms of drug treatment holidays

**Population**: Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass currently receiving osteoporosis drug therapy for fracture prevention.

**Intervention**: Osteoporosis drug treatment discontinuation (placebo drug holiday) for ≥1 year after ≥1 year prior osteoporosis drug treatment

**Comparator**: Continued osteoporosis drug treatment after ≥1 year prior osteoporosis drug treatment

**Health Outcomes and Harms**: See Table 2 below for class-specific harms

**Timing**: >1 year osteoporosis drug discontinuation after ≥1 year prior osteoporosis drug treatment

**Setting**: Any

### KQ 8: Effect modifiers of harms of drug treatment holidays

**Population**: Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass currently receiving osteoporosis drug therapy for fracture prevention.

**Intervention**: Possible effect modifiers of harms with osteoporosis drug treatment holidays will be the same as the possible effect modifiers of incident fractures with drug holidays detailed above for KQ 6.

**Comparator**: Continued osteoporosis drug treatment after ≥1 year prior osteoporosis drug treatment

**Health Outcomes and Harms**: See Table 2 below for class-specific harms

**Timing**: >1 year osteoporosis drug discontinuation after ≥1 year prior osteoporosis drug treatment

**Setting**: Any

---

**Abbreviations**: BMD=bone mineral density; BSAP=bone specific alkaline phosphatase; CKD=chronic kidney disease; CTX=C-terminal telopeptide; CVD=cardiovascular disease; DM=diabetes mellitus; DXA=dual x-ray absorptiometry; FRAX=World Health Organization (WHO) Fracture Risk Assessment Tool; KQ=key question; MOF=major osteoporotic fracture; NTX=N-terminal telopeptide; P1NP=procollagen I intact N-terminal; PICOTS=populations, interventions, comparators, outcomes, timing, and settings/study design; PM=postmenopausal; RCT=randomized controlled trial

*Osteoporosis defined by hip or lumbar spine DXA BMD T-score of ≤-2.5 or worse, past clinical hip or vertebral fracture, or prevalent radiographic vertebral fracture.

†Osteopenia/low bone mass defined by hip or lumbar spine DXA BMD T-score < -1.0 and ≤-2.5.
Table 3: Class-Specific Harms of Drugs Used for Osteoporosis Treatment

<table>
<thead>
<tr>
<th>Drug Class</th>
<th>Harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphosphonates</td>
<td>ONJ*, AFF*, atrial fibrillation*, heart attacks, musculoskeletal pain, upper GI intolerance, esophageal cancer</td>
</tr>
<tr>
<td>Biologic (Denosumab)</td>
<td>ONJ*, AFF*, atrial fibrillation*, heart attacks, musculoskeletal pain, upper GI intolerance, esophageal cancer, infection, fracture after stopping therapy</td>
</tr>
<tr>
<td>PTH related anabolic (recombinant PTH and PTH analogue)</td>
<td>Hypercalcemia, hypercalciuria, osteosarcoma, fracture after stopping therapy, upper GI intolerance</td>
</tr>
<tr>
<td>SERM</td>
<td>Stroke, venous thromboembolic disease (PE, DVT), hot flashes, mild cognitive impairment, dementia, mortality</td>
</tr>
<tr>
<td>Estrogen, Estrogen with Progestin, Estrogen with SERM</td>
<td>Cardiovascular disease (heart attack, stroke), venous thromboembolic disease (PE, DVT), cancer (breast, ovarian, endometrial, colorectal), mild cognitive impairment, dementia, mortality</td>
</tr>
<tr>
<td>Antisclerostin monoclonal antibody</td>
<td>Cardiovascular disease (heart attack, stroke), ONJ*, AFF*</td>
</tr>
</tbody>
</table>

*Investigated as rare harms for this report. Some studies reported subtrochanteric/femoral shaft fractures without radiological AFF features as a proxy for AFF.

Abbreviations: AFF=atypical femoral fracture; DVT=deep venous thrombosis; GI=gastrointestinal; ONJ=osteonecrosis of the jaw; PE=pulmonary embolism; PTH=parathyroid hormone; SERM=selective estrogen receptor modulator
Analytic Framework

Figure 1 outlines the analytic framework used to guide the present review.

Figure 1. Analytic Framework for Long-term Drug Therapy and Drug Holidays for Osteoporosis Fracture Prevention

KQ1

Men & PM women ≥50 yrs of age with osteoporosis* or osteopenia/low bone mass†

KQ2

Short-term osteoporosis drug treatment

KQ3

Long-term osteoporosis drug treatment

KQ4

Osteoporosis drug holiday

KQ5

Treatment outcome predictors
- Patient
- Bone
- Osteoporosis drug

KQ6

Final health outcomes
- Incident clinical fx
  - Any
  - Hip
  - Vertebral
  - Nonhip nonvertebral
  - MOF
- Intermediate health outcomes
  - Primary: Incident radiographic vertebral fx
  - Secondary: BMD change

KQ7

Treatment harms

KQ8

BMD=bone mineral density
Fx=fracture
MOF=major osteoporotic fracture
PM=postmenopausal

KQ9
Chapter 2. Methods

This comparative effectiveness review (CER) follow methods suggested in the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews (http://www.effectivehealthcare.ahrq.gov/methodsguide.cfm); certain methods map to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.81

Topic Refinement and Review Protocol

The NIH ODP Working Group, which included individuals from the ODP, National Institute on Aging (NIA) and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), developed the original Key Questions (KQs). We refined the KQs in collaboration with the NIH ODP Working Group, a NIH Content Area Expert Group, and a Technical Expert Panel (TEP). The resulting KQs were incorporated into the final protocol, which was posted April 12, 2018 at https://effectivehealthcare.ahrq.gov/topics/osteoporosis-fracture-prevention/research-protocol and registered in PROSPERO.

Literature Search Strategy

We searched Ovid Medline, Ovid Embase, and the Cochrane Central Register of Controlled Trials (CENTRAL) to identify previous RCTs, controlled clinical trials (CCTs), and controlled observational studies published and indexed in bibliographic databases. Our search strategy, Appendix A, included relevant medical subject headings and natural language terms for the concepts of osteoporosis and drug treatment. These concepts were combined with validated filters to select study designs. Dates for the search algorithm were 1995 to June 2018. We supplemented our searches with backward citation searches of relevant systematic reviews published from 2012 and onward. We will update searches while the draft report is under public/peer review.

We searched ClinicalTrials.gov to identify additional relevant completed and ongoing studies. AHRQ also opened a Supplemental Evidence and Data for Systematic Reviews (SEADs) portal to solicit pharmaceutical manufacturer protocols with additional information about published or unpublished drug studies. Grey literature search results were used to identify studies, outcomes, and analyses not reported in the published literature to assess publication and reporting bias and inform future research needs.

We reviewed studies relevant to inclusion criteria based on our population, intervention, comparators, timing, and settings (PICOTS) framework outlined in Table 4.

Table 4. Study inclusion criteria

<table>
<thead>
<tr>
<th>Category</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
</table>
| Study Population | Adults aged ≥50 years, including men and postmenopausal women  
                   All key questions: Participants with osteoporosis (osteoporosis defined as hip or vertebral DXA BMD T-score ≤-2.5, past clinical hip or vertebral fracture, or radiographic vertebral fracture) or osteopenia/low bone mass (hip or vertebral BMD T-score >-2.5 and <-1) being treated to prevent fractures.  
                   For rare harms (included in key questions 3, 4, 7 & 8): also include participants without osteoporosis/osteopenia or with unknown osteoporosis/osteopenia status being treated to prevent fractures.  
                   Exclude: Studies focused on populations with known secondary causes of osteoporosis (e.g., transplant, spinal cord injury, exogenous glucocorticoids, hormone suppressive therapy, |
<table>
<thead>
<tr>
<th>Category</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>endogenous hypercortisolism, hyperparathyroidism, hyperthyroidism); though will include studies focused on populations with CKD, DM, or CVD. Exclude studies focused on patients with cancer metastatic to bone or focused on drug effects on acute fracture healing.</td>
</tr>
<tr>
<td>Study Objectives</td>
<td>To evaluate the efficacy and harms of long-term osteoporosis drug treatment (&gt;3 years), the efficacy and harms of osteoporosis drug treatment continuation versus discontinuation (placebo drug holidays), possible effect modifiers of long-term osteoporosis drug treatment on incident fractures and harms, and possible effect modifiers of osteoporosis drug treatment holidays on risk of incident fractures and harms.</td>
</tr>
<tr>
<td>Study Design</td>
<td>All key questions: RCTs, CCTs For assessment of harms (KQ 3, 4, 7 &amp; 8): Also include observational studies with contemporaneous human controls that employed methods to account for selection bias (adjust for demographics and some measure of fracture risk [e.g., past fracture, BMD or fracture risk calculator]). For assessment of rare harms (i.e., AFF, ONJ, atrial fibrillation): Sample size must be ≥100 and may include case-control, retrospective or prospective cohort, or administrative data studies. For nonrare harms: Sample size must be ≥1000 and limited to prospective cohort studies. For all key questions: Exclude case report, case series, post-marketing reports.</td>
</tr>
<tr>
<td>Interventions</td>
<td>Drugs FDA approved for osteoporosis treatment or prevention (bisphosphonates, denosumab, teriparatide, abaloparatide, estrogen*, estrogen/progesterone*, SERM, estrogen/SERM*, romosozumab†) Discontinuation of osteoporosis drug treatment</td>
</tr>
<tr>
<td>Comparisons</td>
<td>For treatment efficacy and harms: Placebo, active contemporaneous control For osteoporosis drug treatment discontinuation (holiday): Continued osteoporosis drug treatment</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Final health outcomes: incident clinical fracture (e.g., any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF) Intermediate health outcomes: Primary: Incident radiographic vertebral fracture, Secondary: Change in BMD (BMD change outcome was assessed only in studies that also reported incident clinical fracture or incident radiographic vertebral fracture outcomes) Harms (Serious adverse events and specific harms as listed in Table 3)</td>
</tr>
<tr>
<td>Outcome predictors (applicable only for KQs 2, 4, 6 &amp; 8)</td>
<td>Patient characteristics: pretreatment age, race, sex, comorbid conditions (DM, CKD, CVD), osteoporosis status (osteoporosis, osteopenia/low bone mass, normal), fracture history (clinical fracture, radiographic vertebral fracture), calculated pre-treatment fracture risk (e.g., FRAX®) Bone characteristics: pretreatment and early treatment (e.g., 1 year) imaging (lumbar spine, total hip &amp; femoral neck DXA BMD) and biochemical markers (CTX, NTX, P1NP, bone-specific ALP) Osteoporosis drug characteristics: For KQ 2 &amp; 4: dose, frequency, treatment duration, delivery route; For KQ 6 &amp; 8: pre-drug holiday agent/class, time between drug initiation and start of drug holiday, duration of drug holiday, post-drug holiday agent/class</td>
</tr>
<tr>
<td>Timing</td>
<td>Long-term osteoporosis drug treatment: treatment duration &gt;3 years. Osteoporosis drug treatment holidays: treatment cessation ≥1 year after prior osteoporosis drug treatment ≥1 year</td>
</tr>
<tr>
<td>Setting</td>
<td>Any</td>
</tr>
<tr>
<td>Publication type</td>
<td>Published in full text in peer reviewed journals. Will use systematic reviews and eligible studies to identify additional references. Data was supplemented by grey literature if it included sufficient information to assess eligibility and risk of bias.</td>
</tr>
<tr>
<td>Language of Publication</td>
<td>English</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=Atypical femoral fracture; ALP=alkaline phosphatase; BMD=bone mineral density; CCT=controlled clinical trial; CKD=chronic kidney disease; CTX=C-terminal telopeptide; CVD=cardiovascular disease; DM=diabetes mellitus; DXA=dual x-ray absorptiometry; FDA=U.S. Food and Drug Administration; FRAX=World Health Organization (WHO) Fracture Risk Assessment Tool; KQ=key question; MOF=major osteoporotic fracture; NTX=N-terminal telopeptide; ONJ=Osteonecrosis of the jaw; P1NP=procollagen I intact N-terminal; RCT=randomized controlled trial *FDA approved for osteoporosis prevention, but not for osteoporosis treatment. †Not currently FDA approved for any indication. Will be included in this review if it receives FDA approval before the close of the draft report peer/public review comment period.
Study Selection and Risk of Bias Assessment

We screened titles and abstracts of all references identified from our bibliographic database search, references from relevant systematic reviews, and grey literature. Studies considered possibly eligible based on our inclusion criteria by at least one of two independent reviewers were flagged for full text screening. Then, two independent reviewers screened the full text publications to determine if inclusion criteria were met. Differences in screening decisions were resolved by discussion between investigators, and, if necessary, consultation with a third investigator or team consensus. For studies excluded at the full text review stage, reasons for ineligibility were documented. Reviewers regularly met to discuss inclusion criteria and ensure consistency between reviewers.

Based on AHRQ guidance,82 we assessed risk of bias (ROB) of eligible studies in their design, analysis and reporting. ROB was assessed by two independent investigators using a modified instrument we created for a previous AHRQ review (Appendix B) as a guide. Differences in ROB decisions were resolved via reviewer discussion and if necessary, team consensus. We selected items most relevant in assessing ROB for this topic, including: (1) participant selection (adequacy of randomization method); (2) attrition/incomplete outcome data (loss to follow-up, both overall and differentially between treatment groups); (3) detection bias (outcome assessor masking, outcome measurement quality); (4) performance bias (intention to treat analysis, adjustment for potential confounding variables, participant masking to treatment assignment); and (5) reporting bias (selective outcomes reporting). Summary ROB assessments for each study were classified as low, medium, or high based upon the collective ROB inherent in each domain and confidence that the results were believable given the study’s limitations (Appendix C).

ROB was assessed separately for the following outcomes: incident clinical fractures (any, nonvertebral, hip, vertebral, nonhip nonvertebral, major osteoporotic fracture [MOF]), incident radiographic vertebral fractures, AFF with confirmed radiological features, subtrochanteric/femoral shaft fractures, ONJ, and incident clinical fracture after stopping therapy (rebound fractures). Differences in ROB ratings between investigators were resolved by discussion, and, if necessary, team consensus. Screening and ROB assessment were performed in Distiller (DistillerSR, Evidence Partners, Ottawa, Canada).

Data Extraction

For studies meeting inclusion criteria, one investigator extracted relevant data into extraction forms created in Microsoft Word. Evidence tables were reviewed and verified for accuracy by a second investigator.

Limited data were extracted from studies with high ROB: author, publication year, study design, intervention, types of efficacy outcomes, and whether adverse effects were reported (Appendix D7). For studies with low or medium ROB, we extracted additional information: participant inclusion criteria, setting, participant baseline characteristics (age, race, sex, comorbid conditions [diabetes mellitus (DM), chronic kidney disease (CKD), cardiovascular disease (CVD)], osteoporosis versus osteopenia/low bone mass, fracture history, calculator estimated fracture risk, BMD, bone turnover marker levels (C-terminal telopeptide [CTX], N-terminal telopeptide [NTX], procollagen I intact N-terminal [P1NP], bone-specific alkaline phosphatase [BSAP]), intervention details (drug class, name, dose and delivery route), control
intervention details, follow-up duration, study funding, and results of efficacy outcomes and adverse effects (Appendix D).

Data Synthesis

Results were organized first by KQ, then treatment comparison, then treatment outcome (incident clinical fractures, incident radiographic vertebral fractures, change in BMD, harms).

For studies with low and moderate risk of bias, we summarized results in evidence tables. Because there were few treatment-outcome comparisons with more than one eligible study and those had substantial clinical and methodological heterogeneity (participant population, intervention, outcome measures), we judged that statistical pooling was inappropriate and conducted qualitative synthesis.

When applicable data was available within individual studies or across multiple studies (i.e., tests of interaction, stratified results, special population), we evaluated a priori selected possible effect modifiers of osteoporosis drug treatment and osteoporosis drug treatment holidays on efficacy and harms outcomes (i.e., age, race, sex, comorbid conditions [DM, CKD, CVD], osteoporosis status, fracture history, calculated estimated fracture risk, BMD, CTX, NTX, P1NP, bone-specific alkaline phosphatase, drug dose, frequency, treatment duration and delivery route, and follow-up duration, pre-drug holiday agent/class, time between osteoporosis drug initiation and start of drug holiday, duration of drug holiday, post-drug holiday agent/class).

Strength of Evidence for Major Comparisons and Outcomes

The overall strength of evidence (SOE) was evaluated by two independent reviewers. Differences were resolved by consultation between investigators, and, if necessary, consultation with a third investigator. SOE was rated individually for the following efficacy outcomes: any incident clinical fracture, incident nonvertebral fracture, incident hip fracture, incident clinical vertebral fracture, incident nonhip nonvertebral fracture, incident MOF, and incident radiographic vertebral fractures. Strength of evidence also was graded for the following harms: serious adverse events, AFF with confirmed radiologic features, subtrochanteric/femoral shaft fractures without confirmed radiologic AFF features, ONJ, and incident clinical fracture after stopping therapy (rebound fractures).

SOE ratings were based on five required domains: (1) study limitations (based on individual study ROB); (2) directness (a single, direct link between intervention and outcome); (3) consistency (similarity of effect direction and size); (4) precision (degree of certainty around an estimate); and (5) reporting bias. Study limitations were rated as low, medium, or high. Consistency among studies was rated as consistent, inconsistent, or unknown/not applicable (e.g., single study). Directness was rated as direct or indirect. Precision was rated as precise or imprecise. An imprecise estimate was one for which the confidence interval was wide enough to include clinically distinct conclusions. Precision was evaluated using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system guidelines. Other factors considered in assessing strength of evidence included dose-response relationships, the presence of confounders, and strength of association.

Based on these factors, the overall evidence for each outcome (Appendix D) was rated as: High: Very confident that estimate of effect lies close to true effect. Few or no deficiencies in body of evidence, findings believed to be stable.

Moderate: Moderately confident that estimate of effect lies close to true effect. Some deficiencies in body of evidence; findings likely to be stable, but some doubt.
**Low**: Limited confidence that estimate of effect lies close to true effect; major or numerous deficiencies in body of evidence. Additional evidence necessary before concluding that findings are stable or that estimate of effect is close to true effect.

**Insufficient**: No evidence, unable to estimate an effect, or no confidence in estimate of effect. No evidence is available or the body of evidence precludes judgment.

An overall rating of high SOE would be assigned when included studies were RCTs with low risk of bias, and the results were consistent, direct, and precise.

**Applicability**

Applicability of findings was determined according to the PICOTS framework. Study characteristics that affected applicability included, but were not limited to, the population (age, race, sex, presence or lack of comorbidities, country from which the study participants were enrolled), narrow eligibility criteria, and patient and intervention characteristics potentially associated with treatment response different than those described by population studies.\(^{85}\)

**Peer Review and Public Commentary**

AHRQ staff and an AHRQ associate editor will review the draft report. After we revise the report based on this review, experts in osteoporosis, primary care, geriatrics, and systematic reviews will be invited to provide external peer review. The draft report will again be revised and will be posted on the AHRQ website for 4 weeks to elicit public comment. We then will address all public reviewer comments, revise the text as appropriate, and document all comments, responses and revisions in a disposition of comments report that we will submit with the final report for posting on the Effective Health Care website.
Chapter 3. Search Results

We identified 8251 unique citations (Figure 2) from 1995 to June 2018 from bibliographic databases addressing osteoporosis and drug treatment. An initial title and abstract review excluded 6,936 publications that were not related to relevant drug treatments for patients with osteoporosis. Full texts of 1,320 publications were reviewed to determine final inclusion. Appendix E provides a list of publications excluded at full text review. We identified 58 unique references eligible for inclusion, which are summarized in Table 5. ROB assessments for these references can be found in Appendix C.

Of 46 publications with low or medium ROB, there were 33 randomized or controlled clinical trials (8 unique studies) and 12 controlled observational studies (10 unique studies) (Appendix C). Most publications were based on three RCTs of alendronate, zoledronic acid and raloxifene, respectively, and their extension studies.

All trials enrolled only postmenopausal women, with most limited to women with osteoporosis as defined by BMD and vertebral fracture history, and a few also including women with osteopenia. The observational studies included between 84 to 100 percent women. Mean participant age was 72 years, with all but two studies reporting a mean age of <80 years. Most of the observational studies presumed that participants had osteoporosis because of past fracture or their use of osteoporosis medications, but none reported BMD status.
Figure 2. Literature flow diagram

Bibliographic database searches
8,251 references

Title and abstract review excluded
6,936 references

Hand search (reference lists of
included studies)
5 references

Pulled for full text review
1,320 references

Excluded
1,262 references
Not available in English = 10
Ineligible study design = 437
Ineligible treatment duration = 352
Ineligible population = 77
No intervention of interest = 43
No outcomes of interest = 343

Eligible references=58
(26 unique studies, including
18 with low or medium risk of bias)
Table 5. Eligible Publications that Compared Long-term Osteoporosis Drug Use versus Control or Osteoporosis Drug Continuation versus Discontinuation (Placebo Drug Holiday) and Reported on Risk of Incident Fractures and/or Harms

<table>
<thead>
<tr>
<th>Drug name</th>
<th>Drug class</th>
<th>Low or medium ROB publications (analyzed)</th>
<th>High ROB publications (not analyzed)</th>
<th>Total number of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate Bisphosphonate</td>
<td>Bisphosphonate</td>
<td>13 RCTs or CCTs23, 30-33, 37, 53, 54, 62-64, 67, 68, 86 (11 FIT/FLEX)</td>
<td>2 RCTs88, 89 (FIT) 4 observational110, 116, 117</td>
<td>21 publications: 15 RCTs 8 observational</td>
</tr>
<tr>
<td>Zoledronic acid Bisphosphonate</td>
<td>Bisphosphonate</td>
<td>3 RCTs65, 66, 68 (2 HORIZON extensions)</td>
<td>1 RCT90 (HORIZON extension)</td>
<td>4 publications: 4 RCTs</td>
</tr>
<tr>
<td>Risedronate Bisphosphonate</td>
<td>Bisphosphonate</td>
<td>0</td>
<td>2 RCTs95, 96</td>
<td>2 publications: 2 RCTs</td>
</tr>
<tr>
<td>Ibandronate Bisphosphonate</td>
<td>Bisphosphonate</td>
<td>0</td>
<td>1 RCT97</td>
<td>1 RCT</td>
</tr>
<tr>
<td>Any bisphosphonate (studies reported results as a class)</td>
<td>Bisphosphonate</td>
<td>5 observational102-114, 115, 116</td>
<td>1 observational98</td>
<td>6 publications: 6 observational</td>
</tr>
<tr>
<td>Denosumab</td>
<td>Biologic</td>
<td>1 RCT118</td>
<td>1 RCT119</td>
<td>2 publications: 2 RCTs</td>
</tr>
<tr>
<td>Teriparatide (recombinant PTH)</td>
<td>PTH related anabolic</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Abaloparatide (PTH analogue)</td>
<td>PTH related anabolic</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Raloxifene SERM</td>
<td>SERM</td>
<td>14 RCTs or CCTs24, 25, 34, 36, 49, 50, 55-61, 62-64, 65 (14 MORE/CORE)</td>
<td>0</td>
<td>14 publications: 14 RCTs or CCTs</td>
</tr>
<tr>
<td>Estrogen and Estrogen/ Progestin Combination products</td>
<td>Biologic</td>
<td>3 RCT27-29</td>
<td>0</td>
<td>3 publications 2 RCTs</td>
</tr>
<tr>
<td>Conjugated estrogens/ bazedoxifene*</td>
<td>SERM</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Romosozumab†</td>
<td>Anti-sclerostin monoclonal antibody</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pooled osteoporosis drugs</td>
<td>Bisphosphonates, raloxifene</td>
<td>3 observational130, 45, 48</td>
<td>0</td>
<td>3 publications: 3 observational</td>
</tr>
<tr>
<td>Total number of publications</td>
<td></td>
<td>46 publications: 33 RCTs 12 observational</td>
<td>12 publications: 7 RCTs 5 observational</td>
<td>58 publications: 40 RCTs 17 observational</td>
</tr>
</tbody>
</table>

**Abbreviations:** CCT=Controlled clinical trial; CORE=Continuing Outcomes Relevant to Evista; FIT=Fracture Intervention Trial; FLEX=Fracture Intervention Trial Long Term Extension; HORIZON=Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly trial; MORE=Multiple Outcomes of Raloxifene Evaluation; PTH = parathyroid hormone; RCT=randomized controlled trials; ROB=risk of bias; SERM=selective estrogen receptor modulator

*FDA approved for osteoporosis prevention, but not for osteoporosis treatment.
†Not currently FDA approved for any indication. Will be an eligible treatment for the review if it receives FDA approval before the end of the draft report peer/public review comment period.
Chapter 4. Efficacy of Long-term Osteoporosis Drug Therapy

Chapter 4 reviews the evidence from eligible studies that addresses Key Questions 1 and 2. Key Question 1 examines the efficacy of long-term (>3 years) osteoporosis drug therapy versus inactive control (e.g., placebo, no therapy) or a different active therapy in reducing risk of incident fracture and on change in BMD. Key Question 2 examines whether efficacy of long-term osteoporosis drug therapy versus control in reducing risk of incident fracture varies as a function of patient, bone or drug characteristics. We report the literature on the effect of continuing osteoporosis drug therapy versus discontinuing it (placebo drug holiday), including on the effect of long-term versus shorter term therapy, on risk of fractures under Key Question 5 in Chapter 6 of this report.

This chapter is organized by drug, with Key Question 1 and Key Question 2 discussed sequentially within each drug subsection. Drug versus placebo or no therapy comparisons are listed first, followed by any comparisons of drug versus another active drug therapy. We provide an overview at the start of specific drug subsections when the literature base was complex. Summary statements about the included studies are below; individual study details are provided in the report tables and appendices.

Alendronate

Key Points

Key Question 1
- In postmenopausal women with FN-BMD T-score \( \leq -1.6 \) (mix of osteoporosis and osteopenia) but no baseline vertebral fracture, between alendronate versus placebo for 4 years:
  - There was no difference in risk of incident clinical fractures (moderate SOE), incident nonvertebral fractures (moderate SOE), or incident hip fractures (low SOE).
  - There was a lower risk of incident radiographic vertebral fractures (high SOE).

Key Question 2
- In postmenopausal women with no baseline vertebral fracture, the effect of alendronate therapy versus placebo for 4 years was modified by baseline BMD:
  - In women with FN-BMD T-score < -2.5 (osteoporosis), alendronate was associated with a lower risk of incident clinical fractures (moderate SOE) and incident radiographic vertebral fractures (moderate SOE).
  - In women with FN-BMD T-score -2 to -2.5 (osteopenia), there was no difference between treatment groups in risk of incident clinical fractures (low SOE) and possibly no difference in incident radiographic vertebral fractures (low SOE).
  - In women with FN-BMD T-score -1.6 to -2 (osteopenia), there was no difference between treatment groups in risk of incident clinical fractures (low SOE), but evidence was insufficient to draw conclusions about risk of incident radiographic vertebral fractures.
• In postmenopausal women with baseline FN-BMD -1.6 to > -2.5 (osteopenia), the effect of therapy with alendronate versus placebo for 4 years:
  o On risk of incident nonvertebral fracture was not modified by history of prior nonvertebral fracture.
  o On risk of incident clinical and radiographic vertebral fracture was not modified by presence of baseline radiographic vertebral fracture.
  o On risks of incident clinical fracture, nonvertebral fracture, major osteoporotic fracture, and radiographic vertebral fracture was not modified by baseline FRAX score.
• In postmenopausal women, whether with osteoporosis (baseline vertebral fracture or FN-BMD < -2.5) or not (no baseline vertebral fracture and FN-BMD -1.6 to > -2.5), the effect of treatment with alendronate versus placebo for 4 years was not modified by pretreatment levels of bone turnover markers.

Eligible Studies

Nine eligible publications of three unique studies compared long-term alendronate treatment versus placebo, no treatment or a different active treatment, and reported on risk of incident fractures. We rated four of these studies as having high ROB and extracted only limited data (Appendix D). For the remaining five studies with low or medium ROB, additional information was extracted in evidence tables and SOE summary tables, available in Appendix D.

All five publications with low or medium ROB were from the Fracture Intervention Trial (FIT), in which participants were randomized to alendronate versus placebo. No eligible RCTs reporting incident fracture compared long-term alendronate versus a different active treatment. Trials that compared the efficacy of long-term alendronate versus shorter term alendronate on risk of incident fractures are reported under Key Question 5 about osteoporosis drug treatment holidays in Chapter 6 of this report.

FIT was conducted in the US and enrolled postmenopausal women aged 55-80 years with osteoporosis or osteopenia/low bone mass (FN-BMD ≤ 0.68g/cm2 [T score < -1.6]). FIT excluded women with recent or severe peptic ulcers, dyspepsia that required daily treatment and significant renal dysfunction. Prior osteoporosis drug treatment was restricted. All participants were randomized to alendronate 5 mg daily versus placebo, with the alendronate dose changed to 10 mg daily at the 24-month clinic visit based on emerging evidence from other trials.

FIT was comprised of two parallel RCTs. FIT-I enrolled 2027 women with baseline radiographic vertebral fracture, had a mean treatment duration of 2.9 years, and was excluded from this review because treatment duration did not exceed three years as required to define long-term treatment for this review. FIT-II enrolled 4432 women without baseline radiographic vertebral fracture, had a mean treatment duration of 4 years, was rated as low ROB and was included in this review. There also were multiple reports of post hoc analyses of FIT-II or pooled analyses of FIT-I plus FIT-II. We included the pooled reports if either mean or median treatment duration was reported as greater than 3 years, or, if neither mean nor median was reported, if more than 50 percent of the pooled sample came from FIT-II and the article otherwise met inclusion criteria. Four publications met these criteria, were rated as medium ROB and were included in this review.

Baseline participant characteristics for the FIT-II RCT and the four related reports are listed in Table 6 (below) and Appendix D2. The mean age of FIT-II women was 68 years and
most women were of white race (97%). Thirty-seven percent of the FIT-II sample had osteoporosis at baseline (FN-BMD T score <-2.5).

Although the 2.9-year FIT-I RCT was excluded from this review of long-term treatment, its baseline participant characteristics are included in Table 6 for comparison, since all or part of FIT-I was pooled in three included reports that contained FIT-II data.30, 32, 33

Table 6. Baseline participant characteristics from included alendronate long-term treatment efficacy publications for KQ 1 and KQ 2 and FIT-I

<table>
<thead>
<tr>
<th></th>
<th>Cummings 1998 FIT-II</th>
<th>Ryder 2008 FIT-II subgroup</th>
<th>Donaldson 2012 FIT-I &amp; II</th>
<th>Quandt 2005 pooled FIT-I &amp; II subset</th>
<th>Bauer 2006 pooled FIT-I &amp; II</th>
<th>FIT-I excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample N</td>
<td>4432</td>
<td>2785</td>
<td>6459</td>
<td>3737</td>
<td>6186</td>
<td>2027</td>
</tr>
<tr>
<td>N (%) from FIT-II</td>
<td>100%</td>
<td>100%</td>
<td>4432 (68.6)</td>
<td>2797 (74.8)</td>
<td>NR</td>
<td>0</td>
</tr>
<tr>
<td>% of FIT-II sample</td>
<td>100%</td>
<td>62.8%</td>
<td>100%</td>
<td>63.1%</td>
<td>ND</td>
<td>0</td>
</tr>
<tr>
<td>Mean follow-up (yr)</td>
<td>4.2 yr</td>
<td>4.2 yr</td>
<td>4 yr*</td>
<td>3.4-5 yr</td>
<td>3.2 yr</td>
<td>2.9 yr</td>
</tr>
<tr>
<td>Mean age (yr)</td>
<td>68</td>
<td>68</td>
<td>68</td>
<td>69</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Age &lt;65 yr (%)</td>
<td>34</td>
<td>38</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>16</td>
</tr>
<tr>
<td>Age 65-74 yr (%)</td>
<td>53</td>
<td>52</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>57</td>
</tr>
<tr>
<td>Age 75-81 yr (%)</td>
<td>13</td>
<td>11</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>27</td>
</tr>
<tr>
<td>Female (%)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>White race (%)</td>
<td>97</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>97</td>
</tr>
<tr>
<td>Baseline RVF (%)</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>25</td>
<td>31</td>
<td>100</td>
</tr>
<tr>
<td>% Prior fracture since age ≥ 45 yr</td>
<td>36</td>
<td>32</td>
<td>44</td>
<td>37</td>
<td>NR</td>
<td>58</td>
</tr>
<tr>
<td>Enrolled T-score</td>
<td>≤ -1.6</td>
<td>-1.0 to -2.5</td>
<td>≤ -1.6</td>
<td>-1.6 to -2.5</td>
<td>≤ -1.6</td>
<td>≤ -1.6</td>
</tr>
</tbody>
</table>

Abbreviations: FIT=Fracture Intervention Trial; KQ=key question; NR=not reported for this study subsample; RVF=radiographic vertebral fracture; yr=year
*Calculated by EPC

Outcomes

All evidence on the long-term efficacy of alendronate versus placebo on risk of incident fractures and change in BMD comes from the single, eligible, low ROB FIT-II RCT, described in detail above.23

Incident Clinical Fractures

In the FIT-II RCT, alendronate for 4 years versus placebo did not significantly reduce risk of incident clinical fracture (12.3% vs. 14.1%; hazard ratio [HR] 0.86 [95% CI 0.73, 1.01]) (moderate SOE), incident nonvertebral fracture (11.8% vs. 13.3%; HR 0.88 [95% CI 0.74, 1.04]) (moderate SOE), or incident hip fracture (0.9% vs. 1.1%, HR 0.79 [95% CI 0.43, 1.44]) (low SOE).23

Incident Radiographic Vertebral Fractures

Compared with placebo, alendronate was associated with a significantly lower risk of incident radiographic vertebral fracture (2.1% vs. 3.8%; HR 0.56 [95% CI 0.39, 0.80]) (high SOE).
Change in BMD

Alendronate for four years was associated with statistically significant increases in BMD compared with placebo, ranging from 3.4 percent for total hip (TH) BMD to 8.3 percent for LS-BMD (p<0.001 for all between-group differences).

Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We looked at whether the efficacy of long-term alendronate therapy versus inactive control on risk of incident fracture varies as a function of patient, bone, and/or drug characteristics. Eligible studies included evidence for the following possible effect modifiers.

Baseline BMD

In a FIT-II subgroup analysis planned before treatment assignments were unblinded, the effect of long-term alendronate versus placebo on risks of any incident clinical fracture and incident radiographic vertebral fracture varied by baseline BMD.23 While risk of incident clinical fracture was significantly lower with four years of alendronate versus placebo in women with a baseline FN-BMD T-score of <-2.5 (13.1% vs. 19.6%; HR 0.64 [95% CI 0.50, 0.82]), there was no difference between treatment groups in women with FN-BMD T-scores between -2.5 to -2.0 or between -2.0 to -1.6 (p-value for interaction 0.01). Risk of incident radiographic vertebral fracture was significantly lower with four years of alendronate versus placebo in women with a baseline FN-BMD T-score of <-2.5 (2.9% vs. 5.8%; HR 0.50 [95% CI 0.31, 0.82]). The reduction in risk appeared similar in magnitude in the FN-BMD T-score -2.5 to -2 subgroup (1.9% vs. 3.6%; HR 0.54 [95% CI 0.28, 1.04]), but was not statistically significant, likely in part to there being fewer women in this BMD subgroup with such fractures. There appeared to be no effect of alendronate versus placebo on risk of incident radiographic vertebral fractures in women with FN-BMD T-scores between -2.0 to -1.6 (1.3% vs. 1.5%; HR 0.82 [0.33, 2.07]) (p-value for interaction was not reported) (Appendix D3).

In a post hoc FIT-II analysis, alendronate was associated with a significantly lower risk of incident hip fracture in women with baseline FN-BMD T-score of <-2.5, (HR 0.44 [95% CI 0.18, 0.97]), but not in those with baseline FN-BMD T-score >-2.5 (HR 1.84 [95% CI 0.70, 5.36]) (p-value for interaction not reported).

Prior Fracture

In a post hoc analysis of the subgroup of FIT-II women with FN-BMD T-score between -1.6 and -2.5 (medium ROB), the effect of 4 years of alendronate versus placebo on risk of incident nonvertebral fracture was insignificant overall, and not different between individuals with and without a prior nonvertebral fracture after age 45 (p-value for interaction 0.37) (Appendix D3).

In a post hoc analysis that pooled FIT-I and FIT-II women with FN-BMD T-scores between -1.6 to -2.5 (medium ROB), the effect of four years of alendronate versus placebo on risks of incident clinical and radiographic vertebral fracture, which favored alendronate overall, was not significantly different between individuals with and without a radiographic vertebral fracture at study baseline (p-value for interactions 0.44 and 0.54, respectively) (Appendix D3).
Combined Baseline BMD and Prevalent Radiographic Vertebral Fracture Status

In a post hoc analysis that pooled data from all FIT-I and FIT-II participants (medium ROB), the effect of four years of alendronate versus placebo on risk of incident fractures was evaluated as a function of baseline osteoporotic status. Participants were categorized at baseline as either osteoporotic (FN-BMD ≤ -2.5 or baseline radiographic vertebral fracture) or not osteoporotic (FN-BMD > -2.5 and no baseline radiographic vertebral fracture). In osteoporotic women, alendronate lowered risk of both incident nonvertebral fracture (HR 0.69 [95% CI 0.58, 0.83]) and incident radiographic vertebral fracture (HR 0.50 [95% CI 0.39, 0.65]) compared with placebo. However, in nonosteoporotic women, there was no between-treatment difference in risk of either fracture outcome. No test for interaction was reported.

FRAX Score

In a post hoc analysis that pooled data from all FIT-I and FIT-II participants (medium ROB), the effect of four years of alendronate versus placebo on risk of incident clinical fracture, nonvertebral fracture, MOF, and radiographic vertebral fracture were not significantly different between individuals with higher versus lower FRAX scores (Appendix D3). For every 10 percent increase in the FRAX 10-year probability of MOF as calculated with FN-BMD, the p-values for the interaction of FRAX score on the association of treatment assignment and these fracture outcomes were 0.38, 0.61, 0.42 and 0.88, respectively. These p-values indicate that the associations between treatment and fracture did not differ by baseline FRAX score. Results were similar when FRAX scores for 10-year probability of MOF were modeled in tertiles and as ≥20 percent versus <20 percent.

Bone Turnover Markers

One post hoc analysis of pooled data from all FIT-I and FIT-II participants (medium ROB) evaluated whether pretreatment levels of several bone turnover markers modified the effect of four years of alendronate versus placebo on risk of incident fractures. Participants were stratified as osteoporotic (FN-BMD T-score ≤ -2.5 or prevalent radiographic vertebral fracture) or not osteoporotic (FN-BMD T-score > -2.5 and no prevalent radiographic vertebral fracture).

In osteoporotic women, the effect of alendronate versus placebo on risk of incident nonvertebral fracture was significantly greater in women with higher versus lower pretreatment levels of P1NP (p for interaction 0.02). However, the effect of treatment on risk of incident nonvertebral fracture did not vary significantly as a function of pretreatment levels of BSAP or CTX (p for interaction not significant). Further, there were no significant interactions with any of the three bone turnover markers on the effect of treatment group on risk of incident radiographic vertebral fracture.

In nonosteoporotic women, the association between treatment group assignment and risk of either incident nonvertebral fracture or incident radiographic vertebral fracture did not vary as a function of pretreatment levels of P1NP, BSAP or CTX (interactions not significant).

Other Effect Modifiers Included in Review

No evidence was found for the following effect modifiers: pretreatment age, race, sex, comorbid conditions (DM, CKD, CVD), treatment dose, treatment frequency, treatment duration, or treatment delivery route.
Zoledronic Acid (Zolendronate)

Key Points

Key Question 1
- Evidence was insufficient to draw conclusions about differences in risk of incident fractures between long-term zoledronic acid and placebo or a different active treatment.

Key Question 2
- Evidence was insufficient to draw conclusions about whether differences in risk of incident fractures between long-term zoledronic acid and placebo or a different active treatment vary as a function of patient, bone or drug characteristics.

Eligible Studies
No eligible studies of long-term treatment with zoledronic acid versus inactive control (e.g., placebo, no treatment) or a different active treatment reported on risk of incident fracture or on whether risk of incident fracture between treatments varied as a function of patient, bone, or zoledronic acid or control characteristics. Reports of RCTs that compared zoledronic acid continuation versus discontinuation (placebo drug holiday) and reported on risk of fractures are reviewed in Chapter 6 of this report.
Denosumab

Key Points

Key Question 1

- In postmenopausal women with BMD T score < -1.8, there was no difference between denosumab therapy for 4 years versus placebo in risk for incident clinical fracture (low SOE).
- Evidence was insufficient to draw conclusions about differences between 4 years of denosumab therapy versus placebo on risk of other incident clinical fractures or incident radiographic vertebral fractures.

Key Question 2

- Evidence was insufficient to draw conclusions about whether differences in risk of incident fractures between long-term denosumab and placebo or a different active treatment vary as a function of patient, bone or drug characteristics.

Eligible Studies

One eligible publication of one study compared long-term denosumab treatment versus placebo, no treatment or a different active treatment, and reported on risk of incident fracture (Table 7). Details of this study with medium ROB were extracted in evidence tables, summary SOE assessments (Appendix D), and ROB assessments (Appendix C).

This multisite RCT randomized 365 postmenopausal women with LS-BMD T-score of -1.8 to -4.0 or proximal femur T-score of -1.8 to -3.5 to placebo (n=46) or one of seven denosumab regimens for 2 years. After 2 years, women initially assigned placebo continued placebo for another 2 years. Based on the denosumab regimen to which they’d been randomized, women initially assigned to denosumab were nonrandomly reallocated to denosumab 60 mg every 6 months (n=231) for two years; placebo for 1 year followed by denosumab 60 mg every 6 months for 1 year (n=41); or placebo for 2 years (n=47).

Table 7. Baseline characteristics of the denosumab trial versus placebo

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number randomized</td>
<td>365</td>
</tr>
<tr>
<td>Age, mean years</td>
<td>62.4</td>
</tr>
<tr>
<td>Gender, woman</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white</td>
<td>86%</td>
</tr>
<tr>
<td>Race, Hispanic</td>
<td>11%</td>
</tr>
<tr>
<td>Mean LS-BMD T-score</td>
<td>-2.16</td>
</tr>
<tr>
<td>Mean hip BMD T-score*</td>
<td>-1.42</td>
</tr>
<tr>
<td>History of clinical fracture</td>
<td>NR</td>
</tr>
</tbody>
</table>
Outcomes

Incident Clinical Fractures

In this trial, clinical and osteoporotic fractures were considered neither primary or secondary outcomes, but rather were reported as adverse events. No information was provided about how incident fractures were identified and defined. Further, incidence of fractures was reported collectively for all women initially assigned to denosumab (n=314), with no separate results provided for those assigned to denosumab for four years, those assigned to denosumab for two years followed by placebo for 2 years, or those assigned to denosumab for 2 years followed by placebo for 1 year and then denosumab again for 1 year. Through 4 years, results showed no difference between the groups initially assigned to denosumab versus placebo in risk of any incident clinical fracture (10.5% vs. 10.9%; risk ratio [RR] 0.97 [95% CI 0.40, 2.35]) (low SOE) or incident osteoporotic fracture (any clinical fracture excluding those of the phalanges, face or those caused by severe trauma) (7.0% vs. 8.7%; RR 0.81 [95% CI 0.29, 2.23]).

Incident Radiographic Vertebral Fractures

No information was reported about risk of incident radiographic vertebral fractures.

Change in BMD

The primary study endpoint was mean percentage change from baseline in dual-energy x-ray absorptiometry (DXA) LS-BMD and TH-BMD. BMD results were reported separately for participants assigned to denosumab for four years, denosumab for 2 years followed by placebo for 2 years, denosumab for 2 years followed by placebo for one year followed by denosumab retreatment for one year, and placebo for 4 years. Among the original 277 participants in the continuous denosumab and placebo groups, 66 percent (n=182) completed the 4-year trial. Mean change in LS-BMD from baseline ranged from 9.4 percent to 11.8 percent in the continuous denosumab group compared with -2.4 percent for the placebo group (between-group p<0.001). Mean change in TH-BMD from baseline ranged from 4.0 percent to 6.1 percent in the continuous denosumab group compared with -3.5 percent for the placebo group (between-group p<0.001).

Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

No eligible trials evaluated whether the efficacy of long-term denosumab versus inactive control on risk of incident fracture varied as a function of patient, bone or drug characteristics.
Raloxifene

Key Points

Key Question 1
- In postmenopausal women with osteoporosis, between raloxifene treatment and placebo for 4 years:
  - There was no difference in risk of incident nonvertebral fracture (high SOE) or incident hip fracture (moderate SOE).
  - There was a lower risk of incident clinical vertebral fracture (high SOE) and incident radiographic vertebral fracture (high SOE).
- In postmenopausal women with osteoporosis, between raloxifene treatment and placebo for 8 years:
  - There was no difference in risk of incident nonvertebral fracture (moderate SOE).
  - Evidence was insufficient to draw conclusions about risk of incident hip, incident clinical vertebral or incident radiographic vertebral fractures.

Key Question 2
- There did not appear to be differences in risk of incident fractures between long-term raloxifene and placebo vary as a function of patient, bone, or drug characteristics, though SOE for these associations was not assessed.

Eligible Studies

Five eligible publications of one unique study compared long-term raloxifene treatment versus placebo and reported on risk of incident fractures (Table 8). All studies had low or medium ROB. Appendix D provides detailed evidence tables and SOE assessments. Summary ROB assessments can be found in Appendix C.

All these publications were from the Multiple Outcomes of Raloxifene Evaluation (MORE) trial and its extension, the Continuing Outcomes Relevant to Evista (CORE) study.

MORE was a multinational RCT that compared raloxifene 60 mg/day and 120 mg/day to placebo in 7705 postmenopausal women with osteoporosis (LS-BMD or FN-BMD T-score ≤ -2.5, or “low BMD” with ≥ 1 moderate or ≥ 2 mild baseline radiographic vertebral fractures, or ≥ 2 moderate baseline radiographic vertebral fractures) for 3 years with a 1-year extension phase. During the extension phase, participants remained blinded to their originally assigned treatment, but were allowed to use other bone-active agents as clinically indicated. For the 3-year core treatment phase, the primary outcome was incident radiographic vertebral fracture, defined as a new or worsened vertebral deformity on at least two of three types of independent radiologic assessments during scheduled study radiographs. Secondary endpoints included incident clinical vertebral fracture identified on study radiographs performed to evaluate back pain suggestive of fracture, and incident self-reported nonvertebral fracture confirmed by nonstudy radiographic reports.

CORE was a 4-year follow-up to the MORE RCT conducted at 130 of the 180 MORE study sites. Of the 6511 participants enrolled at those 130 MORE sites, 4011 chose to enroll in CORE. This nonrandomized CCT assigned individuals who had been randomized to either
raloxifene 60 mg/day or 120 mg/day in MORE to take raloxifene 60 mg/day in CORE. Women who had been randomized to placebo in MORE were assigned to continue placebo in CORE. Most CORE participants had an approximately 1-year interval between the end of the MORE extension and the beginning of CORE, during which they received no study drug. As during the MORE extension phase, CORE participants were allowed to take other bone-active agents as clinically indicated. All 4011 CORE enrollees were followed according to their MORE treatment assignment (raloxifene or placebo), including the 20 percent who did not take study medication during CORE, either by choice or due to contraindication. CORE participants and investigators remained blinded to MORE and CORE treatment assignment. Incident nonvertebral fracture was a secondary endpoint in CORE, and was self-reported by participants and confirmed by radiologic report.

Baseline characteristics for the MORE and CORE populations are described in detail in Appendix D2. The four reports from MORE had low ROB24, 34, 35, 50, and the single report from CORE had medium ROB for fracture outcomes but high ROB for BMD outcomes (available in <10% of CORE population).25

### Table 8. Baseline Participant Characteristics from the MORE trial and CORE extension

<table>
<thead>
<tr>
<th>Characteristic, mean or %</th>
<th>MORE</th>
<th>CORE (Characteristics available only from MORE baseline)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number enrolled and randomized</td>
<td>7705</td>
<td>4011</td>
</tr>
<tr>
<td>Age, mean</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Years post-menopausal</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Gender, woman</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>Mean FN-BMD T-score</td>
<td>-2.33</td>
<td>-2.32</td>
</tr>
<tr>
<td>Mean LS-BMD T-score</td>
<td>-2.57</td>
<td>-2.59</td>
</tr>
<tr>
<td>History of clinical fracture</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Prevalent radiographic vertebral fracture</td>
<td>37%</td>
<td>36%</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>5%</td>
<td>3%</td>
</tr>
</tbody>
</table>

**Abbreviations**: BMD=bone mineral density; CORE= Continuing Outcomes Relevant to Evista; FN=femoral neck; LS=lumbar spine; MORE=Multiple Outcomes of Raloxifene Evaluation; NR=not reported

### Outcomes

#### Incident Clinical Fractures

In analyses in which results for women allocated to raloxifene 60 mg/day and 120 mg/day were pooled and compared with placebo through 4 years of follow-up, there was no difference between treatment groups in risk of incident nonvertebral fracture (12% vs. 11%; RR 0.93 [95% CI 0.81, 1.06]) (high SOE) or incident hip fracture (1% vs. 1%; RR 0.97 [95% CI 0.62, 1.52])
However, raloxifene 60 mg/day reduced the four-year risk of incident clinical vertebral fracture versus placebo (2% vs. 4%; RR 0.58 [95% CI 0.43, 0.79]) (high SOE). Through 8 years of follow-up, there was no difference between raloxifene and placebo in incidence of any nonvertebral fracture (HR 1.00 [95% CI 0.82, 1.21]) (moderate SOE) or in the incidence of a pooled subset of nonvertebral fractures (clavicle, humerus, wrist, pelvis, hip, or lower leg) (HR 1.01 [95% CI 0.81, 1.26]).

Incident Radiographic Vertebral Fractures
Through 4 years of follow-up, raloxifene reduced the risk of incident radiographic vertebral fracture versus placebo at both 60 mg/day (8% vs. 12% as estimated from figure 2; RR 0.64 [95% CI 0.53, 0.76]) and 120 mg/day doses (7% vs. 12%; RR 0.57 [95% CI 0.48, 0.69]) (high SOE). No data were reported for incident radiographic vertebral fractures through eight years.

Change in BMD
Changes in LS-BMD and FN-BMD through four years were significantly greater compared with placebo with both raloxifene 60 mg/day (2.6% for LS-BMD and 2.1% for FN-BMD) and raloxifene 120 mg/day (2.5% for LS-BMD and 2.3% for FN-BMD) (p<0.001 for all comparisons). No data were reported for BMD changes through 8 years.

Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics
We looked at whether the efficacy of long-term raloxifene therapy versus inactive control on risk of incident fracture varies as a function of patient, bone, and/or drug characteristics. Eligible studies included evidence for the following possible effect modifiers.

Age
In the CORE extension population, the relative risk for incident nonvertebral fracture between the raloxifene 60 mg/day and placebo groups through 8 years did not differ as a function of participant age at MORE baseline (p for interaction ≥ 0.10).

Baseline BMD
In the CORE extension population, the relative risk for incident nonvertebral fracture between the raloxifene 60 mg/day and placebo groups through 8 years did not differ as a function of participant BMD (site not specified) at MORE baseline (p for interaction ≥ 0.10).

Prior Fracture
In the MORE population, risks of incident radiographic vertebral fracture and clinical vertebral fracture during 4-year follow-up for raloxifene 60 mg/day or raloxifene 120 mg/day compared with placebo did not differ as a function of baseline radiographic vertebral fracture status. During 8-year follow-up from the start of MORE to the end of CORE, the effect of raloxifene versus placebo on risk of incident nonvertebral fracture risk was insignificant overall, insignificant within women who had a baseline radiographic vertebral fracture, and
insignificant in women who did not have a baseline radiographic vertebral fracture. Nevertheless, the p-value for interaction by baseline radiographic vertebral fracture status was reported at <0.10.\textsuperscript{25}

**Other Effect Modifiers Included in Review**

No evidence was found for the following effect modifiers: race, sex, comorbid conditions (DM, CKD, CVD), osteoporosis status, calculated pre-treatment fracture risk (e.g., FRAX\textsuperscript{®}), biochemical markers (CTX, NTX, P1NP, BSAP), treatment dose, treatment frequency, treatment duration, or treatment delivery route.
Hormone Therapy

Key Points

- In postmenopausal women with past clinical fracture, but unknown baseline BMD or vertebral fracture status, and with prior hysterectomy, there was low strength evidence that compared with placebo, 7 years of unopposed estrogen lowers risk of both incident clinical fractures and incident hip fractures.
- In postmenopausal women with past clinical fracture, but unknown baseline BMD or vertebral fracture status, and with an intact uterus, there was low strength evidence that compared with placebo, 5.6 years of estrogen/progestin lowers risk of incident clinical fractures but not that of incident hip fractures.
- In postmenopausal women with osteoporosis, evidence was insufficient to draw conclusions about differences between 4 years of estrogen/progestin therapy versus nonplacebo control and risk of incident nonvertebral fractures or incident radiographic vertebral fractures.

Eligible Studies

We identified 3 eligible RCTs that compared hormone therapy versus placebo\textsuperscript{28, 29} or nonplacebo control\textsuperscript{27} and reported long-term results for incident fracture (Table 9). Appendix D provides detailed evidence tables and SOE assessments. Summary ROB assessments can be found in Appendix C.

Two of these trials, from the U.S.-based Women’s Health Initiative (WHI), had low ROB and enrolled participants without regard to whether they had osteoporosis or osteopenia. The first WHI trial randomized 16,608 postmenopausal women with an intact uterus to an estrogen/progestin combination (conjugated equine estrogen 0.625 mg/day plus medroxyprogesterone 2.5 mg/day) versus placebo for a mean treatment duration of 5.6 years.\textsuperscript{29} Thirty-nine percent of women reported a history of past clinical fracture. Baseline vertebral radiographs were not collected and baseline DXA BMD was measured in 1024 women (participants at 3 of 40 clinical sites). The second WHI trial randomized 10,739 postmenopausal women with a history of hysterectomy to unopposed estrogen (conjugated equine estrogen 0.625 mg/day) versus placebo for a mean treatment duration of 7.1 years. Thirty-nine percent of women reported a history of past clinical fracture. Baseline vertebral radiographs were not collected and baseline DXA BMD was measured in 938 women (participants at 3 of 40 clinical sites). Participants in both WHI trials were allowed to co-enroll in an RCT of calcium plus vitamin D supplementation versus placebo. Fracture efficacy results for these two trials were analyzed only for individuals who reported a past clinical fracture and in those with measured baseline BMD. A third, unrelated trial, which had medium ROB, enrolled 72 postmenopausal women with osteoporosis, defined as 1 to 4 baseline radiographic vertebral fractures and DXA LS-BMD T-score $\leq -2.0$. Of the 72 women enrolled in this study, 36 were randomized to etidronate alone or to a combination of etidronate plus estrogen/progestin. Because etidronate is not U.S. Food and Drug Administration (FDA) approved, these treatment groups were not analyzed. The remaining 36 women were randomized to an estrogen/progestin combination (conjugated estrogen 0.625 mg/day plus norgestrel 150 mg for 12 days each month) versus
nonplacebo control for 4 years. All participants received daily supplemental calcium 1000 mg and vitamin D 400 units. The country where the trial was performed was not identified.

Table 9. Baseline Characteristics of hormone therapy versus placebo/control RCTs

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Cauley 200329 WHI</th>
<th>Jackson 200628 WHI</th>
<th>Wimalawansa 199827</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment intervention</td>
<td>Estrogen + Progestin vs. PBO</td>
<td>Estrogen vs. PBO</td>
<td>Estrogen + Progestin vs. non-PBO control</td>
</tr>
<tr>
<td>Treatment duration</td>
<td>5.6 yr</td>
<td>7.1 yr</td>
<td>4 yr</td>
</tr>
<tr>
<td>Number randomized</td>
<td>16,608 (BMD measured in 1024)</td>
<td>10,739 (BMD measured in 938)</td>
<td>36*</td>
</tr>
<tr>
<td>Age, mean years</td>
<td>63</td>
<td>64</td>
<td>65</td>
</tr>
<tr>
<td>Gender, woman</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white</td>
<td>84%</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>Time since menopause</td>
<td>10-19 yr = 36% 10-19 yr = 40% 10-19 yr = 32% 20 yr = 24% 20 yr = 50%</td>
<td>Mean 19 yr 10-19 yr = 18% 20 yr = 50%</td>
<td>Median 15 yr</td>
</tr>
<tr>
<td>Mean LS-BMD T-score</td>
<td>-1.3</td>
<td>-1.2</td>
<td>-2.1</td>
</tr>
<tr>
<td>Mean FN BMD T-score</td>
<td>NR</td>
<td>NR</td>
<td>-1.6</td>
</tr>
<tr>
<td>Mean TH BMD T-score</td>
<td>-0.9</td>
<td>-0.8</td>
<td>NR</td>
</tr>
<tr>
<td>History of clinical fracture</td>
<td>39%</td>
<td>39%</td>
<td>NR</td>
</tr>
<tr>
<td>Number of previous RVF per participant</td>
<td>NR</td>
<td>NR</td>
<td>2.3</td>
</tr>
</tbody>
</table>

Abbreviations: BMD=bone mineral density; CEE=conjugated equine estrogen; FN=femoral neck; LS=lumbar spine; NR=not reported; PBO=placebo; RCT=randomized controlled trial; RVF=radiographic vertebral fractures; TH=total hip; WHI=Women’s Health Initiative

*Excludes the 36 women randomized to etidronate or to combined estrogen/progestin and etidronate.

Outcomes

Incident Clinical Fractures

In the first WHI trial, among women with a history of prior clinical fracture, those randomized to estrogen/progestin versus placebo for 5.6 years had a lower risk of incident clinical fracture (11% vs. 14%; HR=0.78 [95% CI 0.68, 0.91]) (low SOE), but no statistically significant difference in risk of incident hip fracture (1.1% vs. 1.4%; HR=0.77 [95% CI 0.48, 1.22]) (low SOE).29 In women with osteoporosis by baseline DXA BMD T-score <-2.5 at any hip or vertebral site (sample size not reported), evidence was insufficient to draw conclusions about a difference in risk of any incident clinical fracture between the estrogen/progestin and placebo groups (1.7% vs. 1.9%; HR=0.87 [95% CI 0.57, 1.34]). In the second WHI trial, among women with a history of prior clinical fracture, those randomized to estrogen versus placebo for 7.1 years had a lower risk of both incident clinical fracture (14% vs. 19%; HR=0.73 [95% CI 0.62, 0.86]) (low SOE) and incident hip fracture (1.0% vs. 1.9%; HR=0.55 [95% CI 0.32, 0.94]) (low SOE).28 In women with baseline DXA BMD T-score <-2.5 (i.e., osteoporosis; n=53) or between -1.0 and -2.5 (i.e., osteopenia; n=363) at any hip or vertebral site, evidence was insufficient to draw conclusions about differences in risk of any incident clinical fracture between the estrogen and placebo groups (HR=0.83 [95% CI 0.17, 3.91] for osteoporosis group, and HR=0.83 [95% CI 0.49, 1.40] for osteopenia group, respectively). In both these studies, analyses conducted in the entire study populations, regardless of past fracture history or baseline...
BMD, showed that compared to placebo, estrogen/progestin and estrogen both significantly reduced the risk of incident clinical fractures, incident hip fractures, and incident clinical vertebral fractures. In the small non-WHI trial, incident fractures were reported as secondary endpoints. Over its 4-year trial period, one participant each in the estrogen/progestin and control groups had an incident nonvertebral fracture (7% vs. 7%; RR=0.93 [95% CI 0.06, 13.5]) (insufficient SOE).

**Incident Radiographic Vertebral Fractures**
Results for incident radiographic vertebral fractures were not reported in the WHI trials. In the non-WHI trial, incident radiographic vertebral fractures occurred in 13 percent of women randomized to estrogen/progestin versus 36 percent in the control group (RR=0.37 [95% CI 0.09, 1.62]) (insufficient SOE).

**Change in BMD**
In both WHI trials, change in BMD was a secondary endpoint. Results for change in BMD between baseline and follow-up were reported together for all women with baseline BMD, but not separately by baseline BMD category (e.g., osteoporosis or osteopenia). Three-year BMD follow-up was 82-83% in one of these trials and was not reported in the other trial. Compared to the placebo groups, individuals in both studies assigned to the hormone groups experienced an approximately 3.5% greater increase in total hip BMD and 4-5% greater increase in lumbar spine BMD between baseline and follow-up at 3 years. In the small, non-WHI trial, mean percentage change from baseline in DXA LS-BMD and TH-BMD were the primary study endpoints. Among the original 36 participants randomized, 81 percent (n=29) completed the 4-year trial. Participants randomized to estrogen/progestin had a statistically significantly larger increase in LS-BMD during the 4-year follow-up period compared with the control, 7 percent vs. -2.5 percent respectively (mean difference [MD] 9.5% [95% CI 7.0, 12.0]). A significant difference between groups also was observed in TH-BMD, favoring the estrogen/progestin group (4.8% vs. -4.4%; MD 9.2% [95% CI 6.8, 11.6]).

**Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis Drug Characteristics**
No eligible trials evaluated whether the effect of any estrogen or estrogen/progestin preparation versus placebo or inactive control on risk of incident fracture varied as a function of patient, bone or drug characteristics.
Other Drugs Investigated for Review

Ibandronate

No eligible studies of long-term treatment with ibandronate versus inactive control (e.g., placebo, no treatment) or a different active treatment reported on risk of incident fracture or on whether risk of incident fracture between treatments varied as a function of patient, bone, or ibandronate or control characteristics.

Risedronate

Two eligible studies of long-term treatment with risedronate versus inactive control (e.g., placebo, no treatment) or a different active treatment reported on risk of incident fracture. These studies were rated high ROB and only limited data was extracted (Appendix D7). We identified no eligible studies that compared risedronate with control, and assessed whether risk of incident fracture between treatments varied as a function of patient, bone, or risedronate or control characteristics.
Chapter 5. Harms of Long-term Drug Therapy

Chapter 5 reviews the evidence from eligible studies that addresses Key Questions 3 and 4. Key Question 3 addresses the harms of long-term (>3 years) osteoporosis drug treatment versus placebo, no treatment or a different active treatment. Key Question 4 addresses whether risk of harms varies by patient, bone, or drug characteristics.

We report the literature on the effect of continuing osteoporosis drug therapy versus discontinuing it (placebo drug holiday), including on the effect of long-term versus shorter term treatment, on risk of harms under Key Question 7 in chapter 7 of this report.

This chapter is organized by drug, with Key Questions 3 and 4 discussed sequentially within each drug subsection. Drug versus placebo or no treatment comparisons are listed first, followed by any comparisons of drug versus another active drug treatment. We provide an overview at the start of specific drug subsections when the literature base was complex. Summary statements about the included studies are below; individual study details are provided in Appendix D.

Alendronate

Key Points

Key Question 3
- In postmenopausal women with FN-BMD T-score ≤-1.6 and without baseline vertebral fracture, between alendronate and placebo for 4 years:
  o There was no difference in risk of upper GI events, mortality (low SOE), or hospitalization due to adverse events (low SOE).
- In postmenopausal women with FN-BMD T-score ≤-1.6, between alendronate and placebo for 3 to 4.5 years:
  o Evidence was insufficient to draw conclusions about differences in risk of subtrochanteric or femoral shaft fractures without confirmed radiologic AFF features.
- Between alendronate and no osteoporosis medication use for up to 11 years:
  o Risk of subtrochanteric or femoral shaft fracture without radiologic confirmation of AFF features appeared higher with alendronate versus no osteoporosis drug prescription in one cohort study (low SOE), but evidence was insufficient to draw conclusions about differences in risk between alendronate and either placebo or no treatment in two other studies.
  o Risk of ONJ appeared higher with alendronate versus no osteoporosis drug prescription in one cohort study (low SOE).
  o There appeared to be no difference in risk of atrial fibrillation, though SOE was not graded.
- Evidence was insufficient to draw conclusions about differences in risk of ONJ with alendronate versus raloxifene or calcitonin.

Key Question 4
- In postmenopausal women with FN-BMD T-score ≤ -1.6, between alendronate and placebo for 3.2 to 4 years:
Risk of upper gastrointestinal (GI) adverse events did not appear to differ as a function of age, prior history of upper GI tract disease, renal function or use of nonsteroidal anti-inflammatory drugs (NSAIDs).

- Risk of coronary heart disease, cerebrovascular disease, cancer and mortality did not appear to differ as a function of renal function.
- We found no evidence about whether risk of AFF, subtrochanteric or femoral shaft fracture without confirmed radiologic AFF features, ONJ, or atrial fibrillation between alendronate and control varies as a function of patient, bone or osteoporosis drug treatment factors.

## Eligible Studies

Fourteen eligible publications of eleven unique studies compared long-term alendronate treatment versus placebo, no treatment or a different active treatment, and reported on risk of harms. We rated four of these studies as having high ROB⁸⁷, ⁹⁰-⁹² and extracted only limited data (Appendix D7). For the remaining ten studies with low or medium ROB, there were four eligible publications from one unique RCT with low²³, ³⁷ and medium⁵³, ⁵⁴ ROB, respectively, and six observational studies with medium ROB.³⁸, ³⁹, ⁴⁶-⁴⁸ We extracted additional information from these studies in evidence tables, SOE summary tables (Appendix D) and ROB summary tables (Appendix C).²³

The eligible RCT reports were based on the Fracture Intervention Trial (FIT), which excluded women with recent or recurrent GI ulcers, significant upper GI tract bleeding in the past five years, esophageal or gastric varices, or use of daily medication for dyspepsia. In the FIT-II RCT, risks of mortality, any adverse event leading to hospitalization, and upper GI events (included abdominal pain, esophagitis, esophageal ulcer, acid regurgitation/reflux, and other) were compared between women randomized for 4 years to alendronate versus placebo.²³ In the second FIT report, a pooled analysis of FIT-I and FIT-II participants, risk was compared between alendronate and placebo for a mean follow-up of 3.8 years for upper GI tract adverse events, diagnoses and symptoms.⁵³ The third FIT report, pooled participants from FIT-I and FIT-II and estimated risk of incident subtrochanteric or femoral shaft fracture.³⁷ Self-reported fractures were centrally confirmed by review of community radiographic reports. For reports noting subtrochanteric or femoral shaft fractures, outside radiographs were sought but rarely available. Descriptions in the radiographic reports of atypical features were recorded. Pathologic fractures, periprosthetic fractures, and fractures due to excess trauma were excluded. Another pooled analysis of FIT-I and FIT-II participants reported on risk of GI events, coronary heart disease, cerebrovascular disease, cancer and mortality as a function of baseline renal function.⁵⁴ None of the reports provided information about the risk of ONJ or atrial fibrillation.

Four publications used Danish national data to examine risk of subtrochanteric and femoral shaft fracture,³⁸, ³⁹ inflammatory jaw events possibly attributable to ONJ,⁴⁵ or atrial fibrillation⁴⁸ with long-term alendronate. In the earliest study, among men and women ≥51 years of age who experienced a hospital treated nonhip fracture, 178 individuals treated and adherent (i.e., medication possession ratio >80%) with alendronate for >6 years were matched 2:1 with age, sex and same index fracture site controls who were not exposed to bisphosphonate treatment.³⁸ Subtrochanteric and femoral shaft fractures were identified by International Statistical Classification of Diseases and Related Health Problems (ICD)-10 diagnosis codes, without radiologic review for AFF features or exclusions for excess trauma or cancer.
The other three Danish studies evaluated 55,090 patients who received an alendronate prescription during an 11-year period. For each exposed participant, three age and gender matched controls were randomly selected from the entire Danish population who had not been prescribed an osteoporosis medication. Mean treatment follow-up was 3.8 years for all osteoporosis medications collectively. Though time on treatment was not reported for alendronate specifically, 53 percent of treated patients were prescribed alendronate (another 38% were prescribed etidronate). One reported on risk of any inflammatory jaw event as defined by diagnostic codes that included osteomyelitis, osteitis, osteoradionecrosis, periostitis and sequestrum, without a requirement that coding exceeded one encounter, and without review of radiology and pathology records.45 In the second of these reports, the primary outcome was subtrochanteric or femoral shaft fracture as defined by ICD8 and ICD10 diagnostic codes, without review of radiology records and without exclusions for trauma or cancer.39 For the third report, the primary outcome was atrial fibrillation or flutter defined as a recorded incident leading to hospitalization or an outpatient contact.48 Participant mean age was 71 years and no information was reported about baseline BMD or prior fracture history.

One retrospective cohort study conducted in Taiwan reported risk of ONJ in 6485 osteoporotic women treated with alendronate versus 1869 using raloxifene.47 ONJ was defined as exposed maxillofacial bone for more than 8 weeks meeting the following criteria: qualifying ICD-9 diagnostic code, at least one dentistry visit claim, radiographic documentation of ill-defined lytic lesions of the jawbone, and pathology records documenting sequestra or osteomyelitis. Participants were excluded for having head and neck cancer before or during the alendronate treatment period, a history of radiotherapy to the jaws, or previously receiving other antiresorptive treatment. Alendronate treatment duration ranged up to 11 years, but mean duration was about 1.8 years, including a mean duration of about 4 years in those diagnosed with ONJ. Women were 50 years of age or older (mean 70 years). Sixteen percent of participants in both treatment groups had diabetes. Women treated with alendronate more often were chronic users of glucocorticoids. The study did not report baseline BMD or prior fracture history.

One retrospective cohort study conducted in Taiwan reported risk of ONJ for alendronate versus raloxifene or calcitonin in men and women who had been started on osteoporosis medication following a recent vertebral or hip fracture.46 ONJ was defined as meeting the following criteria: qualifying ICD-9 diagnostic code for at least three consecutive visits over at least 8 weeks, and documentation of receipt of appropriate broad-spectrum oral antibiotics. Participants were excluded for a history of radiation to the jaws. Radiology and pathology records were not reviewed. Participants were 50 years of age or older (mean 74 years). About 24 percent had diabetes. In the original cohort (n=43,645), the alendronate treated group differed from the raloxifene/calcitonin treated group in less often being female (79% vs. 87%), and more often having had a hip fracture as their pre-treatment fracture (29% vs. 15%) and having baseline gingival or periodontal disease (27% vs. 23%). The study did not report baseline BMD. Baseline prognostic variables were similar between treatment groups in the propensity-matched cohort (n=32,006).

Baseline characteristics for both observational studies are reported in Appendix D1.
Outcomes

Serious Adverse Events
In the FIT-II RCT, although serious adverse events as defined by the U.S. FDA were not reported, there was no difference between women assigned to long-term alendronate versus placebo in risk of mortality (1.7% vs. 1.8%; HR 0.92 [95% CI 0.59, 1.45]) (low SOE) or adverse events leading to hospitalization (29% vs. 27%; RH 1.09 [95% CI 0.98, 1.22]) (low SOE).23

Upper GI Tract Adverse Events
In the FIT-II RCT, alendronate for 4.2 years versus placebo did not significantly increase risk of any upper GI tract event (47.2% vs. 47.5%; RH 1.00 [95% CI 0.92, 1.09]), or of abdominal pain, esophageal ulcer, or acid regurgitation/reflux (Appendix D4).23 Results were not statistically significantly different between treatment groups for risk of esophagitis (0.9% vs. 0.5%; RH 1.90 [95% CI 0.90, 4.26]).

In a pooled analysis of FIT-I and FIT-II data, alendronate was not associated with an increased risk of any upper GI tract adverse event (47.5% vs. 46.2%; RR 1.02 [95% CI 0.95, 1.10]).53 There also was no between group difference in risk of any individual upper GI tract adverse event, including dyspepsia (18.2% vs. 19.1%; RR 0.94 [95% CI 0.84, 1.05]), abdominal pain, nausea, vomiting, or serious upper GI tract adverse event (RR 1.10 [95% CI 0.77, 1.56]). (Appendix D4).

Subtrochanteric and Femoral Shaft Fractures
No eligible long-term alendronate study reported information about risk of radiologically confirmed AFF. Three eligible studies provided conflicting evidence about the association of long-term alendronate use with risk of incident subtrochanteric or femoral shaft fracture.37-39 In a secondary analysis of pooled FIT-I and FIT-II RCT data, one woman each in the alendronate and placebo groups had a subtrochanteric or femoral shaft fracture (0.031% vs. 0.031%; HR 1.03 [95% CI 0.06, 16.46] (insufficient SOE).37 One retrospective cohort study reported five cases of subtrochanteric or femoral shaft fracture in the combined groups of those highly adherent with alendronate >6 years (n=178) and a no treatment group (n=356). (Abrahamsen 2009) They detected no significant difference in risk (HR 1.37 [95% CI 0.22, 8.62]) (insufficient SOE).38 A second retrospective cohort study reported that alendronate prescription for a mean of approximately 3.8 years compared with no osteoporosis drug prescription was associated with an increased risk of subsequent subtrochanteric fracture (0.017% vs. 0.006%; HR 2.41 [95% CI 1.78, 3.27] (low SOE) and femoral shaft fracture (0.012% vs. 0.003%; HR 2.90 [95% CI 1.97, 4.26]) (low SOE).39

ONJ
Three eligible studies provided conflicting results about the risk of ONJ associated with long-term alendronate use.45-47 All were retrospective cohort studies. The first study reported an increased risk of any jaw event after approximately 3.8 years of follow-up after alendronate exposure versus no osteoporosis drug prescription (0.03% vs. 0.01%; HR 3.15 [95% CI 1.44, 6.87]) (low SOE).45 The other two studies reported higher risk of ONJ with alendronate versus raloxifene (HR 7.42; 95% CI 1.02, 54.09) (insufficient SOE)47 and no difference in risk of ONJ
between alendronate and raloxifene or calcitonin (0.15% vs. 0.08%; HR 0.86 [95% CI 0.44, 1.69] in the propensity-matched cohort) (insufficient SOE)\textsuperscript{46}, respectively.

**Atrial Fibrillation**

One eligible study reported that alendronate prescription for a mean of approximately 3.8 years compared with no osteoporosis drug prescription was not associated with risk of subsequent atrial fibrillation or flutter (1.3% vs. 1.0%; HR 1.04 [95% CI 0.98, 1.10]).\textsuperscript{48}
Table 10. Estimated risk of Subtrochanteric/femoral shaft (ST/FS) fracture*, ONJ or Atrial Fibrillation with Alendronate Use from Controlled Observational Studies

<table>
<thead>
<tr>
<th>Author, Year Country</th>
<th>Study design</th>
<th>AFF, ONJ or atrial fibrillation case definition</th>
<th>Alendronate treatment duration</th>
<th>Treatment control group</th>
<th>Results (95% CI) Model Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subtrochanteric/femoral shaft (ST/FS) fracture</strong></td>
<td>Retrospective cohort n=5 N=534</td>
<td>ST/FS cases: Subtrochanteric and femoral shaft fractures defined by ICD10 diagnostic codes. No radiology records were reviewed and there were no exclusions for trauma or cancer.</td>
<td>&gt;6 years (mean not reported)</td>
<td>No prescription for bisphosphonate</td>
<td>HR=1.37 (0.22, 8.62) Age, sex, number of comediations, oral glucocorticoid use, Charlson comorbidity index.</td>
</tr>
<tr>
<td>Abrahamsen 2009&lt;sup&gt;18&lt;/sup&gt; Denmark Medium</td>
<td>Retrospective cohort n=309 N=220,360</td>
<td>ST/FS cases: Subtrochanteric and femoral shaft fractures defined by ICD8 and ICD10 diagnostic codes. No radiology records were reviewed and there were no exclusions for trauma or cancer.</td>
<td>Range 0-11 years. Mean duration ~3.8 years.</td>
<td>No prescription for osteoporosis drug</td>
<td>Subtrochanteric: 0.017% vs. 0.006%; HR 2.41 (1.78, 3.27) Femoral shaft: 0.012% vs. 0.003%; HR 2.90 (1.97, 4.26) History of past fracture, systemic hormone use, systemic corticosteroid use, alcoholism</td>
</tr>
<tr>
<td>Vestergaard 2011&lt;sup&gt;19&lt;/sup&gt; Denmark Medium</td>
<td>Retrospective cohort n=40 N=8354</td>
<td>ONJ case: ICD9 diagnosis codes for inflammatory dental condition, claim for ≥1 dentistry visit, radiographic confirmation, pathologic confirmation.</td>
<td>Range 0-11 years. Mean duration ~1.8 years (~4.0 years in ONJ group)</td>
<td>Raloxifene</td>
<td>HR 7.42 (1.02, 54.09) Reported “adjusting for other possible contributions from other variables,” but specific covariates included in the model is unclear.</td>
</tr>
<tr>
<td>Chiu 2014&lt;sup&gt;27&lt;/sup&gt; Taiwan Medium</td>
<td>Retrospective cohort n=40 N=8354</td>
<td>ONJ case: ICD9 diagnosis codes for inflammatory dental condition, claim for ≥1 dentistry visit, radiographic confirmation, pathologic confirmation.</td>
<td>Range 0-11 years. Mean duration ~1.8 years (~4.0 years in ONJ group)</td>
<td>Raloxifene</td>
<td>HR 7.42 (1.02, 54.09) Reported “adjusting for other possible contributions from other variables,” but specific covariates included in the model is unclear.</td>
</tr>
<tr>
<td>Lin 2014&lt;sup&gt;46&lt;/sup&gt; Taiwan Medium</td>
<td>Retrospective cohort</td>
<td>ONJ case: ICD9 diagnostic codes from at least 3 consecutive visits over at least 8 weeks, and receipt</td>
<td>Up to 6 years (mean and</td>
<td>Non- bisphosphonate osteoporosis</td>
<td>Unmatched cohort: 0.14% vs. 0.08%; HR 0.87 (0.47, 1.58) Propensity matched cohort: 0.15% vs. 0.08%; HR 0.86 (0.44, 1.69)</td>
</tr>
<tr>
<td>Unmatched cohort: n=46; N=43,645</td>
<td>of appropriate broad-spectrum antibiotics.</td>
<td>median not reported</td>
<td>medication (e.g. raloxifene or calcitonin)</td>
<td>Age, gender, year, fracture type, fracture history, fall history, comorbidities (diabetes, hyperlipidemia, pancreatitis, gingival and periodontal diseases, other diseases and conditions of teeth and supporting structures, dentoalveolar surgery, rheumatoid arthritis, systemic lupus erythematosus, renal disease, hypertension, Alzheimer’s disease), and comedications (antiepileptics, β-blockers, benzodiazepines, glucocorticoids, hormone therapy, COX-2 agents, SSRI, thyroid drugs, and sleep/hypnotic agents).</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Propensity matched cohort: n=37; N=32,006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Atrial fibrillation**

<table>
<thead>
<tr>
<th>Vestergaard 2010&lt;sup&gt;48&lt;/sup&gt;</th>
<th>Retrospective cohort n=2364 N=220,360</th>
<th>Atrial fibrillation case: recorded incident leading to hospitalization or an outpatient contact</th>
<th>Range 0-11 years. Mean duration ~3.8 years.</th>
<th>No prescription for osteoporosis drug</th>
<th>1.3% vs. 1.0%; HR 1.04 (0.98, 1.10) Prior atrial fibrillation, heart valve disease, heart failure, hyperthyroidism, diuretic use, other cardiovascular drug use, COPD, COPD medication use, alcoholism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Denmark Medium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; CI=confidence interval; COPD=chronic obstructive pulmonary disease; COX=cyclooxygenase; HR=adjusted hazard ratio; ICD=International Classification of Diseases; N=number; ONJ=osteonecrosis of the jaw; SSRI=selective serotonin reuptake inhibitors.

*No studies reported results for atypical femoral fractures meeting radiographic criteria for ASBMR*
Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We looked at whether the effect of long-term alendronate therapy versus inactive control on risk of harms varies as a function of patient, bone, and/or drug characteristics. Two eligible studies with medium risk of bias included evidence for the following possible effect modifiers of long-term alendronate treatment harms:53, 54

Upper Gastrointestinal Tract Adverse Events

A post hoc pooled analysis of FIT-I and FIT-II data53 reported that the risk of alendronate versus placebo on upper GI adverse events did not differ as a function of age, current use of NSAIDs, or history of an upper GI tract disease that was not an exclusion criterion for study participation (e.g., peptic ulcer, reflux esophagitis). No data were provided to support these statements for most upper GI tract adverse events. However, in a manuscript figure, unadjusted rates for esophageal adverse events and gastroduodenal perforations, ulcerations and bleeding appeared visually similar across age categories (55-65, 66-70, 71-75 and 76-85 years). Also, unadjusted RR between alendronate and placebo groups for these outcomes were similar and 95 percent confidence intervals were widely overlapping between strata defined based on prior history of a nonexclusionary upper GI tract disease and current NSAID use (Appendix D4). A second post hoc pooled analysis of FIT-I and FIT-II data reported that renal function (estimated glomerular filtration rate [eGFR] <45 ml/min vs. ≥45 ml/min based on the Cockcroft Gault formula) did not modify the effect of alendronate versus placebo on risk of GI events, coronary heart disease, cerebrovascular disease, cancer or mortality.54

Other Effect Modifiers Included in Review

No evidence was found for the following effect modifiers: pretreatment age, race, sex, comorbid conditions (DM, CKD, CVD), osteoporosis status, fracture history, calculated pre-treatment fracture risk (e.g., FRAX®), pretreatment and early treatment (e.g. 1 year) imaging (L-spine, total hip & femoral neck DXA BMD), biochemical markers (CTX, NTX, P1NP, BSAP), treatment dose, treatment frequency, treatment duration, or treatment delivery route.
Zoledronic Acid (Zolendronate)

Key Points

Key Question 3
- Evidence was insufficient to draw conclusions about differences in risk of harms between long-term zoledronic acid and placebo or a different active treatment.

Key Question 4
- Evidence was insufficient to draw conclusions about whether differences in risk of harms between long-term zoledronic acid and placebo or a different active treatment vary as a function of patient, bone or drug characteristics.

Eligible Studies
No eligible studies reported on risk of harms of long-term treatment with zoledronic acid versus placebo or another active treatment, or on whether any differences in risk vary as a function of patient, bone, or osteoporosis drug characteristics. Reports of eligible studies that compared zoledronic acid continuation versus discontinuation (placebo drug holiday) and reported on risk of harms are reviewed in Chapter 7 of this report.
Any Bisphosphonate

Key Points

Key Question 3

- There was higher risk of atypical femoral fracture (AFF) with confirmed radiologic features with long-term use of any bisphosphonate versus no bisphosphonate use (low SOE) or past bisphosphonate use (low SOE).
- There was higher risk of subtrochanteric/femoral shaft fractures without radiologic AFF features with long-term use of any bisphosphonate versus no bisphosphonate use (low SOE).
- Evidence was insufficient to draw conclusions about differences in risk of subtrochanteric or femoral shaft fractures with long-term use of any bisphosphonate versus raloxifene or calcitonin use.
- Evidence was insufficient to draw conclusions about differences in risk of ONJ with long-term use of any bisphosphonate versus control.

Key Question 4

- Risks of radiologically confirmed AFF and of subtrochanteric or femoral shaft fracture without radiologic confirmation of AFF features appeared greater with increased duration of long-term bisphosphonate use versus control, though SOE for these association were not graded.

Eligible Studies

Six eligible publications of six unique studies, all observational, compared long-term treatment with any bisphosphonate (reported results for bisphosphonates collectively) versus no bisphosphonate treatment, past bisphosphonate treatment or active control and reported on risk of harms. We assessed one study as high ROB and extracted limited data (Appendix D7). The remaining five studies had medium ROB. Appendix D provides detailed evidence tables and summary SOE assessments; ROB assessments are provided in Appendix C.

Two studies compared the effect of long-term use of any bisphosphonate versus no osteoporosis drug use on risk of harms. The first of these, conducted in Sweden, performed both retrospective cohort and case-control analyses. Participants were men or women aged > 55 years, had a diagnosis code for either subtrochanteric or femoral shaft fracture during a three-year period and no associated excessive trauma or pathological fracture. Participants had radiographs reviewed and those whose met American Society for Bone and Mineral Research (ASBMR) criteria, were categorized as having AFF. Those whose subtrochanteric or femoral shaft fractures were determined to not have AFF radiologic features were categorized as non-AFF controls. In the cohort analysis, the age and sex-stratified risk of AFF was estimated for participants who received bisphosphonates (per the Swedish Prescription Register) compared with that in those who did not. Osteoporosis status was not reported. In the case-control analysis, the multivariable-adjusted odds of AFF versus non-AFF were compared between individuals with long-term bisphosphonate use and those with no osteoporosis drug use. The second of these studies was a nested case-control study conducted in Spain that estimated risk of subtrochanteric
or femoral shaft fracture compared with no hip fracture.\textsuperscript{42} The study population consisted of women aged 65 years or older (mean 82 years) identified in a national primary practice research database over a 3-year study period. Subtrochanteric or femoral shaft fracture cases were identified using diagnostic codes, with exclusions for excessive trauma and cancer, but no review of radiographs for AFF features. Each case was matched by age and calendar year of enrollment to five controls with no history of hip fracture. Information on bisphosphonate use before the fracture or a corresponding index date was obtained from the research database. Osteoporosis status was not reported.

Two studies compared long-term bisphosphonate use to past bisphosphonate use. The first of these was a nested case-control study conducted in Canada that estimated risk with use of any bisphosphonate for \( \geq 5 \) years versus that with past use for \( <100 \) days.\textsuperscript{44} Cases were defined using diagnostic codes for subtrochanteric or femoral shaft fracture, with exclusions for excessive trauma and cancer/pathological fracture, but no review of radiographs for AFF features. Each case was matched to five controls who had not been hospitalized with a subtrochanteric or femoral shaft fracture. The study population consisted of women aged 68 years or older (median age 83 years) included in a provincial research database over a 5-year period who began treatment with a bisphosphonate within the past 3 years. No results were reported for baseline BMD. About 70 percent of cases and 24 percent of controls reported an osteoporotic fracture in the preceding 5 years. The second of these studies was a retrospective, nested case-control study conducted in Korea.\textsuperscript{40} It compared AFF risk between postmenopausal women with use of any bisphosphonate for \( \geq 5 \) years versus that in women with past use for \( \geq 1 \) year that was stopped between 6 months and 5 years prior to study start date. Participants were identified through electronic hospital records over an 8-year study period. AFF cases were defined using diagnostic codes, with exclusions for excessive trauma and cancer, and radiologic confirmation of AFF features. Mean participant age was 68 years, mean pre-fracture (or most recent) DXA BMD T-score at the total hip and femoral neck were -1.33 and -1.10, respectively. Women with AFF were matched by age and sex to three controls without AFF.

The fifth study compared long-term use of any bisphosphonate versus long-term raloxifene or calcitonin use. This retrospective cohort study, conducted in the United States, estimated risk of subtrochanteric or femoral shaft fracture with use of any bisphosphonate for \( >3 \) years versus that with long-term use of raloxifene or calcitonin.\textsuperscript{43} Cases were defined using diagnosis codes for subtrochanteric or femoral shaft fracture, without exclusions for trauma or cancer/pathologic fracture, and no review of radiographs for AFF features. Study participants were Medicare beneficiaries, mean age was 80 years, and 97 percent of participants were women. Individuals were presumed to have osteoporosis, though only 10 to 15 percent had a prior history of hip or vertebral fracture and no results were reported for baseline BMD. The bisphosphonate and raloxifene/calcitonin groups were propensity score-matched one-to-one.

**Outcomes**

**Atypical Femoral Fractures**

Two controlled observational studies provided low strength evidence that long-term bisphosphonate use was associated with a significantly increased risk of radiologically confirmed AFF. When compared to no bisphosphonate use, long-term (\( \geq 3 \) years) bisphosphonate use was associated with RR 40 to 116\textsuperscript{126} or to past bisphosphonate use (HR 3.36 to 5.17)\textsuperscript{40} (Table 11). Only one of these studies permitted estimation of differences in absolute risk, and found a
statistically significant 11 more radiologically confirmed AFF per 10,000 person-years (95% CI 7, 14) for long-term bisphosphonate treatment versus no osteoporosis drug use.

**Subtrochanteric/Femoral Shaft (ST/FS) Fractures**

The relative risk of subtrochanteric or femoral shaft fractures without radiologically confirmed AFF features in individuals who received long-term bisphosphonate treatment compared with controls varied widely (Table 11). Compared to no use of bisphosphonates, risk of ST/FS fractures was significantly increased with use for >3 years (OR 9.46 [95% CI 2.17, 41.3]). Compared to minimal past bisphosphonate use (<100 days), risk of ST/FS fractures was not significantly increased with 3-5 years of bisphosphonate use (OR 1.59 [95% 0.80, 3.15]), but was increased with ≥5 years of bisphosphonate use (OR 2.74 [95% CI 1.25, 6.02]). Compared to raloxifene or calcitonin use, risk of ST/FS fractures was not increased with either 3-5 years of bisphosphonate use or >5 years of bisphosphonate use.

Risks for AFF and ST/FS fractures varied substantially between studies. This variation in results may have been attributable to several differences in study design, including case definition (whether cases were excluded for trauma or cancer and particularly whether radiological review for AFF features was required), bisphosphonate treatment control group (no bisphosphonate use, past bisphosphonate use, raloxifene or calcitonin use), non-AFF comparison group (non-AFF femoral shaft fracture [risk of which bisphosphonate could have lowered], and no hip fracture), study design (case control vs. retrospective cohort), and the covariables included in the adjusted statistical models. The impact of these factors in isolation was challenging to interpret because they generally were not compared within single studies and studies differed in multiple ways. Nevertheless, the two studies that defined AFF using ASBMR task force criteria and excluded pathological fractures reported high relative risk estimates, ranging from HR 3.36 to 5.17 in one study to RR 40 to 126 in a second study. By comparison, the single study that analyzed risk of subtrochanteric or femoral shaft fractures based solely on diagnostic codes, without exclusions for trauma or cancer and without evaluating radiographs for AFF features, found no increase in risk. For the two studies that used diagnostic codes and excluded cases for trauma or cancer, but didn’t evaluate radiographs for AFF features, risk estimates ranged from OR 1.59 to 2.74 and OR 9.46. Further, the two studies that compared long-term bisphosphonate use to no use reported the highest relative estimates of AFF or subtrochanteric or femoral shaft fracture risk, ranging from OR 9.46 to RR 40 and 126. The studies that compared long-term bisphosphonate use to past bisphosphonate use reported intermediately increased risk of AFF. The study that compared long-term bisphosphonate use to long-term raloxifene or calcitonin use reported no increased risk of subtrochanteric or femoral shaft fracture.

**Osteonecrosis of the Jaw**

No eligible studies compared long-term treatment with bisphosphonates as a class versus a control group and reported on risk of ONJ.
### Table 11. Estimated risk of Atypical Femoral Fracture (AFF)* or Subtrochanteric/Femoral Shaft (ST/FS) Fracture† for Any Bisphosphonate Use from Observational Studies

<table>
<thead>
<tr>
<th>Author, Year Country</th>
<th>Risk of bias</th>
<th>Study design</th>
<th>Case and control definitions</th>
<th>BP treatment duration</th>
<th>Treatment control group</th>
<th>RR or OR (95% CI) Model covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schilcher, 2015†</td>
<td>Medium</td>
<td>Retrospective cohort (n<del>98,200 BP users, n</del>2.8 million non-BP users)</td>
<td>AFF case (n=172 total, n=27 with &gt;4 years BP use): Diagnosis code for subtrochanteric or femoral shaft fracture, exclusions for excessive trauma or pathological fracture, met ASBMR radiographic criteria for AFF. Non-AFF control: Diagnosis code for subtrochanteric or femoral shaft fracture, exclusion for excessive trauma or pathological fracture, did not meet ASBMR radiographic criteria for AFF.</td>
<td>≥ 4 years</td>
<td>Non-use of BP</td>
<td>RR 126 (55, 238) (women only) Age</td>
</tr>
<tr>
<td>Koh, 2017†</td>
<td>Medium</td>
<td>Case-control</td>
<td>Case-control AFF case (n=172 total, n=17 with 3-4 years of BP use): Same as above Non-AFF control (n=952): Same as above</td>
<td>3-4 years</td>
<td>Same as above</td>
<td>OR 40 (17, 91) Age, sex, cortisone use, Charlson’s comorbidity index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case-control AFF case (n=172 total, n=16 with 4-5 years of BP use): Same as above Non-AFF control (n=952): Same as above</td>
<td>4-5 years</td>
<td>Same as above</td>
<td>OR 116 (58, 234) Age, sex, cortisone use, Charlson’s comorbidity index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case-control AFF case (n=172 total, n=11 with &gt;5 years of BP use): Same as above Non-AFF control (n=952): Same as above</td>
<td>&gt;5 years</td>
<td>Same as above</td>
<td>OR 93 (66, 132) Age, sex, cortisone use, Charlson’s comorbidity index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AFF case (n=43): Diagnosis codes for subtrochanteric and femoral shaft fracture, exclusions for cancer/pathologic fracture, met ASBMR radiographic criteria for AFF. Non-AFF control (n=129): No history of AFF.</td>
<td>≥5 years</td>
<td>No BP use &gt;6 months to 5 years after ≥1 year of BP treatment</td>
<td>Model 1: HR 5.17 (2.0, 13.36) Model 2: HR 4.37 (1.68, 11.41) Model 3: HR 3.36 (1.77, 11.91) Model 1: BMI, continued bisphosphonate, long-term glucocorticoid use Model 2: BMI, continued bisphosphonate, rheumatoid arthritis Model 3: BMI, continued bisphosphonate, disease-modifying anti-rheumatic drugs</td>
</tr>
<tr>
<td>Author, Year</td>
<td>Study design</td>
<td>Case and control definitions</td>
<td>BP treatment duration</td>
<td>Treatment control group</td>
<td>RR or OR (95% CI)</td>
<td>Model covariates</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>------------------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>shaft fracture (ST/FS fracture)†</td>
<td>Case-control</td>
<td>ST/FS cases (n=44): Diagnosis codes for subtrochanteric and femoral shaft fracture with exclusions for excessive trauma and cancer, but no review of radiographs. Non-ST/FS fracture control (n=220): No hip fracture.</td>
<td>&gt;3 years</td>
<td>Non-use of BP</td>
<td>OR 9.46 (2.17, 41.3)</td>
<td>Age, calendar year of enrollment in primary care database, smoking, alcoholism, BMI, previous fracture, comorbidities, other medications (including raloxifene, hormone therapy)</td>
</tr>
<tr>
<td>Erviti, 2013</td>
<td>Case-control</td>
<td>ST/FS cases (n=204): Hospitalized with diagnosis codes for first subtrochanteric or femoral shaft fracture, with exclusions including for excessive trauma, cancer/pathologic fracture, and use of non-BP osteoporosis medications, but no review of radiographs. Non-ST/FS fracture control (n=1070): Not hospitalized with subtrochanteric or femoral shaft fracture.</td>
<td>3-5 years</td>
<td>&lt;100 total days of past BP use</td>
<td>OR 1.59 (0.80, 3.15)</td>
<td>Socioeconomic status, comedinations, drug count, comorbidities, recent medical visits, prior fall, prior osteoporotic fracture, BMD test past 5 years</td>
</tr>
<tr>
<td>Park-Wyllie, 2011</td>
<td>Case-control</td>
<td>ST/FS cases (n=121): Same as above. Non-ST/FS fracture control (n=460): Same as above.</td>
<td>≥5 years</td>
<td>Same as above</td>
<td>OR 2.74 (1.25, 6.02)</td>
<td>Socioeconomic status, comedinations, drug count, comorbidities, recent medical visits, prior fall, prior osteoporotic fracture, BMD test past 5 years</td>
</tr>
<tr>
<td>Kim, 2011</td>
<td>Retrospective cohort (n=2,591 BP users, n=2,309 raloxifene or calcitonin users)</td>
<td>ST/FS cases (n=26): Diagnosis codes for subtrochanteric and femoral shaft fracture, but no exclusions for trauma or cancer/pathologic fracture, and no review of radiographs. Non-ST/FS fracture control: Not specified.</td>
<td>3-5 years</td>
<td>Raloxifene or calcitonin</td>
<td>HR 1.20 (0.55, 2.61)</td>
<td>Propensity score–matched by demographics, health care utilization, comorbidities, and other medications</td>
</tr>
<tr>
<td></td>
<td>Retrospective cohort (n=2,371 BP users, n=1,726 raloxifene or calcitonin users)</td>
<td>ST/FS cases (n=8): Same as above. Non-ST/FS fracture control: Not specified.</td>
<td>&gt;5 years</td>
<td>Same as above</td>
<td>HR 2.02 (0.41, 10.0)</td>
<td>Propensity score–matched by demographics, health care utilization, comorbidities, and other medications</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; ASBMR=American Society for Bone and Mineral Research; BMI=body mass index; BP=bisphosphonates; CI=confidence interval; HR=adjusted hazard ratio; N=number; OR=odds ratio; RR=risk ratio; ST/FS=subtrochanteric/femoral shaft shaft fracture (ST/FS fracture)† l
*Atypical femoral fractures must have been defined as meeting ASBMR radiographic criteria for AFF.

†Subtrochanteric/femoral shaft fractures were fractures of these sites not stated as meeting ASBMR radiographic criteria for AFF.
Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We looked at whether the harms of long-term bisphosphonate therapy versus control vary as a function of patient, bone, and/or drug characteristics. Eligible studies included evidence for the following potential effect modifiers:

Age
A Swedish retrospective cohort study reported risk of radiologically confirmed AFF with long-term bisphosphonate use versus no bisphosphonate use as a function of age in women only. The age-adjusted risk of AFF for bisphosphonate treatment compared with no use of osteoporosis drugs was OR 100 (95% CI 40, 253) in women <80 years of age versus OR 163 (95% CI 39, 687) in women ≥80 years of age. Authors did not report results for a test of interaction.

Bisphosphonate Treatment Duration
Three studies each reported risk estimates for AFF or subtrochanteric or femoral shaft fractures between long-term bisphosphonate use and control for multiple treatment duration strata. Collectively they suggested that within the range of >3 years of bisphosphonate use, risk of these fractures may be higher with longer treatment duration.

One study reported on risk of AFF. It found that compared with no osteoporosis drug use, risk of AFF with radiologically confirmed AFF features with long-term bisphosphonate use appeared greater with bisphosphonate use of either 4 to 5 years (OR 116 [95% CI 58, 234]) or >5 years (OR 93 [95% CI 66, 132]) than with use for 3 to 4 years (OR 40 [95% CI 17, 91]).

Two studies reported on risk of subtrochanteric/femoral shaft fractures. The first of these reported that compared with past bisphosphonate use for <100 days, risk of subtrochanteric or femoral shaft fracture was not significantly increased with 3 to 5 years of bisphosphonate use (OR 1.59 [95% CI 0.80, 3.15]), but was increased with >5 years of bisphosphonate use (OR 2.74 [95% CI 1.25, 6.02]). The second study reported risk for subtrochanteric or femoral shaft fractures with long-term bisphosphonate use versus raloxifene or calcitonin use. Risk was numerically higher with >5 years use (HR 2.02 [95% CI 0.41, 10.0]) than with 3 to 5 years use (HR 1.20 [95% CI 0.55, 2.61]), but neither result was statistically significant. Confidence intervals overlapped in all studies and none tested for an interaction as a function of treatment duration.

Other Effect Modifiers Included in Review
No evidence was found for the following effect modifiers: race, sex, comorbid conditions (DM, CKD, CVD), osteoporosis status, fracture history, calculated pre-treatment fracture risk (e.g., FRAX®), pretreatment and early treatment (e.g. 1 year) imaging (L-spine, total hip & femoral neck DXA BMD), biochemical markers (CTX, NTX, P1NP, BSAP), treatment dose, treatment frequency, or treatment delivery route.
Denosumab

Key Points

Key Question 3
• In postmenopausal women with osteoporosis, between denosumab versus placebo for 4 years:
  o There was no difference in risk of serious adverse events (low SOE).
  o Evidence was insufficient to draw conclusions about differences in risk of AFF or ONJ.

Key Question 4
• We found no evidence about whether differences in risk of harms between long-term denosumab and placebo vary as a function of patient, bone or drug characteristics.

Eligible Studies
Two eligible publications of two RCTs and no eligible observational studies compared long-term treatment with denosumab versus placebo, no treatment or another active treatment and reported on risk of harms. One publication was rated high ROB99 and only limited data was extracted (Appendix D7). The remaining publication26, which had medium ROB, was extracted in detailed evidence tables, SOE assessments (Appendix D), and summary ROB assessments (Appendix C).

Described in detail previously, this trial randomized postmenopausal women with osteoporosis or low bone mass/osteopenia to placebo (n=46) versus one of seven denosumab regimens (n=319) for 2 years. After 2 years, those initially assigned placebo remained assigned to placebo. Those initially assigned denosumab continued denosumab for 2 years (n=231), switched to placebo for 2 years (n=47), or switched to placebo for 1 year before restarting denosumab for 1 year (n=41).

Outcomes

Serious Adverse Events
Through 4 years, risk of any serious adverse event was not significantly different between women assigned denosumab versus placebo (17.8% vs. 10.9%; RR 1.64 [95% CI 0.69, 3.88]) (low SOE). However, harms results were reported collectively for all women initially assigned to denosumab, with no separate results provided for those assigned to denosumab for four years, those assigned to denosumab for 2 years followed by placebo for 2 years, or those assigned to denosumab for two years followed by placebo for 1 year and then denosumab again for 1 year. As a result of this data pooling, no direct comparison between the 4 years of continuous denosumab and placebo groups was possible.
Cardiovascular Events
   One fatal cardiovascular event (stroke) was reported in the denosumab group.

Mortality
   There were four deaths (1.3%) in the denosumab group and none in the placebo group.

Atypical Femoral Fracture
   No information was reported about incidence of AFF.

Osteonecrosis of the Jaw
   No information was reported about incidence of ONJ.

Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics
   We identified no eligible studies that compared long-term denosumab treatment versus placebo, no treatment or a different active treatment, and assessed whether risk of harms varied as a function of patient, bone or drug characteristics on risk of harms.
Raloxifene

Key Points

Key Question 3

- In postmenopausal women with osteoporosis, between raloxifene and placebo:
  - There was no difference in risk of serious adverse events at 8 years (low SOE).
  - There was a higher risk of venous thromboembolism at both 4 years and 8 years.
- Evidence was insufficient to draw conclusions about differences between long-term raloxifene and either placebo or no osteoporosis drug treatment for risk of AFF or ONJ.

Key Question 4

- Evidence was insufficient to draw conclusions about whether differences in risk of harms between long-term raloxifene and placebo, no osteoporosis drug treatment or a different active treatment vary as a function of patient, bone, or drug characteristics.

Eligible Studies

Twelve eligible publications from one RCT and its extension\(^{24, 35, 36, 49, 50, 55-61}\) and three eligible publications from one observational study\(^{39, 45, 48}\) compared long-term raloxifene treatment with placebo, no treatment or a different active treatment, and reported on risk of harms. All studies had low or medium ROB, and information from these studies was extracted in evidence tables, SOE summary tables (Appendix D) and ROB summary tables (Appendix C).

Eight reports from the MORE RCT with low ROB compared raloxifene to placebo through 4 years. Two compared raloxifene 60 mg/day to placebo (n=5114 to 5133)\(^{35, 55}\), three compared a pooled raloxifene group (60 mg/day and 120 mg/day) to placebo (n=7617 to 7705)\(^{49, 50, 57}\), and three separately compared raloxifene 60 mg/day and 120 mg/day to placebo (n=6828 to 7705).\(^{24, 36, 56}\) Four medium ROB studies from the CORE extension compared raloxifene 60 mg/day to placebo through 8 years (n=4011 to 5133). Details of the MORE and CORE study designs are reported above.

The three reports from one observational study compared raloxifene (any dose) to no osteoporosis treatment.\(^{39, 45, 48}\) The study sample was drawn from the Danish general population and consisted of 4,831 individuals who filled a prescription for raloxifene and 14,493 age and gender matched controls. Mean age was 64 years. Participants generally were healthy with few comorbidities. In a larger study sample, that also included individuals who received other osteoporotic drugs and their matched controls, 85 percent of participants were women and mean follow-up was 3.8 years.

Outcomes

Serious Adverse Events

Among women who enrolled in CORE, there was no difference in risk of serious adverse events between raloxifene and placebo during the four years of CORE follow-up (years 5-8 from their MORE enrollment; 23% vs. 25%, p=0.22).\(^{60}\) Similarly, among CORE enrollees, there was no difference in risk of treatment-emergent serious adverse events between raloxifene and
placebo during their combined 8 years of follow-up in MORE and CORE (42% vs. 46%; RR 0.93 [95% CI 0.86, 1.00]) (low SOE).  

**Venous Thromboembolism**

After 4 years in the MORE trial, which excluded individuals with a venous thromboembolic event in the last 10 years, incidence of deep vein thrombosis (DVT) during 4 years of follow-up was higher in the pooled raloxifene group (60 mg/day and 120 mg/day) compared with placebo.  

However, estimates of effect between publications ranged slightly from RR 2.8 (95% CI 1.3, 5.9) to RR 3.1 (95% CI 1.4, 6.9). Five MORE papers suggested that raloxifene (all but one paper pooled the 60 mg/day and 120 mg/day doses) was associated with an increased risk of pulmonary embolism (PE) versus placebo. However, results only were statistically significant in one paper (RR 4.5 [95% CI 1.1, 19.5]), and were borderline significant in the others. The number of PE cases varied between reports (0.31% to 0.43% for raloxifene and 0.08% to 0.23% for placebo). Through 8 years, among women who participated in the CORE extension (n=4011), there was no difference between raloxifene 60 mg/day and placebo in risk of DVT, but risk of PE was higher in the raloxifene group compared with placebo.

**Hot Flashes**

Through 8 years, among women who participated in the CORE extension (n=4011), risk of hot flashes was higher in the raloxifene group compared with placebo.

**Cardiovascular Disease**

Through eight years, among women who participated in the CORE extension (n=4011), there was no difference between raloxifene 60 mg/day and placebo in risk of stroke or cardiovascular mortality.

**Mortality**

Risk of all-cause mortality was not different between treatment groups when considering follow-up only during the CORE extension period. However, mortality was lower with raloxifene versus placebo when considering only the raloxifene group originally randomized to 60 mg/day and considering follow-up through both the MORE and CORE phases (HR 0.68 [95% CI 0.46, 0.99]).

**Atypical Femoral Fractures**

In data from one observational study, compared with age and gender-matched controls who received no osteoporosis treatment, evidence was insufficient about whether individuals treated with raloxifene had a difference in risk of either subtrochanteric (HR 1.06 [95% CI 0.34, 3.32]) or femoral shaft (HR 0.82 [95% CI 0.21, 3.20]) fracture. No data were reported about AFF in either the MORE trial or CORE extension.
Osteonecrosis of the Jaw

In data from one observational study, compared with age and gender-matched controls who received no osteoporosis treatment, evidence was insufficient about whether individuals treated with raloxifene had a difference in risk of any inflammatory jaw event (none of 4,831 and two of 14,493 participants in the raloxifene and no osteoporosis treatment groups, respectively). No data were reported about ONJ in either the MORE trial or CORE extension.

Atrial Fibrillation

In data from one observational study, compared with age and gender-matched controls who received no osteoporosis treatment, individuals treated with raloxifene appeared to have no increased risk of atrial fibrillation (OR 0.98 [95% CI 0.72, 1.33]), though SOE was not assessed. No data were reported about atrial fibrillation in either the MORE trial or CORE extension.

Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We looked at whether the harms of long-term raloxifene therapy versus placebo, no treatment or a different active treatment vary as a function of patient, bone, and/or drug characteristics. Eligible studies included evidence for the following effect potential modifiers.

Baseline Risk for Cardiovascular Disease

In one publication, though risk of incident stroke was no different between raloxifene and placebo overall, risk was lower versus placebo in the subgroup of women at increased baseline cardiovascular risk (RR 0.38 [95% CI 0.15, 0.94]). A second publication found that risk of VTE, mostly comprised of DVT or PE, was not significantly different between raloxifene 60 mg/day and placebo in the overall study population (HR 1.86 [95% CI 0.97, 3.56]). It also was not significantly different in a subgroup with increased cardiovascular risk (HR 1.32 [95% CI 0.37, 4.69]).

Prior Fracture

In one publication, the effect of raloxifene 60 mg/day versus placebo on risk of stroke, PE, DVT, hot flashes and mortality did not differ as a function of baseline radiographic vertebral fracture status.

Raloxifene Dose and Treatment Duration

One publication reported that there was no significant difference in risk of DVT or PE between the raloxifene 60 mg/day and 120 mg/day groups, and that risks of VTE and DVT (but not of PE) were higher in the pooled raloxifene groups versus placebo through 2 years of treatment but not in later years.

Other Effect Modifiers Included in Review

No evidence was found for the following effect modifiers: pretreatment age, race, sex, comorbid conditions (DM, CKD), osteoporosis status, calculated pre-treatment fracture risk (e.g.,
FRAX®, pretreatment and early treatment (e.g. 1 year) imaging (L-spine, total hip & femoral neck DXA BMD), biochemical markers (CTX, NTX, P1NP, BSAP), treatment dose, treatment frequency, or treatment delivery route
Hormone Therapy

Key Points

- In postmenopausal women with osteoporosis, evidence was insufficient to draw conclusions about any differences in risk for long-term harms between estrogen/progestin therapy and control.

 Eligible Studies

Described previously, we identified 1 small eligible RCT of estrogen/progestin reporting long-term harms. Postmenopausal women with osteoporosis, defined as 1 to 4 baseline radiographic vertebral fractures and LS-BMD T-score ≤-2.0, were randomized to receive estrogen/progestin (n=18) or control (n=18) over 4 years. Both groups received supplemental calcium 1000 mg/day and vitamin D 400 units/day.

Outcomes

Serious Adverse Events

Incidence of serious adverse events was not reported (insufficient evidence). Through 4 years, there was no apparent difference between the estrogen/progestin and control groups in withdrawals due either to adverse events or other medical problems (17% vs. 17%, p=1.0).

Mortality

No deaths were reported for either the estrogen/progestin or the control groups.

ONJ or AFF

No information was reported about incidence of ONJ or AFF.

Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We identified no eligible studies that compared estrogen or estrogen/progestin with control, and assessed whether risk of harms varied as a function of patient, bone or drug characteristics.
Other Drugs Investigated for Review

Ibandronate

One eligible study reported on harms with long-term treatment with ibandronate versus placebo or another active treatment. This study was rated high ROB and only limited data was extracted (Appendix D7). We identified no eligible studies that compared ibandronate with control, and assessed whether risk of harms varied as a function of patient, bone or drug characteristics.

Risedronate

Two eligible studies of long-term treatment with risedronate versus inactive control (e.g., placebo, no treatment) or a different active treatment reported on harms. These studies were rated high ROB and only limited data was extracted (Appendix D7). We identified no eligible studies that compared risedronate with control, and assessed whether risk of harms varied as a function of patient, bone or drug characteristics.
Chapter 6. Effects of Drug Holidays

Chapter 6 discusses Key Questions 5 and 6. Key Question 5 addresses the effects of drug therapy continuation versus discontinuation (placebo drug holiday) on risk of incident fracture. Key Question 6 examines whether the effect on risk of fracture varies by patient, bone, or drug characteristics.

Alendronate

Key Points

Key Question 5

- In postmenopausal women who had received 5 years of alendronate for FN-BMD T-score \(\leq -1.6\), alendronate continuation for 5 more years versus alendronate discontinuation was associated with:
  - No difference in risk of incident clinical fractures (moderate SOE), incident nonvertebral fractures (moderate SOE), incident hip fractures (low SOE), or incident radiographic vertebral fractures (low SOE).
  - Lower risk of incident clinical vertebral fractures (moderate SOE).
- In postmenopausal women who had received 5 years of alendronate for LS-BMD T-score \(\leq -2.5\):
  - Alendronate continuation for 2 more years versus alendronate discontinuation was associated with no difference in risk of incident nonvertebral fractures (low SOE) or incident clinical vertebral fractures (low SOE).
  - Between alendronate continuation for 5 more years versus alendronate discontinuation, evidence was insufficient to draw conclusions about differences in risk of incident nonvertebral fracture during follow-up years 8-10.

Key Question 6

- In postmenopausal women who had received 5 years of alendronate for FN-BMD T-score \(\leq -1.6\):
  - The effect of alendronate continuation for 5 more years versus alendronate discontinuation on risk of incident nonvertebral fracture and incident clinical fracture did not appear to vary as a function of baseline BMD or prevalence of baseline radiographic vertebral fracture.
  - Among those with a baseline radiographic vertebral fracture, the effect of alendronate continuation for 5 more years versus alendronate discontinuation on risk of incident nonvertebral fracture, incident clinical fracture and incident radiographic vertebral fracture did not appear to vary as a function of baseline BMD.
  - Among those without a baseline radiographic vertebral fracture, the effect of alendronate continuation for 5 more years versus alendronate discontinuation:
    - On risk of incident nonvertebral fracture, appeared to vary as a function of baseline BMD.
On risk of incident clinical fracture and incident radiographic vertebral fracture, did not appear to vary as a function of baseline BMD.

Eligible Studies

Four eligible publications of two unique studies compared alendronate continuation versus discontinuation and reported on risk of incident fractures. We rated three of these publications as having low ROB and one as having medium ROB. We extracted detailed information from these studies in evidence tables, SOE summary tables (Appendix D), and ROB summary tables (Appendix C).

Two low ROB publications were from the Fracture Intervention Trial Long Term Extension (FLEX), an extension of the previously described U.S.-based FIT study. Eligibility in FLEX was limited to women assigned to alendronate in either FIT-I or FIT-II (n=3236 total from both studies) who received at least 3 years (mean 5.0 years) of alendronate during combined follow-up in FIT (mean 3.8 years) and a subsequent open-label period (mean 1.9 years). Further, they must have had a TH-BMD T-score at FLEX baseline that was ≥-3.5 and no worse than at their FIT baseline. The 1099 FLEX enrollees were randomized to continuation of alendronate 10 mg/day for 5 years versus discontinuation (placebo drug holiday). Change in TH-BMD was the primary efficacy outcome and incident fracture was a secondary outcome. At FLEX baseline, mean participant age was 73 years, 60 percent of participants reported a clinical fracture since age 45, 34 percent had a prevalent radiographic vertebral fracture, and 78 percent were taking alendronate (Appendix D5). Though 29 percent of women had FN-BMD T-scores of ≤-2.5, mean FN-BMD T-scores were in the osteopenic/low bone mass range.

Two publications were extensions of a pair of nearly identical dose-ranging RCTs, one conducted in the US and the other in several non-U.S. countries. Together, these studies enrolled 994 postmenopausal women aged 45-80 years with LS-BMD T-score ≤-2.5. Women were randomized to one of three alendronate regimens (5 mg/day for 3 years, 10 mg/day for 3 years, or 20 mg/day for 2 years followed by 5 mg/day for 1 year) or placebo. All results for the two RCTs were pooled. At baseline, mean age was 63 years and 21 percent of participants had a prevalent radiographic vertebral fracture. Among the 598 women initially assigned alendronate, 439 participated in an extension in which they continued their original alendronate regimen for 2 years (5 mg/day, 10 mg/day, and continuing 5 mg/day for the third group). Though all women knew they were receiving alendronate, they and the investigators were blinded to their dose regimen. After 5 years, 350 of these women entered another 2-year extension, in which women originally assigned alendronate 5 mg/day and 10 mg/day continued these regimens and the group originally assigned alendronate 20 mg/day and later changed to 5 mg/day was switched to blinded placebo. This study was rated as having low ROB. Although participants remained blinded to their treatment assignments, these assignments were not random. In a final extension, 247 women continued these blinded treatment assignments an additional 3 years. This last extension study was rated as having medium ROB. Change in LS-BMD was the primary efficacy outcome. Secondary outcomes included change in FN-BMD. Incident fractures were reported as safety outcomes. Authors did not report participant characteristics from the baseline of the extension studies.
Outcomes

Incident clinical fractures

In FLEX, whose participants previously were treated with alendronate for a mean of five years, women randomized to continued alendronate for 5 years versus discontinuation (placebo) had no difference in risk of any incident clinical fracture (19.9% vs. 21.3%; RR 0.93 [95% CI 0.71, 1.21]) (moderate SOE), nonvertebral fracture (18.9% vs. 19.0%; RR 1.00 [95% CI 0.76, 1.32]), (moderate SOE) or hip fracture (3.0% vs. 3.0%; RR 1.02 [95% CI 0.51, 2.10]) (low SOE). However, continued alendronate reduced risk of incident clinical vertebral fracture (2.4% vs. 5.3%; RR 0.45 [95% CI 0.24, 0.85]) (moderate SOE).62

In the dose-ranging extension study, in results pooling both alendronate continuation arms versus the discontinuation arm, for the year 6-7 extension, neither incidence of nonvertebral fracture (RR 0.87 [95% CI 0.40, 1.91]) nor incidence of clinical vertebral fracture (RR 0.92 [95% CI 0.40, 2.10]) was different between treatment groups (both low SOE).64 For the year 8-10 extension, evidence was insufficient about whether incidence of nonvertebral fracture differed between the pooled alendronate continuation and discontinuation groups (RR 0.81 [95% CI 0.38, 1.71]).63

Incident radiographic vertebral fractures

In FLEX, continued alendronate versus discontinuation (placebo) did not reduce risk of incident radiographic vertebral fractures (9.8% vs. 11.3%; RR 0.86 [95% CI 0.60, 1.22]) (low SOE).62 In the dose-ranging extension study, the proportion of women with an incident radiographic vertebral fracture between years 6-10 did not appear to differ significantly between the continued alendronate 5 mg/day, alendronate 10 mg/day and placebo (discontinuation) groups (13.9%, 5.0% and 6.6%, respectively). Again, this study reported no formal statistical testing.63

Bone Mineral Density

In FLEX, during the 5-year follow-up on continued alendronate versus discontinuation (placebo), though both treatment groups experienced a mean decline in TH-BMD, this BMD loss was significantly smaller in those assigned to continue alendronate (-1.02% vs. -3.38%; MD 2.36 [95% CI 1.81, 2.90]).62 Results for other hip sites showed essentially no change in BMD in women assigned alendronate for an additional five years versus a significantly greater 2 to 3 percent decline from FLEX baseline in those randomized to discontinuation. Although both treatment groups had an increase in LS-BMD during FLEX, the gain was significantly greater in the alendronate continuation group (MD 3.74 [95% CI 3.03, 4.45]). Authors stated that sensitivity analyses in which data were not carried forward to replace missing values yielded similar results.

The dose-ranging extension studies63, 64 did not report between treatment group differences in BMD change, but did report results within groups. For study years 6-10, the discontinuation group experienced a mean decline of approximately 2 percent in TH-BMD and FN-BMD and no change in LS-BMD. By comparison, neither TH-BMD or FN-BMD significantly changed while LS-BMD increased in both continuing alendronate dose groups.63
Variation in Effects of Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We looked at whether the effects of alendronate continuation versus discontinuation on risk of incident fracture varies as a function of patient, bone, and/or drug characteristics. Eligible studies included evidence for the following potential effect modifiers:

Baseline BMD
One post hoc FLEX analysis reported that risk of incident nonvertebral fracture and incident clinical vertebral fracture, neither of which differed between the alendronate continuation and discontinuation groups overall, also did not differ as a function of baseline FN-BMD (interaction p-values 0.40 and 0.72, respectively). A second post hoc FLEX analysis stratified further and reported that within the subset of women with a prevalent radiographic vertebral fracture at FLEX baseline, the effect of continued alendronate versus discontinuation on risk of incident nonvertebral fracture, clinical vertebral fracture and radiographic vertebral fracture did not differ as a function of baseline FN-BMD. In women without a prevalent radiographic vertebral fracture at FLEX baseline, the effect of continued alendronate versus discontinuation on risk of incident nonvertebral fracture differed as a function of baseline FN-BMD (interaction p-value 0.019). These results suggested that compared to women with higher baseline FN-BMD, women with the lowest FN-BMD levels may have greater reduction in risk of incident nonvertebral fracture with continued versus discontinued alendronate. However, the effect of continued versus discontinued alendronate on risk of incident clinical vertebral fracture or incident radiographic vertebral fracture did not appear to differ as a function of baseline FN-BMD (interactions not significant).

History of Prior Fracture
One post hoc FLEX analysis reported that risk of incident nonvertebral fracture between alendronate continuation and discontinuation did not differ as a function of prevalent radiographic vertebral fracture (interaction p-value 0.23). Similarly, risk of incident clinical vertebral fracture neither differed with treatment overall, nor as a function of prevalent radiographic vertebral fracture (interaction p-value 0.86).

Other Effect Modifiers Included in Review
No evidence was found for the following effect modifiers: age, sex, race, osteoporosis status, calculated fracture risk [e.g. FRAX®], comorbid conditions (DM, CKD, CVD), biomarkers, pre-drug holiday agent/class, time between drug initiation and start of drug holiday, duration of drug holiday, or post-drug holiday agent/class.
Zoledronic Acid (Zolendronate)

Key Points

Key Question 5

- In postmenopausal women with osteopenia, between zoledronic acid continued for 2 years and zoledronic acid stopped after 1 year and switched to placebo for 1 year:
  - Evidence was insufficient to draw conclusions about differences in risk of incident clinical fractures.

- In postmenopausal women with osteoporosis, between zoledronic acid continued for 6 years and zoledronic acid stopped after 3 years and switched to placebo for 3 years:
  - There was no difference in risk of incident clinical fractures (low SOE) or incident nonvertebral fractures (low SOE).
  - There was a lower risk of incident radiographic vertebral fractures (moderate SOE).
  - Evidence was insufficient to draw conclusions about differences in risk of incident clinical vertebral fractures or risk of incident hip fractures.

- In postmenopausal women with osteoporosis, between zoledronic acid continued for 9 years and zoledronic acid stopped after 6 years and switched to placebo for 3 years:
  - Evidence was insufficient to draw conclusions about differences in risk of incident clinical fractures or incident radiographic vertebral fractures.

Key Question 6

- We identified no evidence about whether differences in risk of incident fractures between zoledronic acid continuation and discontinuation vary as a function of patient, bone or drug characteristics.

Eligible Studies

Four eligible publications of two unique RCTs compared continuation of zoledronic acid versus discontinuation of zoledronic acid (placebo drug holiday) and reported on risk of incident fractures. We rated one of these studies as having high ROB and extracted only limited data (Appendix D7). The remaining three publications of two unique RCTs all were rated as having low ROB. We extracted additional information from these studies in evidence tables and SOE summary tables (Appendix D).

The earliest study, by McClung and colleagues, compared 2 years of 5 mg zoledronic acid (Z2) to one year of 5 mg zoledronic acid followed by 1 year of placebo (Z1/P1) in postmenopausal women with osteopenia (meeting all the following criteria: LS-BMD < -1 and > -2.5, FN-BMD > -2.5, no grade two or three baseline radiographic vertebral fracture, zero to one grade one baseline radiographic vertebral fracture). The other two publications were extensions of the Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly-Pivotal Fracture Trial (HORIZON), an international, multisite trial of 7765 postmenopausal women with osteoporosis (FN-BMD T-score ≤ -2.5, or FN-BMD T-score ≤ -1.5 with at least two mild or one moderate baseline radiographic vertebral fracture) that compared annually administered intravenous zoledronic acid 5 mg to placebo. Following the initial 3-year trial period, women
initially randomized to zoledronic acid were randomized to continue zoledronic acid 3 more
years for a total of 6 years (Z6) or to switch to placebo for three more years (Z3/P3).65 A second
extension was conducted in which women randomized to zoledronic acid for 6 years were then
randomized to remain on zoledronic acid 3 more years for a total of 9 years (Z9) or switch to
placebo for 3 years (Z6/P3).66

Table 12. Baseline characteristics of the HORIZON extension trials at the start of their extensions

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>HORIZON, extension 1&lt;sup&gt;e&lt;/sup&gt;</th>
<th>HORIZON, extension 2&lt;sup&gt;e&lt;/sup&gt;</th>
<th>McClung&lt;sup&gt;f&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number enrolled and randomized</td>
<td>N=1233</td>
<td>N=190</td>
<td>N=379</td>
</tr>
<tr>
<td>Age, mean years</td>
<td>76</td>
<td>78</td>
<td>60</td>
</tr>
<tr>
<td>Gender, women</td>
<td>100%</td>
<td>NR</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white</td>
<td>NR (mostly European)</td>
<td>NR (mostly European)</td>
<td>93%</td>
</tr>
<tr>
<td>Mean FN-BMD T-score</td>
<td>-2.57</td>
<td>-2.44</td>
<td>-1.42</td>
</tr>
<tr>
<td>Mean TH-BMD T-score</td>
<td>-2.07&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-1.98&lt;sup&gt;*&lt;/sup&gt;</td>
<td>NR</td>
</tr>
<tr>
<td>FN-BMD T-score ≤-2.5</td>
<td>55%</td>
<td>45%</td>
<td>0%</td>
</tr>
<tr>
<td>FN-BMD T-score &gt;-2.5 to -1.5</td>
<td>41%</td>
<td>47%</td>
<td>NR</td>
</tr>
<tr>
<td>FN-BMD T-score &gt;-1.5</td>
<td>4%</td>
<td>7%</td>
<td>NR</td>
</tr>
<tr>
<td>History of clinical fracture</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Prevalent radiographic vertebral fracture, none</td>
<td>39%</td>
<td>44%</td>
<td>NR</td>
</tr>
<tr>
<td>Prevalent radiographic vertebral fracture, one</td>
<td>28%</td>
<td>28%</td>
<td>NR</td>
</tr>
<tr>
<td>Prevalent radiographic vertebral fracture, &gt;2</td>
<td>33%</td>
<td>28%</td>
<td>0%</td>
</tr>
<tr>
<td>Comorbid conditions</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: BMD=bone mineral density; FN-BMD=femoral neck bone mineral density; HORIZON= Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly-Pivotal Fracture Trial; NR=not reported; TH-BMD=total hip bone mineral density
*Calculated by EPC

Outcomes

Incident Clinical Fractures

In the McClung trial, incident fractures were neither a primary nor secondary endpoint, but
were reported as adverse events.68 Incidence of fractures was not separately reported for the
second year of the trial, during the time when treatment groups were receiving different
treatments, but only for the 2-year trial period as a whole. During this 2-year period, evidence
was insufficient about whether risk of any incident clinical fracture differed between the Z2 and
Z1/P1 groups (3.0% vs. 2.2%; RR=1.37; 95% CI 0.39, 4.78).

Incident fractures were secondary endpoints in both HORIZON zoledronic acid extension
RCTs. During the first extension study (years 4-6), there were no differences between the Z6 and
Z3/P3 treatment groups in risk of any incident clinical fractures (HR 1.04 [95% CI 0.71, 1.54]) (low SOE) or incident nonvertebral fractures (7.6% vs. 8.2%; HR 0.99 [0.7, 1.5]) (low SOE). However, evidence was insufficient about whether risks between groups differed for incident hip fracture (1.3% vs. 1.4%; HR=0.90 [95% CI 0.33, 2.49]), or incident clinical vertebral fracture (HR 1.81 [95% CI 0.53, 6.2]).

During the second HORIZON extension period (years 7-9), 190 women enrolled and few incident fractures occurred. Evidence was insufficient about whether risk of incident clinical fractures differed between the Z9 and Z6/P3 treatment groups (11% vs. 9%; HR 1.11 [95% CI 0.45, 2.73]).

### Incident Radiographic Vertebral Fractures

In the first HORIZON extension study (years 4-6), risk of incident radiographic vertebral fracture was lower in the Z6 group versus the Z3/P3 group (3.0% vs. 6.2%; OR 0.51 [95% CI 0.26, 0.95]) (moderate SOE). During the second HORIZON extension (years 7-9), there were few incident radiographic vertebral fractures and risk was not different between the Z9 and Z6/P3 treatment groups (3.2% vs. 5.3%; OR 0.58 [95% CI 0.13, 2.55]) (insufficient SOE).

### Change in BMD

The primary study endpoints in the McClung trial and the two HORIZON extension studies were change in LS-BMD, FN-BMD, and TH-BMD, respectively. Study results are reported here, but strength of evidence was not assessed for BMD outcomes in this report.

In the McClung trial, compared with the Z1/P1 group, the Z2 group had a small but statistically significantly larger increase in LS-BMD during the 2-year follow-up period (5.2% vs. 4.4%; MD 0.76% [95% CI 0.70, 0.82]). Small but statistically significant differences in TH-BMD and FN-BMD also favored the Z2 group.

In the first HORIZON extension, 75 percent (n=921) of participants randomized at year 3 completed follow-up BMD measures at year 6. During this interval, BMD levels in the Z3/P3 group regressed slightly but did not fall below pretreatment levels, while BMD levels in the Z6 group remained constant. This resulted in a small but statistically significant difference between treatment groups in change in FN-BMD that favored the Z6 group (MD 1.04% [95% CI 0.43, 1.65]). Results similarly favored the Z6 group for change in TH-BMD (MD 1.22% [95% CI 0.75, 1.70]). In the second HORIZON extension, 72 percent (n=136) of participants randomized at year 6 completed follow-up BMD measures at year 9. Mean changes from year 6 to year 9 in FN-BMD and TH-BMD were similar in the Z9 and Z6/P3 groups.

### Variation in Effects of Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We identified no eligible trials that compared zoledronic acid continuation versus discontinuation (placebo drug holiday) and assessed whether risk of incident fracture varied as a function patient, bone or drug characteristics.
Denosumab

Key Points

Key Question 5
- Evidence was insufficient to draw conclusions about whether risk of incident clinical fractures differed between denosumab continuation and discontinuation (placebo drug holiday).

Key Question 6
- We identified no evidence about whether differences in risk of incident fractures between denosumab continuation and discontinuation vary as a function of patient, bone or drug characteristics.

Eligible Studies
One eligible publication of one RCT compared denosumab continuation versus discontinuation (placebo drug holiday) and reported on risk of fractures. Data from this study, which had medium ROB, was extracted in detailed evidence tables, SOE assessments (Appendix D), and summary ROB assessments (Appendix C).

Described in detail previously, this trial randomized postmenopausal women with osteoporosis or low bone mass/osteopenia to placebo (n=46) versus one of seven denosumab regimens (n=319) for 2 years. After 2 years, those initially assigned placebo were assigned to continue placebo for another 2 years. Those initially assigned to denosumab continued denosumab for 2 years (n=231), switched to placebo for 2 years (n=47), or switched to placebo for 1 year before restarting denosumab for 1 year (n=41).

Outcomes

Incident Clinical Fractures
Incident fractures were not reported as primary or secondary outcomes, but only as adverse events. Authors reported that during 4 years of follow-up, 10.5 percent of 314 women in the denosumab group had an incident clinical fracture (the discrepancy of this denominator from the 319 the study reported were initially assigned denosumab was not explained). They stated that no increase in fracture incidence was observed among the small number of patients who discontinued denosumab treatment, but reported no numerical data comparing results between women assigned to continue denosumab and those assigned to discontinue it.

Incident Radiographic Vertebral Fractures
No information was reported about risk of incident radiographic vertebral fractures.

Change in BMD
Among women assigned 4 years of denosumab, LS-BMD increased 9.4 to 11.8 percent from baseline and TH-BMD increased 4.0 to 6.1 percent from baseline. In women who received
denosumab for 2 years, followed by placebo for 1 year and then 1 year of denosumab, LS-BMD increased 9.0 percent between baseline and year 4, while TH-BMD increased 3.9 percent during this interval. In women randomized to denosumab for 2 years followed by 2 years of placebo, authors reported no numerical data, but stated that BMD after 4 years returned to near its pre-denosumab treatment baseline but was greater than that in women assigned 4 years of placebo.

**Variation in Effects of Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics**

We identified no eligible trials that compared denosumab continuation versus discontinuation (placebo drug holiday) and assessed whether the risk of incident fracture varied as a function of patient, bone, or drug characteristics.
Other Drugs Investigated for Review

**Raloxifene**
We identified no studies that compared raloxifene continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between raloxifene continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Ibandronate**
We identified no studies that compared ibandronate continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between ibandronate continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Risedronate**
We identified no studies that compared risedronate continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between risedronate continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Teriparatide**
We identified no studies that compared teriparatide continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between teriparatide continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Abaloparatide**
We identified no studies that compared abaloparatide continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between abaloparatide continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Hormone Therapy**
We identified no studies that compared hormone therapy continuation with discontinuation and reported on risk of incident fracture; or whether differences in risk of incident fractures between hormone therapy continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.
Chapter 7. Harms of Drug Holidays

Chapter 7 discusses Key Questions 7 and 8. Key Question 7 addresses the effect of osteoporosis drug treatment continuation versus discontinuation (placebo drug holiday) on risk of harms. Key Question 8 examines whether the effect on risk of harms varies by patient, bone or drug characteristics.

Alendronate

Key Points

Key Question 7
- In postmenopausal women who had received 5 years of alendronate for FN-BMD T-score ≤ -1.6:
  - Between alendronate continuation for 3 more years or 5 more years versus discontinuation (placebo drug holiday), evidence was insufficient to draw conclusions about differences in risk of serious adverse events.
  - Between alendronate continuation for 5 more years versus discontinuation, evidence was insufficient to draw conclusions about differences in risk of ONJ, AFF, or subtrochanteric or femoral shaft fracture.
- In postmenopausal women who had received 5 years of alendronate for LS-BMD T-score ≤ -2.5:
  - Between alendronate continuation for 2 years versus discontinuation, there was no difference in risk for risk of serious adverse events (low SOE).
  - Between alendronate continuation for 5 years versus alendronate continuation for 3 more years followed by discontinuation, evidence was insufficient to draw conclusions about differences in risk of serious adverse events.
  - Between alendronate continuation up to 5 years versus discontinuation, evidence was insufficient to draw conclusions about differences in risk of ONJ, AFF, or subtrochanteric or femoral shaft fracture.

Key Question 8
- We identified no evidence about whether the effect of alendronate continuation versus discontinuation on harms varied as a function of patient, bone or drug characteristics.

Eligible Studies
Six eligible publications of three unique studies compared alendronate continuation versus discontinuation and reported on risk of harms. We rated three of these publications as having low ROB\textsuperscript{37, 62, 86} and three as having medium ROB\textsuperscript{63, 64, 93} We extracted detailed information from these studies in evidence tables, SOE summary tables (Appendix D) and ROB summary tables (Appendix C).

As described in detail above, these studies enrolled postmenopausal women who had osteoporosis\textsuperscript{63, 64} or a mix of osteoporosis and low bone mass\textsuperscript{37, 62, 67} when they had been randomly assigned to the alendronate arm of a placebo controlled RCT 5 years earlier. After 5 years of alendronate, one study randomized 1099 women to continuation versus
discontinuation\textsuperscript{37, 62, 67} and the other study nonrandomly assigned two of three alendronate groups to continued alendronate and the third alendronate group to placebo.\textsuperscript{63, 64}

**Outcomes**

**Serious Adverse Events**

In FLEX, authors reported that during 5 years of follow-up there were no significant differences between the alendronate discontinuation and continuation groups in risk of serious adverse events or mortality, but no numerical data were provided.\textsuperscript{62} In an interim report, after 3 years of FLEX follow-up, serious adverse events data were not reported, and there was no between-group difference in risk of mortality or adverse event causing hospitalization.\textsuperscript{86} In the earlier dose-ranging study, during the year 6-7 extension, risks of any serious adverse event appeared similar between continuation and discontinuation groups (11.9% vs. 11.3%; RR 1.05 [95% CI 0.57, 1.96]) (low SOE).\textsuperscript{64} Evidence was insufficient about whether risk of serious adverse events differed between alendronate continuation and discontinuation groups during the year 8-10 extension (RR 1.21 [95% CI 0.75, 1.96]).\textsuperscript{63}

**Upper GI Adverse Events**

Participants randomized to alendronate continuation for 5 years versus discontinuation in FLEX did not differ in risk of upper GI tract adverse events, but again no numerical data were provided.\textsuperscript{62} However, in results reported after 3 years of FLEX follow-up, the discontinuation group had a higher risk of any upper GI event (36% vs. 30%, p=0.04).\textsuperscript{86} In the earlier dose-ranging study, risk of any upper GI event appeared similar between continuation and discontinuation groups during the year 6-7 extension (15.9% and 17.2% in the two continuation groups vs. 18.3% in the discontinuation group).\textsuperscript{64} During the year 8-10 extension of the dose ranging study, authors again reported that risk of upper GI events did not differ between treatment groups (24.1%, 14.1% and 27.9% in discontinuation, alendronate 5 mg/day and alendronate 10 mg/day groups, respectively).\textsuperscript{65}

**Subtrochanteric/Femoral Shaft Fracture**

In a secondary analysis of FLEX data, investigators reviewed radiology reports and medical records for all hip and femur fractures to identify subtrochanteric and femoral shaft fractures. Pathologic, periprosthetic, and high-energy trauma fractures were excluded. Radiographs were rarely available to review for atypical features, though when atypical features were described in the report, these were recorded. Two women in the alendronate continuation group and one in the discontinuation group had a subtrochanteric or femoral shaft fracture (0.030% vs. 0.023%; HR 1.33 [95% CI 0.12, 14.67] (insufficient SOE).\textsuperscript{37}

A retrospective cohort study including women at least 45 years of age and with ≥3 years of bisphosphonate use, compared the incidence of osteoporosis-related fractures in those who had discontinued use of bisphosphonates for ≥1 year (drug holiday group) to those who continued bisphosphonate use.\textsuperscript{93} Alendronate accounted for 99 percent of the bisphosphonate use among the cohort. Continual use was defined as persistent (use with ≥50% adherence) or non-persistent (use with <50% adherence). Fractures were identified using ICD-9 diagnosis codes only, with no documentation that radiographs were reviewed for features of atypical femoral fractures. Incidence of subtrochanteric/femoral shaft fractures was rare, with three (0.03%) in the drug
holiday group compared with 44 (0.15%) in the combined persistent and non-persistent continual use groups (insufficient SOE).

**Osteonecrosis of the Jaw**
In FLEX, authors reported that there were no cases of ONJ.62

**Atrial Fibrillation**
No eligible publication reported information about incidence of atrial fibrillation.

**Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics**
We identified no eligible studies of alendronate continuation versus discontinuation (drug holiday) that assessed whether the risk of harms vary as a function of patient, bone or drug characteristics.
Zoledronic Acid (Zolendronate)

Key Points

Key Question 7

- In postmenopausal women with osteopenia:
  - Between zoledronic acid for 2 years versus treatment for 1 year followed by placebo for 1 year, there was no difference in risk of serious adverse events (moderate SOE).
- In postmenopausal women with osteoporosis:
  - Between zoledronic acid for 6 years versus treatment for 3 years followed by placebo for 3 years, there was no difference in risk of serious adverse events (moderate SOE).
  - Between zoledronic acid for 9 years versus treatment for 6 years followed by placebo for 3 years, there was no difference in risk of serious adverse events (low SOE).
- Evidence was insufficient to draw conclusions about differences in risk of AFF or ONJ between zoledronic acid continuation and discontinuation.

Key Question 8

- We identified no evidence about whether differences in risk of harms between zoledronic acid continuation and discontinuation vary as a function of patient, bone or drug characteristics.

Eligible Studies

Three eligible publications of two unique RCTs and no observational studies compared continuation of zoledronic acid treatment with discontinuation (placebo drug holiday) and reported on risk of harms. All were rated as having low ROB and were described in detail previously. Appendix D provides detailed evidence tables and SOE tables; and Appendix C provides summary ROB assessments for these studies.

Outcomes

Serious Adverse Events

In the McClung trial, risk of harms was not separately reported for the period during the second year of the trial, during which the groups were on different treatments, but only for the 2-year trial period as a whole. During the entire 2-year period, risk of serious adverse events was not significantly different between participants assigned to zoledronic acid continuation versus discontinuation (9.4% vs. 10.6%; RR 0.91 [95% CI 0.50, 1.67]) (moderate SOE). One death due to sepsis was reported in the continuation group. In the HORIZON trial extensions, there also was no difference in risk of serious adverse events between the zoledronic acid continuation and discontinuation groups, neither during years 4 to 6 (31% vs. 27%; RR 1.14 [95% CI 0.96, 1.36]) (moderate SOE) nor years 7 to 9 (26% vs. 30%; RR 0.86 [95% CI 0.54, 1.36]) (low SOE).
Atypical Femoral Fractures

Both HORIZON extension studies reported that no cases of AFF occurred in either treatment group during follow-up,\textsuperscript{65,66} with documentation that radiographs were reviewed for possible AFF features in just one of these reports\textsuperscript{66} (insufficient SOE for both comparisons).

Osteonecrosis of the Jaw

The McClung trial reported that no cases of ONJ occurred in either treatment group (insufficient SOE).\textsuperscript{68} In the HORIZON extension studies, one case of ONJ was reported in the continuation group and none in the discontinuation group during years 4 to 6 (insufficient SOE),\textsuperscript{65} and no cases of ONJ occurred in either group during years 7 to 9 (insufficient SOE).\textsuperscript{66}

Atrial Fibrillation

The McClung trial reported that no cases of atrial fibrillation occurred in either treatment group. In the HORIZON extension studies, incidence of atrial fibrillation was not statistically significantly different between the continuation and discontinuation groups during years 4 to 6 (3.4\% vs. 2.1\%, \(p=0.17\))\textsuperscript{65} or between years 7 to 9 (5.4\% vs. 1.1\%, \(p=0.11\))\textsuperscript{66}

Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We identified no eligible studies that compared zoledronic acid continuation with discontinuation (placebo drug holiday) and assessed whether risk of harms varied as a function of patient, bone, or drug characteristics.
Denosumab

Key Points

Key Question 7
- In postmenopausal women with osteoporosis or osteopenia, between denosumab continuation and discontinuation (placebo drug holiday), evidence was insufficient to draw conclusions about differences in risk of harms.

Key Question 8
- We identified no evidence about whether differences in risk of harms between denosumab continuation and discontinuation vary as a function of patient, bone or drug characteristics.

Eligible Studies
- One eligible publication of one RCT compared denosumab continuation versus discontinuation (placebo drug holiday) and reported on risk of harms. Appendix D provides detailed evidence tables and strength of evidence tables; and Appendix C provides summary ROB assessments for this study.
  - Described in detail previously, this trial randomized postmenopausal women with osteoporosis or low bone mass/osteopenia to placebo (n=46) versus one of seven denosumab regimens (n=319) for 2 years. After 2 years, those initially assigned placebo were assigned placebo for another 2 years. Those initially assigned denosumab were assigned to continue denosumab for 2 years (n=231), switch to placebo for 2 years (n=47), or switch to placebo for 1 year before restarting denosumab for 1 year (n=41).

Outcomes

Serious Adverse Events
- Through 4 years, risk of any serious adverse event was not significantly different between women initially assigned denosumab versus placebo (17.8% vs. 10.9%; RR 1.64 [95% CI 0.69, 3.88]) (low SOE). However, this result was difficult to interpret because the denosumab group included all women initially assigned to denosumab. Results were not separated between women assigned to denosumab for 4 years, women assigned to denosumab for 2 years followed by placebo for 2 years, or those assigned to denosumab for 2 years followed by placebo for 1 year and then denosumab again for 1 year.

Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics
- We identified no eligible studies that compared denosumab continuation with discontinuation (placebo drug holiday) and assessed whether risk of harms vary as a function of patient, bone or drug characteristics.
Other Drugs Investigated for Review

**Raloxifene**

We identified no studies that compared raloxifene continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between raloxifene continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Ibandronate**

We identified no studies that compared ibandronate continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between ibandronate continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Risedronate**

We identified no studies that compared risedronate continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between risedronate continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Teriparatide**

We identified no studies that compared teriparatide continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between teriparatide continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Abaloparatide**

We identified no studies that compared abaloparatide continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between abaloparatide continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.

**Hormone Therapy**

We identified no studies that compared hormone therapy continuation with discontinuation and reported on risk of harms; or whether differences in risk of harms between hormone therapy continuation and discontinuation (placebo drug holiday) vary as a function of patient, bone, or drug characteristics.
Chapter 8. Discussion

Overview

We conducted a systematic review to evaluate the evidence on (1) the efficacy and harms of long-term osteoporosis drug treatment, (2) the efficacy and harms of continuing versus discontinuing osteoporosis drug treatment (i.e., drug holidays), and (3) whether the effects of these interventions vary as a function of patient, bone, or osteoporosis drug characteristics.

We defined long-term osteoporosis drug treatment as exceeding three years in duration. We limited the review of efficacy to RCTs and CCTs, but broadened the evidence base we considered for harms to also include controlled observational studies. We analyzed data only from studies judged to have low or medium ROB. We graded SOE for the most clinically important, patient-centered outcomes of incident clinical fractures (any, nonvertebral, hip, major osteoporotic fracture, vertebral), incident radiographic vertebral fractures, serious adverse events (SAEs), AFF, and ONJ.

Key Findings and Strength of Evidence

Key Question 1: Efficacy of Long-term Osteoporosis Drug Treatment

For this review, efficacy of long-term (>3 years) osteoporosis drug treatment versus placebo, no treatment or another active agent was considered separately from efficacy of long-term osteoporosis drug treatment versus shorter-term treatment with the same agent. We found few eligible placebo-controlled RCTs. All participants in these studies were postmenopausal women. Collectively, these studies showed that in women with osteoporosis, compared to placebo there was an approximately 35-45 percent relative reduction (2-5% absolute reduction) in risk of incident radiographic vertebral fractures (high SOE for 4 years of alendronate or raloxifene, insufficient SOE for other treatments). In addition, raloxifene for four years was associated with an approximately 40 percent relative reduction (2% absolute reduction) in risk of incident clinical vertebral fractures (moderate SOE). However, there was no reduction in several categories of incident nonvertebral fractures (low to moderate SOE for 4 years of alendronate and raloxifene, insufficient SOE for other treatments).

Key Question 2: Variation in Efficacy of Long-term Osteoporosis Drug Treatment as a Function of Patient, Bone or Osteoporosis Drug Treatment Characteristics

Analyses found that efficacy of long-term alendronate varied as a function of baseline BMD. Between long-term alendronate and placebo, there was a significant 35 percent relative reduction (6-7% absolute reduction) in risk of incident clinical fractures among women with FN-BMD <-2.5 (osteoporosis), but not in women with higher BMD (osteopenia). Similarly, women with FN-BMD ≤-2.5 (osteoporosis) had a 55 percent relative reduction (1% absolute reduction) in risk of incident hip fracture with long-term alendronate versus placebo, but women with higher BMD had no decreased risk. For incident radiographic vertebral fractures, the pattern was a bit different. Women with osteoporosis had a significant 50 percent relative reduction (3% absolute reduction) in risk of incident radiographic vertebral fractures (moderate SOE). However, there was no reduction in several categories of incident nonvertebral fractures (low to moderate SOE for 4 years of alendronate and raloxifene, insufficient SOE for other treatments).
were wider, and results were not statistically significant. Women with higher BMD appeared to have no reduction in risk of these fractures. Neither past nonvertebral fracture, FRAX 10-year MOF probability calculated with BMD, nor pretreatment levels of bone turnover markers modified the effect of long-term alendronate versus placebo on risk of any incident fracture outcome. Neither age, baseline BMD nor baseline radiographic vertebral fracture modified the effect of long-term raloxifene versus placebo on risk of incident fractures. Strength of evidence was not assessed for effect modifiers. We found no other evidence for effect modification of any other drugs included in this review.

**Key Question 3: Harms of Long-term Osteoporosis Drug Treatment**

Long-term treatment with raloxifene versus placebo did not increase risk of serious adverse events (low SOE). This outcome was not specifically reported for long-term alendronate. Results about risk of harms with long-term denosumab were difficult to interpret because participants assigned long-term denosumab were pooled with those assigned denosumab discontinuation. In a trial population selected for low risk of upper GI adverse events, long-term alendronate did not increase risk of these outcomes versus placebo. Large magnitude risk estimates suggest that compared to placebo, long-term raloxifene may increase risk of DVT (2.8 to 3.1-fold relative increase and 0.5% absolute risk increase) and PE (2.8 to 4.5-fold relative increase and 0.2% absolute risk increase). Controlled observational studies suggested that alendronate and bisphosphonates as a class increase risk of subtrochanteric and femoral shaft fractures (low SOE), radiologically confirmed AFF (low SOE), and ONJ (insufficient to low SOE). While magnitudes of relative risk often were high, increases in absolute risk were low and results were highly variable between studies. We found no evidence from eligible studies about risk of these harms with long-term denosumab.

**Key Question 4: Variation in Harms of Long-term Osteoporosis Drug Treatment as a Function of Patient, Bone or Osteoporosis Drug Treatment Characteristics**

We found little evidence about factors that modify long-term osteoporosis drug treatment risks. One study was inconclusive about whether relative risk for AFF associated with bisphosphonate use increased with age. Several observational studies suggested that >5 years of bisphosphonate use increased risk of subtrochanteric or femoral shaft fractures, or radiologically confirmed AFF, more versus a control exposure than did 3 to 5 years of use. For long-term raloxifene versus placebo, one study reported that risk of DVT and PE did not vary as a function of baseline cardiovascular risk and another that risk of incident stroke was lower with raloxifene versus placebo in the subgroup of women at increased baseline cardiovascular risk. Strength of evidence was not assessed for effect modifiers. We found no other evidence for effect modification of any other drugs included in this review.

**Key Question 5: Effects of Osteoporosis Drug Treatment Holidays**

In postmenopausal women with osteopenia or osteoporosis who previously received 5 years of alendronate, one of two trials reported that continued alendronate for up to an additional 5 years versus discontinuation reduced risk of incident clinical vertebral fractures (approximately 55% relative reduction and 3% absolute reduction) (low SOE). However, that trial reported no difference in risk of incident radiographic vertebral fractures (low SOE). In one trial conducted in postmenopausal women with osteopenia, evidence was insufficient about whether zoledronic...
acid for two years reduced risk of incident fractures compared to a drug holiday after one year of zoledronic acid. A trial comparing six years of zoledronic acid versus three years followed by a three-year placebo drug holiday reported a reduction in risk of radiographic vertebral fractures (approximately 50% relative reduction and 3% absolute reduction) (moderate SOE), but no difference in any incident clinical fracture endpoint (insufficient to low SOE). In an extension study that compared 9 years of zoledronic acid versus 6 years of zoledronic acid followed by 3 years of placebo drug holiday, there were too few incident fractures to draw any conclusions about differences in risk between treatment groups (insufficient SOE).

Key Question 6: Variation in Effect of Osteoporosis Drug Treatment Holidays as a Function of Patient, Bone or Osteoporosis Drug Treatment Characteristics
Neither baseline FN-BMD nor baseline radiographic vertebral fracture status modified the risk of incident nonvertebral fracture or incident clinical vertebral fracture between alendronate continuation versus discontinuation. Strength of evidence was not assessed for effect modifiers. We found no evidence about possible effect modifiers of continuing any other osteoporosis drug treatment versus discontinuation on risk of incident fracture.

Key Question 7: Harms During Osteoporosis Drug Treatment Holidays
Trials of alendronate and zoledronic acid continuation versus discontinuation found no difference in risk of serious adverse events (insufficient to low SOE). Atrial fibrillation appeared more frequent with zoledronic acid continuation versus discontinuation in both extension studies, but events were uncommon and possible differences between treatment groups were not statistically significant. Too few cases of ONJ occurred in these trials to draw conclusions and studies either provided no information about AFF or reported too few events to draw conclusions about differences in risk of these outcomes between treatment groups (insufficient SOE).

Key Question 8: Variation in Harms During Osteoporosis Drug Treatment Holidays as a Function of Patient, Bone or Osteoporosis Drug Treatment Characteristics
We found no evidence about whether patient, bone or drug characteristics modify the risk of harms between continuation of any osteoporosis drug treatment and discontinuation (placebo drug holiday).

Applicability
It was the aim of the review to apply to the general population of adults aged ≥50 years with osteoporosis or osteopenia/low bone mass. Studies that focused on adults <50 years of age or on individuals with known secondary causes of osteoporosis (e.g., exogenous glucocorticoids) were excluded. Studies that focused on patients with cancer metastatic to bone or acute fracture healing also were excluded. Review findings may not apply to these populations.

Trial inclusion criteria also limited the generalizability of the review findings. All trials were restricted to postmenopausal women, only a small proportion of participants were among the oldest old (e.g., ≥80 years), and participants generally had low levels of comorbidity. Therefore, the applicability of review findings to younger women, men, the elderly, and patients with
multiple comorbid conditions as may be seen in typical primary care settings is unknown. The evidence on the efficacy of long-term alendronate, denosumab, raloxifene, and estrogen versus placebo on risk of incident fracture came from trials conducted entirely in the U.S., and nearly all data on the efficacy of long-term estrogen/progestin versus placebo came from one U.S. trial, so applicability to other geographic settings is uncertain. Trial populations often were selected to be low-risk for specific treatment harms (e.g., upper GI adverse events for alendronate and venous thromboembolism for raloxifene), so the risk of harms reported in trials may be lower than what would be expected in clinical populations typically treated with these agents. Because most extension studies in the review enrolled only a minority of the individuals who participated in the preceding study phase, their participants may be less representative of the general osteoporosis population than were original trial enrollees.

Findings in Relation to What is Already Known

Efficacy of Short-term Osteoporosis Drug Treatment

As reported in a 2014 AHRQ review69 and related 2017 American College of Physicians Clinical Practice Guideline on pharmacologic treatment of osteoporosis to prevent fractures,79 there is strong SOE from short-term RCTs (18 to 36 months) in postmenopausal women with osteoporosis that alendronate, risedronate, ibandronate, zoledronic acid, denosumab, teriparatide and raloxifene lower risk of incident vertebral fractures. All these agents except ibandronate and raloxifene also lower short-term risk of incident nonvertebral fractures (strong SOE), while alendronate, zoledronic acid and denosumab also lower short-term risk of incident hip fractures. Trials have not shown reductions in risk of clinical fractures with short-term treatment of patients without osteoporosis, even in those with low bone mass/osteopenia. RCT evidence on short-term fracture protection in men comes primarily from a single 2-year trial of zoledronic acid in men with osteoporosis that showed a reduction in risk of incident radiologic vertebral fractures.

The current review did not address efficacy and harms of short-term osteoporosis drug treatment or evaluate possible effect modifiers of these treatment outcomes.

Efficacy of Long-term Osteoporosis Drug Treatment

Most patients diagnosed with osteoporosis (for whom short-term osteoporosis drug trial results are applicable) remain at elevated risk for fractures at the completion of three years of treatment, raising the question as to how long patients should be treated with drugs to prevent fracture.

The last AHRQ review did not highlight the few placebo-controlled trials that exceeded three years in duration, presumably in part because it did not use three years duration as a threshold to define short-term versus long-term treatment, as we did in the current review. Nevertheless, the findings of these >3-year placebo-controlled trials were limited. A 5-year risedronate trial was an extension of an earlier three-year trial that had such high loss to follow-up that results were rated high ROB and it was not analyzed in the present review. A 4-year raloxifene trial showed a reduction in clinical vertebral fractures and radiographic vertebral fractures but no reduction in nonvertebral fractures. A raloxifene versus placebo extension phase to 8 years showed no fracture benefit with longer treatment. In postmenopausal women with a past hysterectomy and unknown BMD, in both the overall group and the subgroup with prior clinical fracture, 7 years of
estrogen versus placebo significantly reduced risk of incident clinical fracture and incident hip fracture. In postmenopausal women with an intact uterus and unknown BMD, in both the overall group and the subgroup with prior clinical fracture, 5.6 years of estrogen/progestin versus placebo significantly reduced risk of incident clinical fracture but not incident hip fracture. A 4-year denosumab trial pooled results from denosumab continuation and discontinuation arms to compare to placebo, so effects of continuous denosumab versus placebo could not be determined. A 4-year alendronate trial, conducted in postmenopausal women without baseline radiographic vertebral fracture and with osteopenia or osteoporosis by BMD, showed a reduction in incident radiographic vertebral fractures, but not in incident clinical fractures. In analyses planned before study unblinding, the subgroup with FN-BMD T-score <-2.5 had a lower risk of incident clinical fractures with treatment.

Collectively, these trials suggest that for postmenopausal women with osteoporosis, alendronate and raloxifene each reduce risk of incident vertebral fracture through four years versus placebo, that alendronate for four years also reduces risk of incident nonvertebral fractures compared to placebo, and that estrogen and estrogen/progestin reduce risk of incident clinical fracture and estrogen also reduces risk of hip fracture. Further, other than for estrogen and estrogen/progestin, there is almost no evidence about treatment beyond four years. Our review also found a complete absence of placebo-controlled trial data on the efficacy of long-term treatment with any sequential osteoporosis drug treatment regimen (e.g., anabolic followed by anti-resorptive) for reducing risk of incident fracture.

Harms of Long-term Osteoporosis Drug Treatment

The current review found few trials that provided information about risk of harms associated with long-term treatment with osteoporosis drugs compared to placebo, particularly compared to the number of trials reporting harms in the 2014 AHRQ review. While the prior AHRQ review reported a significant 10 percent relative increase versus placebo in risk of mild upper GI events, the current review of long-term treatment found no difference in risk for any upper GI events. The prior review reported a statistically significant 1.6-fold increase in thromboembolic events with raloxifene compared to placebo, but the current review, deriving long-term data from just the MORE trial and follow-up CORE study found a borderline significant 3 to 4-fold increase in risk of DVT and PE. A 2013 meta-analysis estimated that bisphosphonate exposure was associated with a significant increase in risk of subtrochanteric, femoral shaft and AFF considered as a pooled outcome (RR 1.70 [95% CI 1.22, 2.37]). In stratified analysis, that meta-analysis reported that risk in studies that used the ASBMR criteria to define AFF was RR 11.78 (95% CI 0.39, 359.69) compared to RR 1.62 (95% CI 1.18, 2.22) for studies that used the outcome of subtrochanteric or femoral shaft fractures defined primarily by diagnosis codes. In the current review, we noted that in addition to important differences in how the studies defined and identified these fractures, they differed in other ways that might have affected their risk estimates (e.g., selection of exposure control, selection of fracture control, model covariables). Therefore, we elected to not mathematically pool the studies, but to separate them into two categories (AFF with ASBMR radiographic criteria vs. subtrochanteric/femoral shaft fractures without confirmed radiographic AFF features) and then describe them qualitatively. Prior reviews have reported a more than 2-fold increased risk of ONJ with bisphosphonates used for osteoporosis, but also important differences in ONJ case definitions. We identified these same challenges when limiting our review to studies reporting long-term osteoporosis drug treatment.
Effect of Osteoporosis Drug Treatment Holidays

The results of the largest trials of osteoporosis drug treatment continuation versus discontinuation, detailed in the current review, also were described in the 2014 AHRQ review. In one trial, postmenopausal women randomized to alendronate for 10 years versus 5 years had a lower risk of clinical vertebral fractures but no reduction in risk of radiographic vertebral fractures or in any other type of clinical fractures. In a second trial, postmenopausal women randomized to continue zoledronic acid for a total of 6 years versus 3 years had a lower risk of radiographic vertebral fractures but no difference in risk of any type of clinical fracture. An older trial of long-term alendronate continuation versus discontinuation reported no reduction in incident clinical fracture or incident radiographic vertebral fractures with longer versus shorter alendronate treatment.

The explanation for the variability in risk of incident vertebral fractures between these long-term bisphosphonate trials is unclear. The studies used comparable definitions and ascertainment methods. The magnitude of relative risk reductions in clinical vertebral fracture with continued alendronate in the FLEX trial and in risk of incident radiographic vertebral fracture with continued zoledronic acid in the HORIZON extension are large. However, because they are based on a small absolute number of vertebral fractures and results aren’t corroborated by all other vertebral fracture results within the same or other trials, our confidence in these findings is limited. The current review did not identify additional trials that clarified this uncertainty. More studies are needed to clarify the effects of long-term bisphosphonate treatment on risk of incident clinical and radiographic vertebral fractures.

Variation in Osteoporosis Drug Treatment Outcomes

The current review highlights the relative paucity of information available to help target which patients are most and least likely to experience reductions in risk of incident fractures and/or increases in risk of harms from long-term osteoporosis drug treatment or continuing versus discontinuation of osteoporosis drug treatment. Because so few trials examined the efficacy/effects and harms of long-term or continuing osteoporosis drug treatment, opportunities to investigate factors that predict long-term trial treatment outcomes have been limited. Opportunities to independently validate them in a different trial dataset have been rare. Validation is important for guiding clinical decision making because nearly all subgroup analyses and tests for interaction in these studies were conducted post hoc, often when overall results were negative, and therefore are prone to type I error.

Because of sparse data on fracture prevention from long-term treatment trials and the high cost of trials adequately powered to evaluate incident fracture outcomes, many investigators have sought to identify appropriate surrogate endpoints. Some have advocated using change in BMD during treatment as an important clinical outcome. However, we found no trials that reported on whether differences in risk of incident fractures with long-term treatment varied as a function of changes in BMD during that treatment. Similarly, we found no trials that reported on whether differences in risk of incident fractures between osteoporosis drug continuation and discontinuation varied as a function of changes in BMD before or during the discontinuation period. Preliminary data from the Foundation for the National Institutes of Health (FNIH) Bone Quality project suggests that early treatment changes in BMD and, to a lesser extent, bone formation markers, may predict the anti-fracture efficacy of antiresorptive osteoporosis drugs.17.
We found no data about whether any patient, bone or drug characteristics modify the effect of drug holidays on risk of harms.

Limitations of the CER

Possible limitations of the review methodology include that information about possible effect modifiers of osteoporosis drug treatment outcomes was not always well described in study statistical analysis sections and was variably reported in tables, figures and narrative text. Therefore, it is possible that some information on effect modifiers that was relevant for the review was overlooked. For this review, we excluded non-English language studies based primarily on concerns about available resources. However, as the literature in this field is almost exclusively published in English-language, and we solicited recommendations for additional studies that might meet eligibility criteria from an NIH ODP Working Group, an NIH Content Area Expert Group, and a Technical Expert Panel, it is unlikely that important studies were missed. Because we extracted information on change in BMD only from eligible studies that also reported incident fracture outcomes, we likely missed some studies that reported change in BMD. This may have caused a selection bias for this outcome.

Strengths of the review methodology include its emphasis on patient-important efficacy outcomes, primarily incident clinical fractures and secondarily incident radiographic vertebral fractures and harms. Another strength is that SOE was assessed for patient-important outcomes of incident fracture and harms and not for change in BMD. A further strength is that this review required observational studies of AFF, ONJ or atrial fibrillation to have an exposure control group. Although uncontrolled studies might have provided information about incidence of these harms, they would not account for the nonzero risk of these harms in comparable populations not receiving these medications, and likely would have overestimated the risk attributable to these treatments.

Limitations of the Evidence Base

The first limitation of the evidence base was the small number of eligible, placebo-controlled long-term trials, including only one in which incident fracture was the primary outcome. There also were few eligible trials of osteoporosis drug continuation versus discontinuation, none of which designated incident fracture as the primary outcome. Consequently, for many incident clinical fracture outcomes, there were relatively few events, and risk estimates were imprecise, limiting strength of evidence. Harms reporting in published studies seemed inconsistent and incomplete despite their likely standard collection and reporting to the pharmaceutical manufacturers and the U.S. FDA. This raised concerns about possible reporting bias. Because there were not multiple long-term or drug holiday studies with comparable populations, interventions, and outcomes, consistency of findings could not be determined. The maximum duration of these trials was another limitation. Though three continuation versus discontinuation studies lasted 9-10 years, two had too few remaining participants by that time to draw meaningful conclusions about the risk of incident clinical fractures.

Another limitation is that studies used different definitions and methods for ascertaining outcomes. For example, incident clinical vertebral fractures in FIT and FLEX required initial recognition and imaging in the community that was then confirmed by study comparison of community and study vertebral radiographs. In contrast, in MORE and CORE, community recognition was not required and incident clinical vertebral fractures were identified when study participants reported back pain at scheduled study visits, which prompted investigators to order a
study vertebral radiograph. Outcome definitions for observational studies that evaluated risk of AFF, ONJ and atrial fibrillation were far more variable. For AFF, studies differed in whether cases were defined solely based on diagnosis codes, whether cases were excluded for trauma or cancer, and whether radiologic review for AFF features was required. These differences likely affected specificity. Moreover, to the extent individual osteoporosis drugs differ in their effects on the risk of femur fractures associated with excess trauma, cancer, and with and without radiologic AFF features, these differences probably also affected risk estimates. Other differences in study design may also affect risk estimates. Examples include differences in treatment control group (e.g., no bisphosphonate use, past bisphosphonate use, raloxifene or calcitonin use), in non-AFF comparison group (no hip fracture, femoral shaft fracture documented to not have radiologic AFF features), and in the covariates included in the adjusted statistical models. Since individual studies did not evaluate the impact of these different analytic approaches on results, and studies often differed in multiple ways, the differences in the results between these studies were difficult to interpret. These issues apply similarly to the outcome of ONJ.

The evidence base also was limited in that few studies evaluated possible effect modifiers on risks for incident fracture and harms with long-term osteoporosis drug treatment. Even when these were investigated, nearly all analyses were conducted post hoc, tests of interaction infrequently were reported, and results were not adjusted for testing of multiple outcomes, raising the likelihood of type 1 errors. We found virtually no data examining the effect of age on differences between treatments in risk for incident fractures and harms. This means that the literature did not provide evidence to inform clinical decisions about whether there is an optimal age to start osteoporosis drug treatment or to stop it.

Inconsistency of results reported by different publications from the same study was another limitation of the evidence base. For example, multiple MORE and CORE study publications reported modestly different risks of DVT and PE associated with long-term raloxifene that were difficult to reconcile, even after considering variation in their analytic cohorts and follow-up periods.

**Future Research**

Table 13 summarizes the areas needing future research based on the gaps identified in this review.

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gap</th>
<th>Future Research Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>All eligible RCTs of long-term osteoporosis drug treatment and of osteoporosis drug continuation versus discontinuation were conducted in postmenopausal women and most were conducted in the U.S. None included men and few were multinational.</td>
<td>Future RCTs of long-term osteoporosis drug treatment and of osteoporosis drug continuation versus discontinuation should include men and be conducted in a variety of geographic regions.</td>
</tr>
<tr>
<td>Published trials of long-term osteoporosis drug treatment and of osteoporosis drug continuation versus discontinuation have been conducted in populations that were generally healthy. Drug treatment benefits and harms in these populations may not be</td>
<td>Future trials of long-term osteoporosis drug treatment and of osteoporosis drug continuation versus discontinuation should be performed in more diverse populations, including those with multiple comorbidities.</td>
<td></td>
</tr>
<tr>
<td>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td><strong>KQ1:</strong> Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, what is the efficacy of long-term (&gt;3 years) osteoporosis drug therapy for reducing risk of incident fracture and on change in BMD?</td>
<td>Eligible long-term osteoporosis drug treatment trials have had relatively few fracture events at longer follow-up times, resulting in imprecise estimates of incident fracture risk. This limits confidence that reported estimates of risk represent true effects.</td>
<td>Future RCTs of long-term osteoporosis drug treatment should be designed with adequate power to compare risk of clinical fracture endpoints.</td>
</tr>
<tr>
<td></td>
<td>There were no eligible long-term placebo-controlled trials with adequate power to compare risk of clinical fracture endpoints for risedronate, ibandronate, denosumab, and teriparatide. There have been no long-term trials of sequential therapy of anabolic agents followed by antiresorptive agents compared to antiresorptive agents alone with adequate power to assess risk of clinical fracture endpoints.</td>
<td>Future RCTs of long-term osteoporosis drug treatment designed with adequate power to compare risk of clinical fracture endpoints should be performed for ibandronate, risedronate, denosumab, and teriparatide, as well as for anabolic therapy (e.g. teriparatide or abaloparatide) followed by antiresorptive therapy versus antiresorptive therapy alone.</td>
</tr>
<tr>
<td>Although future long-term placebo-controlled osteoporosis drug trials will be expensive and time consuming, adequate surrogates for incident fracture risks with these treatments are not established.</td>
<td>Analyses should clarify the extent to which individual and combinations of potential surrogate markers (e.g., changes in BMD, bone turnover markers, and/or QCT measures) account for differences in risk of incident fractures between long-term osteoporosis drug treatment and placebo. The Foundation for the NIH (FNIH) Bone Quality Project, a collaboration of Pharma, the federal government, academia, and an osteoporosis specialty society, is evaluating pooled patient-level data from 140,000 participants in osteoporosis drug trials to address these questions. A recently published FNIH meta-regression reported that early treatment changes in bone formation markers (but not bone resorption markers) explain a modest percentage of the effect of treatment on incident vertebral fractures, but not on nonvertebral fractures. Preliminary data presented only at scientific meetings suggests early changes in BMD explain a larger percentage of the treatment effect, and do so for the effect of treatment on both vertebral and nonvertebral fractures. Observational studies may be faster and cheaper for estimating fracture risk with long-term osteoporosis drug treatments, but are prone to bias. These biases may be reduced by use of methods to account for treatment selection, but confounding by indication cannot entirely be eliminated and these studies should be interpreted cautiously to avoid false inferences.</td>
<td></td>
</tr>
<tr>
<td>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Past placebo-controlled trials of long-term osteoporosis drug treatment that included different arms for different regimens of the osteoporosis drug, including continuous use, but did not report results separately for the different osteoporosis drug regimens did not allow direct comparisons of the effects of continuous long-term treatment of these agents versus placebo on risk of incident fracture.</td>
<td>Such past placebo-controlled trials of long-term osteoporosis drug treatment that included a continuous treatment arm should separately compare results for the different osteoporosis regimens to placebo.</td>
<td></td>
</tr>
<tr>
<td>Given strong evidence that several FDA approved medications reduce short-term risk (18 months to 3 years) of new vertebral and nonvertebral fractures in postmenopausal women with osteoporosis, it may not be ethical to randomize postmenopausal women with osteoporosis to placebo in future long-term fracture prevention trials.</td>
<td>Future placebo-controlled trials of long-term osteoporosis drug treatment may need to target patient populations different than those included in the short-term trials (e.g., not postmenopausal women) and/or who decline or have contraindications to medications with documented short-term nonvertebral fracture protection and/or allow participants in both treatment groups to also take other osteoporosis drugs during the trial.</td>
<td></td>
</tr>
<tr>
<td>KQ2: Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, does efficacy of long-term (&gt;3 years) osteoporosis drug therapy for reducing risk of incident fracture vary as a function of patient, bone, or osteoporosis drug characteristics?</td>
<td>Many long-term osteoporosis drug treatment trials were conducted in populations that were not diverse. Such study populations may include little variability to facilitate analyses regarding whether treatment results vary as a function of patient characteristics.</td>
<td>Future trials of long-term osteoporosis drug treatment should be performed in more diverse populations, including those with multiple comorbidities, not just to improve the applicability of the evidence base, but to facilitate analyses regarding whether treatment results vary as a function of patient characteristics.</td>
</tr>
<tr>
<td>Relatively few analyses have evaluated possible effect modification of patient, bone or osteoporosis drug characteristics on the efficacy of long-term osteoporosis drug treatment for preventing fractures.</td>
<td>Future studies should investigate the possible effect modification of characteristics that could help guide clinical decision making about which patients are most and least likely to experience reductions in fracture risk with long-term treatment.</td>
<td>Future studies should specify analysis plans to investigate possible effect modification a priori. Such analyses should be replicated in different trial datasets.</td>
</tr>
<tr>
<td>Most published analyses that have examined possible effect modifiers of long-term treatment efficacy have been post hoc. Statistical power to detect significant interactions and significant associations within subgroups have been limited. A few trials have examined many factors in multiple publications. There has been minimal independent evaluation of the same factors in different trials. This could increase risk of type I errors.</td>
<td>Future long-term osteoporosis drug trials should systematically collect and report Harms reporting in published long-term osteoporosis drug trials have not always</td>
<td>Future long-term osteoporosis drug trials should systematically collect and report Harms reporting in published long-term osteoporosis drug trials have not always</td>
</tr>
<tr>
<td>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, what is the risk of harms associated with long-term (&gt;3 years) osteoporosis drug therapy?</td>
<td>appeared systematic and sometimes has not been quantitative, raising the risk of reporting bias and preventing pooling of harms data across multiple trials. Rare harms potentially associated with antiresorptive agents sometimes have not been mentioned.</td>
<td>harms data, including both potential harms common across agents and required by the FDA (SAEs), drug-specific adverse effect, and rare harms of concern (e.g., AFF, ONJ). Results should be reported including numerators and denominators to facilitate interpretations about absolute risk and potential pooling across studies.</td>
</tr>
<tr>
<td></td>
<td>Published observational studies that evaluated risk of AFF with long-term osteoporosis drug treatment often have used definitions of AFF that may not distinguish AFF from typical femoral fractures or from femoral fractures caused by excess trauma, cancer or other pathological conditions, potentially biasing estimates of the risk of AFF with these long-term treatments.</td>
<td>Future observational studies investigating risk of AFF with long-term antiresorptive osteoporosis drugs should use standard definitions for AFF, including appropriate diagnostic coding, exclusions for excess trauma, cancer and other pathologic fractures, and radiologic confirmation that fractures have ASBMR task force criteria for AFF.</td>
</tr>
<tr>
<td></td>
<td>Published observational studies that evaluated risk of ONJ with long-term osteoporosis drug treatment often have used definitions of ONJ that may not distinguish ONJ from unrelated jaw conditions.</td>
<td>Future observational studies investigating risk of ONJ with long-term antiresorptive osteoporosis drugs should use standard definitions for ONJ, as per ASBMR Task Force criteria, including appropriate diagnostic coding, documented persistence of exposed bone in the maxillofacial region persistent for &gt;8 weeks after recognition by a health care provider, no history of craniofacial radiation.</td>
</tr>
<tr>
<td></td>
<td>Most observational studies that have evaluated the association of osteoporosis drug treatment with rare harms are case-control studies. Though this study design is efficient for estimating relative risks of the rare harms associated with these drug exposures, they cannot be used to estimate the absolute risk of these harms.</td>
<td>Future prospective cohort studies should be conducted to estimate and compare the risk of rare harms associated with antiresorptive osteoporosis drug treatment to best place these risks in the context of the effects of these treatments on other potentially beneficial and harmful outcomes.</td>
</tr>
<tr>
<td>KQ4: Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass†, does the risk of harms associated with long-term (&gt;3 years) osteoporosis drug therapy vary as a function of patient, bone, or osteoporosis drug characteristics?</td>
<td>There is minimal data from completed trials or observational studies examining possible effect modifiers of long-term osteoporosis drug treatment harms.</td>
<td>Ongoing or completed trials of long-term osteoporosis drug treatment versus placebo that are collecting or have collected harms should perform post hoc exploratory analyses of possible effect modifiers of harms to generate hypotheses to test in future studies, with possible coordination between different ongoing or completed studies to use common methods and/or to validate findings.</td>
</tr>
<tr>
<td>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>KQ5: Among men and postmenopausal women aged ≥50 years currently receiving</td>
<td>Past trials of osteoporosis drug treatment continuation versus discontinuation</td>
<td>Larger trials are needed that compare osteoporosis drug treatment continuation versus discontinuation and are powered to look at incident clinical fracture outcomes.</td>
</tr>
<tr>
<td>drug therapy started for osteoporosis* or osteopenia/low bone mass† to</td>
<td>have been too small to be adequately powered to evaluate the effects on</td>
<td>Past or ongoing trials (RCTs or CCTs) that assigned participants to different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment should report their incident fracture data in a manner which allows direct comparison of these fracture outcomes between these different treatment groups, including comparison with a continuous osteoporosis drug treatment if it is available. New trials should be conducted that assign participants (RCTs preferable, but CCTs also would be informative) to different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment and compare risk of incident fractures between treatment groups. The trials also should include an arm of continuous osteoporosis drug treatment (i.e., drug holiday duration of zero). RCTs are preferable, but CCTs also would be informative.</td>
</tr>
<tr>
<td>prevent fracture, what is the effect of osteoporosis drug treatment holidays</td>
<td>incident fracture risk and improving BMD?</td>
<td>No published trials have directly compared different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment and reported on differences in incident fracture risk between these groups.</td>
</tr>
<tr>
<td>on incident fracture risk and improving BMD?</td>
<td></td>
<td>Larger trials are needed that compare osteoporosis drug treatment continuation versus discontinuation and are powered to look at incident clinical fracture outcomes and evaluate whether differences in incident fracture risk vary as a function of a priori identified patient, bone or osteoporosis drug characteristics.</td>
</tr>
<tr>
<td>KQ6: Among men and postmenopausal women aged ≥50 years currently receiving</td>
<td>Past trials of osteoporosis drug treatment continuation versus discontinuation</td>
<td>New trials that compare osteoporosis drug continuation versus discontinuation (as detailed in KQ5 section) and systematically collect incident fracture outcomes, should plan a priori to examine possible effect modifiers of incident fracture risk. Factors to explore as possible effect modifiers include age, BMD, before and during the drug holiday, and levels of bone turnover markers before and during the drug holiday.</td>
</tr>
<tr>
<td>drug therapy started for osteoporosis* or osteopenia/low bone mass† to</td>
<td>have been too small to be adequately powered to evaluate for possible effect</td>
<td>No published trials have directly compared different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment, reported on differences in incident fracture risk between these groups, and evaluated whether differences in incident fracture risk vary as a function of patient, bone or osteoporosis drug characteristics.</td>
</tr>
<tr>
<td>prevent fracture, does the effect of osteoporosis drug treatment holidays</td>
<td>modification of treatment effects on risk of incident clinical fracture</td>
<td>New trials should be conducted that assign participants (RCTs preferable, but CCTs also would be informative) to different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment, and systematically collect and report risk of harms between treatment groups, including both common and rare harms.</td>
</tr>
<tr>
<td>on incident fracture risk vary as a function of patient, bone or osteoporosis</td>
<td>outcomes.</td>
<td></td>
</tr>
<tr>
<td>drug characteristics?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KQ7: Among men and postmenopausal women aged ≥50 years currently receiving</td>
<td>No published trials have directly compared different durations of osteoporosis</td>
<td>New trials should be conducted that assign participants (RCTs preferable, but CCTs also would be informative) to different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment, and systematically collect and report risk of harms between treatment groups, including both common and rare harms.</td>
</tr>
<tr>
<td>drug therapy started for osteoporosis* or osteopenia/low bone mass† to</td>
<td>drug treatment holidays with or without restarting osteoporosis drug</td>
<td>No published trials have directly compared different durations of osteoporosis drug treatment holidays with or without restarting osteoporosis drug treatment and reported on differences in risk of harms between these groups.</td>
</tr>
<tr>
<td>prevent fracture, what is the risk of harms</td>
<td>treatment and reported on differences in risk of harms between these groups.</td>
<td></td>
</tr>
<tr>
<td>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>of osteoporosis drug treatment holidays?</td>
<td>We found no evidence from eligible studies about whether patient, bone or osteoporosis drug characteristics modify risk for harms associated with osteoporosis drug continuation versus discontinuation.</td>
<td>New trials that compare osteoporosis drug continuation versus discontinuation (as detailed in KQ7 section) and systematically collect harms outcomes, should plan a priori to examine possible effect modifiers of harms outcomes.</td>
</tr>
<tr>
<td>KQ8: Among men and postmenopausal women aged ≥50 years currently receiving drug therapy started for osteoporosis or osteopenia/low bone mass† to prevent fracture, does risk of harms associated with osteoporosis drug treatment holidays vary as a function of patient, bone, or osteoporosis drug characteristics?</td>
<td>Ongoing or completed trials of osteoporosis drug continuation versus discontinuation that are collecting or already collected harms did not plan a priori to evaluate possible effect modification of treatment harms.</td>
<td>Ongoing or completed trials of osteoporosis drug continuation versus discontinuation that are collecting or have collected harms should perform post hoc exploratory analyses of possible effect modifiers of harms to generate hypotheses to test in future studies, with possible coordination between different ongoing or completed studies to use common methods and/or to validate findings.</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; ASBMR=American Society of Bone and Mineral Research; BMD=bone mineral density; CCT=controlled clinical trials; FDA=U.S. Food and Drug Administration; HRpQCT=high-resolution peripheral quantitative computed tomography; ONJ=osteonecrosis of the jaw; RCT=randomized clinical trials; SAE=serious adverse event.

**Conclusions**

Few osteoporosis treatment drugs have been shown to reduce risk of clinical fractures with long-term treatment. In postmenopausal women with osteoporosis and no prior osteoporosis drug treatment, both alendronate and raloxifene reduce risk of incident vertebral fracture with up to 4 years of treatment, while 4 years of alendronate also lowers risk of incident clinical fractures. Though trial data on estrogen and estrogen/progestin are not available for women with osteoporosis defined using standard criteria (i.e., BMD or past hip or vertebral fractures), their reduction in risk of incident clinical fractures in postmenopausal women with past fractures, as a proxy for osteoporosis, suggests they’d be effective in women known to have osteoporosis. However, due to their counterbalancing increases in risk of cardiovascular disease, dementia and breast cancer, and the failure of secondary analyses to identify any subpopulation in whom fracture benefits outweigh these harms, there seems no indication for the long-term use of estrogen and estrogen/progestin for fracture prevention. In patients who already have completed a course of osteoporosis drug treatment (i.e., 3 to 5 years), the most favorable trial evidence for continued treatment is for an approximately 3 percent absolute reduction in risk of incident clinical vertebral fractures with 10 years of alendronate versus 5 years. Observational data suggests that the absolute increases in risk for AFF and ONJ with long-term bisphosphonate treatment are far smaller. Future research to help patients and clinicians make decisions about long-term osteoporosis drug treatment should help refine these absolute risk estimates. It should examine how these risks vary as a function of patient, bone, and drug treatment characteristics (e.g., age, sex, pre-drug holiday BMD, duration of prior osteoporosis drug treatment). Further, because patients don’t experience these events dichotomously, quantifying their morbidity in a common measure may help patients weigh their trade-offs more easily.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Adverse event</td>
</tr>
<tr>
<td>AFF</td>
<td>Atypical femoral fracture</td>
</tr>
<tr>
<td>AFib</td>
<td>Atrial fibrillation</td>
</tr>
<tr>
<td>aHR</td>
<td>Adjusted hazard ratio</td>
</tr>
<tr>
<td>AHRQ</td>
<td>Agency for Healthcare Research &amp; Quality</td>
</tr>
<tr>
<td>ALP</td>
<td>Alkaline phosphatase</td>
</tr>
<tr>
<td>aOR</td>
<td>Adjusted odds ratio</td>
</tr>
<tr>
<td>ASBMR</td>
<td>American Society for Bone and Mineral Research</td>
</tr>
<tr>
<td>BSAP</td>
<td>Bone specific alkaline phosphatase</td>
</tr>
<tr>
<td>BMD</td>
<td>Bone mineral density</td>
</tr>
<tr>
<td>BMI</td>
<td>Body mass index (kg/m²)</td>
</tr>
<tr>
<td>CCT</td>
<td>Controlled clinical trial</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>Cochrane Central Register of Controlled Trials</td>
</tr>
<tr>
<td>CER</td>
<td>Comparative Effectiveness Review</td>
</tr>
<tr>
<td>CFx</td>
<td>Clinical fracture</td>
</tr>
<tr>
<td>CHF</td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CKD</td>
<td>Chronic kidney disease</td>
</tr>
<tr>
<td>CORE</td>
<td>Continuing Outcomes Relevant to Evista</td>
</tr>
<tr>
<td>COX</td>
<td>Cyclooxygenase</td>
</tr>
<tr>
<td>CTX</td>
<td>C-terminal telopeptide of type 1 collagen</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>CVFx</td>
<td>Clinical vertebral fracture</td>
</tr>
<tr>
<td>DM</td>
<td>Diabetes mellitus</td>
</tr>
<tr>
<td>DVT</td>
<td>Deep vein thrombosis</td>
</tr>
<tr>
<td>DXA</td>
<td>Dual-energy x-ray absorptiometry</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated glomerular filtration rate</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic health record</td>
</tr>
<tr>
<td>EPC</td>
<td>Evidence-Based Practice Center</td>
</tr>
<tr>
<td>FDA</td>
<td>U.S. Food and Drug Administration</td>
</tr>
<tr>
<td>FIT</td>
<td>Fracture Intervention Trial</td>
</tr>
<tr>
<td>FLEX</td>
<td>Fracture Intervention Trial Long-Term Extension</td>
</tr>
<tr>
<td>FN</td>
<td>Femoral neck</td>
</tr>
<tr>
<td>FN-BMD</td>
<td>Femoral neck bone mineral density</td>
</tr>
<tr>
<td>FRAX</td>
<td>World Health Organization Fracture Risk Assessment Tool</td>
</tr>
<tr>
<td>Fx</td>
<td>Fracture</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendations Assessment, Development, and Evaluation</td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
</tr>
<tr>
<td>g/cm²</td>
<td>Grams per centimeter squared</td>
</tr>
<tr>
<td>HORIZON</td>
<td>Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly Trial</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>ITT</td>
<td>Intention to treat</td>
</tr>
<tr>
<td>IV</td>
<td>Intravenous</td>
</tr>
<tr>
<td>KQ</td>
<td>Key Question</td>
</tr>
</tbody>
</table>
LS  Lumbar spine
LS-BMD  Lumbar spine bone mineral density
MD  Mean difference
mg  Milligrams
MOF  Major osteoporotic fracture
MORE  Multiple Outcomes of Raloxifene Evaluation
N  Number
NA  Not applicable
NIA  National Institute on Aging
NIAMS  National Institute of Arthritis and Musculoskeletal and Skin Diseases
NIH  National Institute of Health
NR  Not reported
NS  Not significant
NSAIDS  Nonsteroidal anti-inflammatory drugs
NTX  N-terminal telopeptide of type 1 collagen
ODP  Office of Disease Prevention
ONJ  Osteonecrosis of the jaw
OR  Odds ratio
P2P  Pathways to Prevention
P1NP  Procollagen type 1 amino-terminal propeptide
PE  Pulmonary embolism
PICOTS  Population, Interventions, Comparators, Outcomes, Timing, and Settings
PM  Postmenopausal
PRISMA  Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTH  Parathyroid hormone
HRpQCT  High resolution peripheral quantitative computed tomography
QD  Per day
RCT  Randomized controlled trial
RD  Risk difference
ROB  Risk of bias
RR  Risk ratio
RVFx  Radiographic vertebral fracture
SAE  Serious adverse event
SC  Subcutaneous
SD  Standard deviation
SE  Standard error
SEADS  Supplemental Evidence and Data for Systematic Reviews
SERM  Selective estrogen receptor modulator
SOE  Strength of Evidence
SSRI  Selective Serotonin Reuptake Inhibitors
TEP  Technical Expert Panel
TH  Total hip
TH-BMD  Total hip bone mineral density
VFx  Vertebral fracture
VTE  Venous thromboembolism
WHO  World Health Organization
References


42. Erviti J, Alonso A, Oliva B, et al. Oral bisphosphonates are associated with increased risk of subtrochanteric and


70. Gedmintas L, Solomon DH, Kim SC. Bisphosphonates and risk of subtrochanteric,


75. Black D. Change in BMD as a Surrogate for Fracture Risk Reduction in Osteoporosis Trials: Results from Poolled, Individual-level Patient Data from the FNIH Bone Quality Project [abstract]. American Society for Bone and Mineral Research Annual Meeting, September 28-October 1, Montréal, Québec, Canada. 2018.


