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

Contract No. 290-2015-00008-I

Prepared by:
Minnesota Evidence-based Practice Center
Minneapolis, MN

Investigators:
Howard A. Fink, M.D., M.P.H.
Roderick MacDonald, M.S.
Mary L. Forte, Ph.D., D.C.
Christina E. Rosebush, M.P.H.
Kristine E. Ensrud, M.D., M.P.H.
John T. Schousboe, M.D., Ph.D.
Victoria A. Nelson, M.Sc.
Kristen Ullman, M.P.H.
Mary Butler, Ph.D., M.B.A.
Carin M. Olson, M.D.
Brent C. Taylor, M.P.H., Ph.D.
Michelle Brasure, Ph.D., M.S.P.H., M.L.I.S.
Timothy J. Wilt, M.D., M.P.H.

AHRQ Publication No. 19-EHC016-EF
April 2019
Key Messages

Purpose of Review
To summarize the effects of long-term osteoporosis drug treatment and of osteoporosis drug treatment discontinuation and holidays.

Key Messages
- Evidence on the effects of long-term osteoporosis drug treatment and drug continuation versus discontinuation is mostly limited to white, healthy, postmenopausal women.
- Long-term alendronate reduces radiographic vertebral and nonvertebral fractures in women with osteoporosis; long-term zoledronate reduces vertebral and nonvertebral fractures in women with osteopenia or osteoporosis.
- Long-term bisphosphonates may increase atypical femoral fractures and osteonecrosis of the jaw, although both are rare.
- In women with osteoporosis, long-term raloxifene reduces vertebral fractures, but not hip or nonvertebral fractures, and increases venous thromboembolism.
- Long-term oral hormone therapies reduce hip and clinical fractures but increase multiple serious harms.
- Evidence is insufficient about the effects of long-term denosumab, risedronate, ibandronate, teriparatide, and abaloparatide on fractures and harms.
- Continuing bisphosphonates after 3–5 years versus discontinuation reduces some measures of vertebral fractures, but not nonvertebral fractures.
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 healthcare in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new healthcare 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 healthcare system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the website (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.

If you have 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.
Director
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, Evidence-based Practice Program
Center for Evidence and Practice Improvement
Agency for Healthcare Research and Quality

Arlene S. Bierman, M.D., M.S.
Director
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

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

We thank many individuals for their contributions to this project: Laura Pincock and Jill Huppert, our AHRQ Task Order Officers; Keisha Shropshire and Carrie Klabunde, of the National Institutes of Health (NIH) Office of Disease Prevention (ODP); and Lyndon Joseph, National Institute on Aging (NIA), and Faye Chen, National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the NIH nominators of this review topic. All provided support, direction, and editorial comments. We also thank Jeannine Ouellette and Cheryl Cole-Hill for their copy editorial help bringing the report to completion.

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 $5,000 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:

Douglas C. Bauer, M.D.
Professor
University of California, San Francisco
School of Medicine
San Francisco, CA

Carolyn J. Crandall, M.D., M.S., FACP*
Professor of Medicine
David Geffen School of Medicine at University of California, Los Angeles
Los Angeles, CA

Cynthia M. Boyd, M.D., M.P.H.*
Professor
Johns Hopkins School of Medicine
Baltimore, MD

Andrea Z. LaCroix, Ph.D.
Professor and Chief of Epidemiology
Family Medicine and Public Health at University of California, San Diego
San Diego, CA

Jane A. Cauley, Dr.P.H.*
Distinguished Professor of Epidemiology
University of Pittsburgh
Pittsburgh, PA

Eric S. Orwoll, M.D.
Professor of Medicine
Oregon Health & Science University
Portland, OR

Angela M. Cheung, M.D., Ph.D.
Senior Scientist
Departments of Medicine and Medical Imaging of University of Toronto
Toronto, Ontario, Canada

*Provided input on Draft 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 $5,000 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:

Bart Clarke, M.D.
Division of Endocrinology and Metabolism
Mayo Clinic
Rochester, MN

Gail A. Greendale, M.D.
Professor of Medicine
David Geffen School of Medicine at
University of California, Los Angeles
Los Angeles, CA

Nancy Lane, M.D.
Director and Distinguished Professor
Center for Musculoskeletal Health and
Department of Internal Medicine
University of California at Davis, School of Medicine
Sacramento, California

Jasvinder Singh, M.D., M.P.H.
Division of Rheumatology
University of Alabama
Birmingham, AL
Long-Term Drug Therapy and Drug Holidays for Osteoporosis Fracture Prevention: A Systematic Review

Structured Abstract

Objective. To summarize the effects of long-term osteoporosis drug treatment (ODT) and ODT discontinuation and holidays on fractures and harms.

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

Review methods. We defined long-term ODT as >3 years and ODT holidays as discontinuation for ≥1 year after ≥1 year of use. Trials were used for incident fractures and harms, and controlled observational studies were included for additional harms. Two investigators rated risk of bias. For studies with low or medium risk of bias, one investigator extracted data and a second verified accuracy. Two investigators graded strength of evidence (SOE).

Results. Sixty-one English-language studies were included. In women with osteoporosis, 4 years of alendronate reduced clinical fractures (hazard ratio [HR] 0.64 [95% confidence interval (CI) 0.50, 0.82]) (moderate SOE) and radiographic vertebral fractures (HR 0.50 [95% CI 0.31, 0.82]) (moderate SOE), while 4 years of raloxifene reduced clinical vertebral fractures (relative risk 0.58 [95% CI 0.43, 0.79]) (high SOE), but not hip (moderate SOE) or nonvertebral fractures (high SOE). In women with osteopenia or osteoporosis, 6 years of zoledronate reduced incident clinical fractures (HR 0.73 [95% CI 0.60, 0.90]) (moderate SOE) and clinical vertebral fractures (HR 0.41 [95% CI 0.22, 0.75]) (moderate SOE). In postmenopausal women with unknown osteoporosis or osteopenia status, both long-term oral estrogen and estrogen/progestin reduced clinical fractures (high SOE) and hip fractures (moderate SOE). After 3–5 years of prior treatment, continuation of zoledronate or alendronate versus drug holiday inconsistently reduced incident vertebral fracture outcomes (radiographic only for zoledronate [low SOE], clinical only for alendronate [moderate SOE]), but did not reduce nonvertebral fractures (low SOE). Hormone therapies increased cardiovascular events, mild cognitive impairment or dementia, and other harms. Observational studies showed that long-term bisphosphonates may increase atypical femoral fractures (AFF) (low SOE) and osteonecrosis of the jaw (low SOE in 2 comparisons, insufficient in 1).

Limitations. Most data were from white, healthy, postmenopausal women, limiting generalizability. Trials often had low power for incident clinical fractures. No trials compared active treatments, sequential treatments, or different durations of drug holidays. Harms and controls were inconsistently defined.

Conclusions. Long-term alendronate, zoledronate, and oral hormone therapy reduced nonvertebral fractures in older women, with oral hormone therapy also reducing hip fractures. While absolute reductions in typical fractures with long-term bisphosphonates are large relative to increases in AFF, reduced hip fracture risk with oral hormone therapy appears offset by increased risk of serious harms. Evidence is limited regarding ODT holidays for fractures and...
harms. Future research is needed, including randomized trials comparing ODT holiday durations and sequential treatments powered for clinical fractures, and controlled cohort studies of ODT holidays to estimate rare harms.
# Contents

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

Chapter 1. Introduction .................................................................................................................... 1
  Background .................................................................................................................................. 1
  Scope and Key Questions .............................................................................................................. 4
    Key Questions ............................................................................................................................. 4
  PICOTS ....................................................................................................................................... 6
  Harms ...................................................................................................................................... 11
    Analytic Framework ................................................................................................................... 12

Chapter 2. Methods ........................................................................................................................ 13
  Topic Refinement and Review Protocol ....................................................................................... 13
  Literature Search Strategy .......................................................................................................... 13
  Study Selection ........................................................................................................................... 15
  Risk of Bias Assessment ............................................................................................................ 15
  Data Extraction ........................................................................................................................... 16
  Data Synthesis ............................................................................................................................ 16
  Strength of Evidence for Major Comparisons and Outcomes .................................................. 17
  Applicability ................................................................................................................................ 17
  Peer Review and Public Commentary ......................................................................................... 18

Chapter 3. Search Results ............................................................................................................... 19

Chapter 4. Efficacy of Long-Term Osteoporosis Drug Treatment ............................................... 22
  Alendronate ................................................................................................................................ 22
    Key Points ................................................................................................................................. 22
    Eligible Studies ......................................................................................................................... 23
    Outcomes ................................................................................................................................. 24
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
      Drug Characteristics .............................................................................................................. 25
  Zoledronic Acid (Zoledronate) .................................................................................................. 27
    Key Points ................................................................................................................................. 27
    Eligible Studies ......................................................................................................................... 27
    Outcomes ................................................................................................................................. 28
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
      Drug Characteristics .............................................................................................................. 28
  Denosumab ................................................................................................................................ 29
    Key Points ................................................................................................................................. 29
    Eligible Studies ......................................................................................................................... 29
    Outcomes ................................................................................................................................. 30
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
      Drug Characteristics .............................................................................................................. 31
  Raloxifene .................................................................................................................................. 32
    Key Points ................................................................................................................................. 32
    Eligible Studies ......................................................................................................................... 32
    Outcomes ................................................................................................................................. 33
    Variation in Long-Term Treatment Efficacy as a Function of Patient, Bone, or Osteoporosis
      Drug Characteristics .............................................................................................................. 34
  Hormone Therapy ....................................................................................................................... 35
### Chapter 6. Effects of Osteoporosis Drug Holidays

**Alendronate** .......................... 68

- Key Points ........................................... 68
- Eligible Studies .................................... 69
- Outcomes ............................................. 70
  - Variation in Effects of Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics .................................................. 71

**Zoledronic Acid (Zoledronate)** .................................. 73

- Key Points ........................................... 73
- Eligible Studies .................................... 73
- Outcomes ............................................. 74
  - Variation in Effects of Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics .................................................. 75

**Denosumab** ........................................ 76

- Key Points ........................................... 76
- Eligible Studies .................................... 76
- Outcomes ............................................. 76

**Other Drugs Investigated for Review** ................................... 78

- Raloxifene ........................................ 78
- Ibandronate ........................................ 78
- Risedronate ........................................ 78
- Teriparatide ........................................ 78
- Abaloparatide ..................................... 78
- Hormone Therapy ............................... 78

### Chapter 7. Harms of Drug Holidays

**Alendronate** .................................. 79

- Key Points ........................................... 79
- Eligible Studies .................................... 79
- Outcomes ............................................. 80
  - Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics .................................................. 81

**Zoledronic Acid (Zoledronate)** .................................. 82

- Key Points ........................................... 82
- Eligible Studies .................................... 82
- Outcomes ............................................. 82
  - Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics .................................................. 83

**Denosumab** ........................................ 84

- Key Points ........................................... 84
- Eligible Studies .................................... 84
- Outcomes ............................................. 84
  - Variation in Harms During Drug Holiday as a Function of Patient, Bone, or Osteoporosis Drug Characteristics .................................................. 85

**Other Drugs Investigated for Review** ................................... 86
Chapter 8. Discussion

Overview ................................................................................................................................... 87
Key Findings and Strength of Evidence ............................................................................... 87
Applicability ................................................................................................................................ 91
Findings in Relation to What Is Already Known ................................................................. 91
Efficacy of Short-Term Osteoporosis Drug Treatment ........................................................... 91
Efficacy of Long-Term Osteoporosis Drug Treatment ............................................................ 91
Harms of Long-Term Osteoporosis Drug Treatment ............................................................ 92
Effect of Osteoporosis Drug Treatment Holidays ................................................................ 93
Variation in Osteoporosis Drug Treatment Outcomes .......................................................... 93
Limitations of the Comparative Effectiveness Review ............................................................ 94
Limitations of the Evidence Base ............................................................................................. 95
Future Research ........................................................................................................................ 96
Conclusions................................................................................................................................. 103

References .................................................................................................................................. 104

Abbreviations ............................................................................................................................... 112

Tables
Table A. Evidence on efficacy of long-term (>3 years) osteoporosis drug treatment .......... ES-4
Table B. Evidence on harms of long-term (>3 years) osteoporosis drug treatment ............... ES-8
Table C. Evidence on effects of osteoporosis drug continuation versus discontinuation on incident fractures................................................................................................................ ES-13
Table D. Evidence on harms of osteoporosis drug continuation versus discontinuation ..... ES-14
Table 1. U.S. FDA approved drugs used for osteoporosis treatment and prevention ............ 3
Table 2. PICOTS ............................................................................................................................. 7
Table 3. Harms evaluated for possible association with long-term osteoporosis drug therapy... 11
Table 4. Study inclusion criteria ................................................................................................. 14
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 ........................................................................... 21
Table 6. Baseline participant characteristics from included alendronate long-term treatment efficacy publications for KQ 1 and KQ 2 and FIT-1 ................................................................. 24
Table 7. Baseline characteristics of the zoledronate trial versus placebo ................................ 27
Table 8. Baseline characteristics of the denosumab trial versus placebo ................................. 29
Table 9. Baseline participant characteristics from the MORE trial and CORE extension ....... 33
Table 10. Baseline characteristics of hormone therapy versus placebo/control RCTs .......... 36
Table 11. Estimated risk of ST/FS fracture, ONJ, or atrial fibrillation with alendronate use from controlled observational studies ................................................................. 45
Table 12. Estimated risk of AFF or ST/FS fracture for any bisphosphonate use from observational studies ........................................................................................................................... 54
Table 13. Trial participant characteristics at the time of assignment to alendronate continuation versus discontinuation................................................................................................................... 70
Table 14. Participant characteristics at the baseline of the zoledronate continuation versus discontinuation trials..................................................................................................................... 74
Table 15. Future research recommendations .......................................................................................................................... 96

Figures
Figure 1. Analytic framework for long-term osteoporosis drug treatment and osteoporosis drug treatment holidays for fracture prevention................................................................................................. 12
Figure 2. Literature flow diagram.......................................................................................................................... 20

Appendixes
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 References
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 frequently cause pain, disability, and impaired quality of life; and hip and clinical vertebral fractures, specifically, are associated with increased mortality. Because risk of most fractures rises steeply with age, and the population is aging, fracture burden is projected to increase in coming decades.

In short-term (18 to 36 months) randomized controlled trials (RCTs) for osteoporosis treatment, bisphosphonates (alendronate, zoledronate, risedronate, ibandronate), denosumab, teriparatide, and abaloparatide lower risk of nonvertebral fractures, clinical vertebral fractures (usually diagnosed in the community because of back pain, with study comparison of study and community radiographs), and radiographic vertebral fractures (identified in studies by comparing vertebral heights on scheduled serial vertebral radiographs; mostly unrecognized in the community). Several bisphosphonates (alendronate, zoledronate, risedronate) and denosumab also lower risk of hip fractures.

Despite the evidence on the efficacy of short-term osteoporosis drug treatment for reducing fracture risk in appropriate patients, there is uncertainty about the balance of benefits and harms of long-term or continued treatment. A recent American College of Physicians clinical practice guideline recommended treatment of osteoporotic women with a bisphosphonate or denosumab for 5 years to reduce risk of hip and vertebral fractures, and suggested that high-risk patients may benefit from more than 5 years of treatment.

However, 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 treatment be discontinued periodically. Several groups advocate bisphosphonate “drug holidays” to minimize harms while preserving as much fracture benefit as possible. But, there is no consensus about who should get them, when they should start, how long they should last, and the criteria for restarting treatment. By contrast, drug holidays are not recommended after denosumab, because bone loss increases rapidly after discontinuation, possibly increasing risk of radiographic vertebral fractures.

Uncertainties about the most appropriate use of long-term osteoporosis drug treatment and of osteoporosis drug holidays led to scheduling a National Institutes of Health (NIH) Office of Disease Prevention (ODP) Pathways to Prevention (P2P) workshop. The goals of the workshop were to present an evidence-based synthesis of the pertinent research base, and to suggest future research to assist patients, clinicians, and other healthcare decision makers.

To further these aims, we conducted this systematic review to address the following questions: (1) What are the effects of long-term (>3 years) osteoporosis drug treatment versus control on risks of incident fractures and harms; (2) Do effects of long-term osteoporosis drug treatment vary as a function of patient, bone, or osteoporosis drug characteristics; (3) Among individuals receiving osteoporosis drug treatment to prevent fracture, what are the effects of continuing versus discontinuing treatment (i.e., osteoporosis drug holiday) on risks of incident fractures and harms; and (4) Do these outcomes of drug holidays vary as a function of patient, bone, or osteoporosis drug characteristics?
Methods

The review was conducted following the Agency for Healthcare Research and Quality (AHRQ) methods guidance. The protocol is available at 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 8,356 unique publications through October 2018, of which 61 met eligibility criteria and were included in the review. Of 48 publications with low or medium risk of bias (ROB), there were 35 randomized or controlled clinical trials (9 unique studies) and 13 controlled observational studies (11 unique studies) (Appendix C of the full report). 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 bone mineral density (BMD) and vertebral fracture history, and some including women with osteopenia. Observational studies included 84 to 100 percent women. Mean participant age was 72 years, with all but two studies reporting mean age <80 years. Most observational studies presumed participants had osteoporosis because of past fracture or use of osteoporosis drugs, but none reported BMD status.

Long-Term Osteoporosis Drug Treatment

Efficacy

Seven eligible placebo-controlled RCTs with low or medium ROB examined the effect of long-term treatment, one each for alendronate, zoledronate, raloxifene, denosumab, and estrogen, and two for estrogen/progestin.

In women with osteopenia or osteoporosis by BMD, but with no past vertebral fracture, 4 years of alendronate versus placebo reduced incident radiographic vertebral fractures (hazard ratio [HR] 0.56 [95% confidence interval (CI), 0.39, 0.80]) (high strength of evidence [SOE]) (Table A), while absolute risk reductions for incident hip and nonvertebral fracture were small and not statistically significant (low SOE). In women with osteoporosis by BMD or past fracture, 4 years of raloxifene versus placebo reduced incident radiographic (relative risk [RR] 0.64 [0.53, 0.76]) and clinical vertebral fractures (RR 0.58 [95% CI 0.43, 0.79]) (both high SOE). However, raloxifene did not reduce incident hip or nonvertebral fracture (moderate and high SOE, respectively). In older women with osteopenia or osteoporosis, 6 years of zoledronate versus placebo reduced incident clinical fractures (HR 0.73 [95% CI 0.60, 0.90]) (low SOE), incident nonvertebral fractures (HR 0.66 [95% CI 0.51, 0.85]) (moderate SOE) and incident clinical vertebral fractures (HR 0.41 [95% CI 0.22, 0.75]) (moderate SOE). Nonvertebral fractures appeared similarly reduced in the subset of women with osteopenia. In women with unknown osteoporosis or osteopenia status, incident clinical fractures (high SOE) and incident hip fractures (moderate SOE) both were reduced with hormone therapy compared to placebo, with 5.6 years of oral estrogen/progestin in women with an intact uterus, and with 7 years of unopposed oral estrogen in women with a hysterectomy.
Evidence was insufficient to compare fracture risk between women on long-term denosumab versus placebo, and there were no data from eligible trials about the long-term fracture efficacy of sequential osteoporosis drug therapy (e.g., anabolic followed by anti-resorptive, or denosumab followed by bisphosphonate).

Alendronate, zoledronate, denosumab, and raloxifene for long-term treatment each increased hip and lumbar spine BMD compared to placebo.

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

Efficacy of long-term alendronate appears to vary as a function of baseline BMD. Relative risk of incident clinical fractures was significantly reduced in women with osteoporotic BMD (femoral neck BMD ≤-2.5 (moderate SOE), but not in women with osteopenic BMD (femoral neck BMD -1.6 to >-2.5) (low SOE) (Table A; Appendix Table D8). In women with osteoporosis, relative risk of incident radiographic vertebral fracture with long-term alendronate was halved versus placebo (HR 0.50 [95% CI 0.31, 0.82]) (moderate SOE). Although women with femoral neck BMD -2.5 to -2 had a similar relative reduction in these fractures versus placebo (HR 0.54 [95% CI 0.28, 1.04]) (low SOE), a lower proportion of women with osteopenia had incident radiographic vertebral fractures, and results were not statistically significant. No tests of interaction were reported for these BMD stratified results.

In a post hoc analysis, women with osteoporosis had a reduced risk of incident hip fracture with long-term alendronate versus placebo, but women with osteopenia had no reduced risk (p-value for interaction not reported). In additional post hoc analyses, some conducted in women with osteopenia, neither past nonvertebral fracture, 10-year major osteoporotic fracture probability calculated with femoral neck BMD, nor pretreatment levels of bone turnover markers significantly 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. Two large trials of oral hormone therapy versus placebo in women with unknown osteoporosis or osteopenia status reported inconsistent findings about whether treatment effect on risk of incident hip and clinical fractures differed as a function of age or time since menopause. However, authors minimized their one significant interaction for age because of the many interactions examined.
<table>
<thead>
<tr>
<th>Comparison # Studies by Design Treatment Duration</th>
<th>Participant Characteristics</th>
<th>Incident Fracture Outcome</th>
<th>Relative and Absolute Risk Differences (95% CI)</th>
<th>Strength of Evidence* (Justification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs. placebo 1 RCT 4 yr</td>
<td>4,432 PM women with osteopenia or osteoporosis (T-score ≤-1.6) and no RVF</td>
<td>CF</td>
<td>No difference: HR=0.86 [0.73, 1.01]; ARR=-2 [-4, 0]</td>
<td>Low (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVF</td>
<td>No difference: HR=0.88 [0.74, 1.04]; ARR=-1 [-3, 0]</td>
<td>Low (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hip</td>
<td>No difference: HR=0.79 [0.43, 1.44]; ARR=-0.2 [-0.8, 0.4]</td>
<td>Low (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RVF Lower risk: HR=0.56 [0.39, 0.80]; ARR=-2 [-3, -1]</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>1,631 PM women with osteoporosis by BMD (T-score ≤-2.5) and no RVF</td>
<td></td>
<td>CF Lower risk: HR=0.65 [0.50, 0.82]; ARR=-7 [-10, -3]</td>
<td>Moderate (RB)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RVF Lower risk: HR=0.50 [0.31, 0.82]; ARR=-3 [-5, -1]</td>
<td>Moderate (RB)</td>
<td></td>
</tr>
<tr>
<td>Zoledronate vs. placebo 1 RCT 6 yr</td>
<td>2,000 PM women ≥65 with osteoporosis or osteopenia</td>
<td>CF Lower risk: HR=0.73 [0.60, 0.90]; ARR=-5 [-9, -2]</td>
<td>Moderate (IM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NVF Lower risk: HR=0.66 [0.51, 0.85]; ARR=-5 [-8, -2]</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hip No difference: HR=0.66 [0.27, 1.16]; ARR=-0.4 [-1, 0.5]</td>
<td>Low (h-IM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CVF Lower risk: HR=0.41 [0.22, 0.75]; ARR=-2 [-3, -1]</td>
<td>Moderate (IM)</td>
<td></td>
</tr>
<tr>
<td>Denosumab vs. placebo 1 RCT 4 yr</td>
<td>365 PM women with osteopenia or osteoporosis by BMD</td>
<td>CF RR=0.97 [0.40, 2.35]; ARR=-0.4 [-10, 9]</td>
<td>Insufficient (RB, IN, h-IM)</td>
<td></td>
</tr>
<tr>
<td>Raloxifene vs. placebo 1 RCT with 1 CCT extension 16, 17, 26-37 4 to 8 yr</td>
<td>6,828 PM women with osteoporosis by BMD or RVF</td>
<td>NVF No difference: 4 yr: RR=0.93 [0.81, 1.06]; ARR NA 8 yr: HR=1.00 [0.82, 1.21]; ARR NA</td>
<td>4 yr: High 8 yr: Moderate (RB)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hip No difference: 4 yr: RR=0.97 [0.62, 1.52]; ARR=0 [-0.6, 0.5]</td>
<td>Moderate (IM)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CVF Lower risk: 4 yr: RR=0.58 [0.43, 0.79]; ARR=-2 [-3, -1]</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RVF Lower risk: 4 yr: RR=0.64 [0.53, 0.76]; ARR=-5 [-6, -3]</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Estrogen vs. placebo 10,739 PM women with past hysterectomy</td>
<td></td>
<td>CF Lower risk: HR=0.71 [0.64, 0.80]; ARR=-4 [-5, -3]</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Comparison # Studies by Design Treatment Duration</td>
<td>Participant Characteristics</td>
<td>Incident Fracture Outcome</td>
<td>Relative and Absolute Risk Differences (95% CI)</td>
<td>Strength of Evidence* (Justification)</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----------------------------</td>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>1 RCT&lt;sup&gt;10&lt;/sup&gt; 7.1 yr (mean)</td>
<td>Hip</td>
<td>Lower risk: HR=0.65 [0.45, 0.94]; ARR= -0.5 [-0.9, -0.08]</td>
<td>Moderate (IM)</td>
<td></td>
</tr>
<tr>
<td>3,816 PM women with past hysterectomy and past clinical fracture</td>
<td>CF</td>
<td>Lower risk: HR=0.73 [0.62, 0.86]; ARR= -5 [-7, -2]</td>
<td>Low (RB, IM)</td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td>Lower risk: HR=0.55 [0.32, 0.94]; ARR= -1 [-2, 0]</td>
<td>Low (RB, IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53 PM women with past hysterectomy and osteoporosis by BMD</td>
<td>CF</td>
<td>HR=0.83 [0.17, 3.91]; ARR NA</td>
<td>Insufficient (RB, h-IM)</td>
<td></td>
</tr>
<tr>
<td>363 PM women with past hysterectomy and osteopenia by BMD</td>
<td>CF</td>
<td>HR=0.83 [0.49, 1.40]; ARR NA</td>
<td>Insufficient (RB, h-IM)</td>
<td></td>
</tr>
<tr>
<td>Estrogen/progestin vs. placebo 1 RCT&lt;sup&gt;21&lt;/sup&gt; 5.6 yr (mean)</td>
<td>16,608 PM women with intact uterus</td>
<td>CF</td>
<td>Lower risk: HR=0.76 [0.69, 0.83]; ARR= -2.4 [-3.3, -1.5]</td>
<td>High</td>
</tr>
<tr>
<td>Hip</td>
<td>Lower risk: HR=0.67 [0.47, 0.96]; ARR= -0.3 [-0.6, -0.03]</td>
<td>Moderate (IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVF</td>
<td>Lower risk: HR=0.65 [0.46, 0.92]; ARR= -0.3 [-0.5, -0.02]</td>
<td>Moderate (IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5,897 PM women with intact uterus and past clinical fracture</td>
<td>CF</td>
<td>Lower risk: HR=0.78 [0.68, 0.91]; ARR= -3 [-5, -1]</td>
<td>Low (RB, IM)</td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td>No difference: HR=0.77 [0.48, 1.22]; ARR= -0.3 [-0.9, 0.2]</td>
<td>Low (RB, IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM women with intact uterus and osteoporosis by BMD n not reported</td>
<td>CF</td>
<td>HR=0.53 [0.25, 1.10]; ARR NA</td>
<td>Insufficient (RB, h-IM)</td>
<td></td>
</tr>
<tr>
<td>Estrogen/progestin vs. nonplacebo control 1 RCT&lt;sup&gt;20&lt;/sup&gt; 4 yr</td>
<td>36 PM women with osteoporosis by BMD (T-score &lt;-2) and RVF</td>
<td>NVF</td>
<td>RR=0.93 [0.06, 13.5]; ARR= -0.5 [-19, 18]</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>RVF</td>
<td>RR=0.37 [0.09, 1.62]; ARR= -22 [-53, 8]</td>
<td>Insufficient (RB, h-IM)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** ARR=absolute risk reduction; BMD=bone mineral density; CCT=controlled clinical trial; CF=clinical fracture; CI=confidence intervals; CVF=clinical vertebral fracture; h-IM=highly imprecise; HR=hazard ratio; IM=imprecise; IN=indirect; NA=not available; NVF=nonvertebral fracture; PM=postmenopausal; RB=medium risk of bias; RCT=randomized controlled trial; RR=risk ratio; RVF=radiographic vertebral fracture;

*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.
‡Results reported for raloxifene 60 mg/d dose group.
**Analyses pooled all participants initially assigned to denosumab, which included both those who received long-term and short-term denosumab.
Harms

Due to few events, RCT data provided insufficient evidence about whether long-term alendronate or zoledronate increase risk of radiologically confirmed atypical femoral fracture (AFF), subtrochanteric or femoral shaft fractures without radiologically confirmed AFF features, or osteonecrosis of the jaw (ONJ).

Data from controlled, long-term observational studies suggest that alendronate and bisphosphonates as a class increase both radiologically confirmed AFF (low SOE), and subtrochanteric and femoral shaft fractures without radiologically confirmed with AFF features (low SOE) (Table B). Relative risks for these outcomes varied from 1 to >100 across studies, likely related to heterogeneity in designs. Relative risks appeared higher for radiologically confirmed AFF than for cases defined only by diagnostic codes. However, some AFF risk estimates were calculated using controls with subtrochanteric or femoral shaft fractures without AFF features. In those cases, risk estimates reflect the probability that a subtrochanteric or femoral shaft fracture will have AFF features, and not the relative risk of sustaining an AFF. Studies also differed in whether fractures cases were excluded for cancer and excess trauma; in whether current bisphosphonate use was compared to no use, limited past use, or nonbisphosphonate osteoporosis drug use; and in how they addressed potential confounding.

Few observational studies provided data about risk of ONJ. Different studies provided low and insufficient strength evidence, respectively, about whether long-term alendronate increases risk of ONJ. Relative risk estimates varied widely between studies, likely due to heterogeneity in case definitions, treatment control groups, and covariate modeling.

We found insufficient evidence about whether long-term zoledronate increases risk of AFF or ONJ. The single trial of long-term zoledronate versus placebo reported no cases of AFF or ONJ in either treatment group, and we identified no eligible observational studies that evaluated risk of these harms with long-term zoledronate.

Due to its pooling of results for both short- and long-term denosumab treatment, it was not possible to conclude anything about the risk of harms of long-term denosumab compared with placebo from the one study that met eligibility for this review. In long-term trials of oral hormone therapy, specifically estrogen/progestin and estrogen versus placebo in postmenopausal women with unknown osteoporosis or osteopenia status, risk was significantly increased for cardiovascular disease, and mild cognitive impairment or dementia. Risk for a composite outcome measure defined to weigh risk of incidence of any of several serious harms (coronary heart disease, invasive breast cancer, stroke, pulmonary embolism, endometrial cancer, colorectal cancer, or death due to other causes) versus hip fracture did not differ between estrogen/progestin and placebo or between estrogen and placebo. Results were similar regardless of participants’ baseline fracture risk, indicating that risk of one or more of these harms offset the reduction in hip fractures even in participants at highest baseline fracture risk. In addition, estrogen/progestin was associated with an increased risk of invasive breast cancer. Long-term raloxifene versus placebo in treatment for osteoporosis significantly increased the risk of deep vein thrombosis and pulmonary embolism by about 3 to 4-fold.
Variation in Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics

We found little evidence about factors that modify risk of harms with long-term osteoporosis drug treatment. One study was inconclusive about whether relative risk for AFF associated with bisphosphonate use increased with age.38 Three controlled observational studies reported that >5 years of bisphosphonate use increased risk of subtrochanteric or femoral shaft fractures (ST/FSF) or radiologically confirmed AFF more than did 3-5 years of use.38, 48, 49 However, none of these studies reported tests for interaction by treatment duration. For long-term raloxifene versus placebo, one study reported that risk of deep venous thrombosis and pulmonary embolism did not vary as a function of baseline cardiovascular risk,32 and another that risk of incident stroke was lower with raloxifene versus placebo in women with increased cardiovascular risk.31 Trials of long-term oral hormone therapy evaluated whether risk of harms varied by a long list of patient characteristics.44-47 Though results suggested that risk of breast cancer with estrogen/progestin compared to placebo may be greater with increased duration of prior postmenopausal hormone use, this was the only significant result out of many examined, and may have been due to chance. Strength of evidence was not assessed for effect modifiers.
Table B. Evidence on harms of long-term (>3 years) osteoporosis drug treatment

<table>
<thead>
<tr>
<th>Comparison # Studies by Design Treatment Duration</th>
<th>Participant Characteristics</th>
<th>Harms</th>
<th>Relative and Absolute Risk Differences (95% CI)</th>
<th>Strength of Evidence* (Justification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs. placebo 1 RCT&lt;sup&gt;30&lt;/sup&gt; 3 to 4.5 yr</td>
<td>6,459 PM women with osteopenia or osteoporosis (T-score ≤-1.6) with or without RVF</td>
<td>ST or FS fracture DX with rare x-ray review for confirmation of AFF features (n=2 cases)</td>
<td>HR=1.03 [0.06, 16.46]; ARR=0 [-0.09, 0.09]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td>Alendronate vs. no osteoporosis drug treatment 2 retrospective cohort observational studies&lt;sup&gt;41, 51, 52&lt;/sup&gt; 3.8 yr (mean) and ≥6 yr</td>
<td>534 adults ≥60 yr with nonhip fracture (90% women)</td>
<td>ST or FS fracture DX codes without x-ray confirmation of AFF features (n=5 cases)</td>
<td>≥6 yr; HR=1.37 [0.22, 8.62]; ARR NA</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>Alendronate vs. raloxifene 1 retrospective cohort observational study&lt;sup&gt;45&lt;/sup&gt; 4 to 6 yr</td>
<td>8,354 women aged ≥50 yr from database of 1 hospital</td>
<td>ONJ DX codes with x-ray and pathology features (n=40 cases)</td>
<td>Higher risk: 3.8 yr; ST: 0.17% vs. 0.06%; HR=2.41 [1.78, 3.27]; ARR=0.11 [0.08, 0.15]</td>
<td>Low (RB, IM, LE)</td>
</tr>
<tr>
<td>Alendronate vs. raloxifene or calcitonin 1 retrospective cohort observational study&lt;sup&gt;52&lt;/sup&gt; Up to 6 yr</td>
<td>43,645 adults ≥50 yr (84% women) with recent hip or vertebral fracture now on osteoporosis drug treatment from national database</td>
<td>ONJ DX codes without x-ray or pathology review (n=46 cases)</td>
<td>HR=0.86 [0.44, 1.69]; ARR NA</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>Zoledronate vs. placebo 1 RCT&lt;sup&gt;15&lt;/sup&gt; 6 yr</td>
<td>2,000 PM women ≥65 with osteoporosis or osteopenia</td>
<td>SAE</td>
<td>No difference: OR=0.84 [0.70, 1.00]; ARR= -4 [-9, 0]</td>
<td>Low (IM)</td>
</tr>
<tr>
<td>Bisphosphonate‡ vs. no bisphosphonate 3 observational studies&lt;sup&gt;38, 39, 55&lt;/sup&gt; &gt;3 yr</td>
<td>~2.8 million (retrospective cohort) and 1,124 (case-control) adults aged ≥55 yr from national database (87% women cases and 52% women controls in cohort)</td>
<td>AFF with radiologic features (n=172 cases)</td>
<td>Higher risk: Cohort ≥4 yr: RR=126 [55, 288]; ARR NA Case-control 3-4 yr: OR=40 [17, 91]; ARR=NA</td>
<td>Low (RB, CO, LE)</td>
</tr>
<tr>
<td>Comparison # Studies by Design</td>
<td>Participant Characteristics</td>
<td>Harms</td>
<td>Relative and Absolute Risk Differences (95% CI)</td>
<td>Strength of Evidence* (Justification)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>-------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td># Studies by Design</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>analysis; 86% women in case-control analysis)</td>
<td></td>
<td></td>
<td>4-5 yr: OR=116 [58, 234]; ARR=NA  &gt;5 yr: OR=93 [66, 132]; ARR=NA</td>
<td></td>
</tr>
<tr>
<td>264 women aged ≥65 yr from national primary practice database (case-control)</td>
<td>ST or FS fracture DX codes without x-ray confirmation of AFF features (n=44 cases)</td>
<td>Higher risk:  &gt;3 yr: OR=9.46 [2.17, 41.3]; ARR=NA</td>
<td>Low (RB, LE)</td>
<td></td>
</tr>
<tr>
<td>6,644 women aged ≥50 yr with hip or femoral fracture from 8 hospital medical records databases (nested case-control)</td>
<td>AFF with radiologic features (n=196 cases)</td>
<td>Higher risk: Mean use 5.2 yr: OR=25.65 [10.74, 61.28]; ARR=NA</td>
<td>Low (RB, LE)</td>
<td></td>
</tr>
<tr>
<td>Current vs. past bisphosphonates‡</td>
<td>2 case-control observational studies† 3 yr</td>
<td>AFF with radiologic features (n=43 cases)</td>
<td>Higher risk with current bisphosphonate: HR=3.36 [1.77, 11.91] to 5.17 [2.0, 13.36]; ARR NA</td>
<td>Low (RB, LE)</td>
</tr>
<tr>
<td>1,855 women aged ≥68 yr from a provincial database</td>
<td>ST or FS fracture DX codes without x-ray review (n=325 cases)</td>
<td>Higher risk with current bisphosphonate: 3-5 yr: OR=1.59 [0.80, 3.15]; ARR=NA &gt;5 yr: OR=2.74 [1.25, 6.02]; ARR=NA</td>
<td>Low (RB, IM)</td>
<td></td>
</tr>
<tr>
<td>Bisphosphonates† vs. pooled raloxifene or calcitonin 1 retrospective cohort observational study‡ ≥3 yr</td>
<td>4,097 Medicare beneficiaries (97% women)</td>
<td>ST or FS fracture DX codes without x-ray confirmation of AFF features (n=34 cases)</td>
<td>3-5 yr: HR=1.20 [0.55, 2.61]; ARR=0.1 [0.3, 0.5] &gt;5 yr: HR=2.02 [0.41, 10.0]; ARR=0.1 [-0.1, 0.4]</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>Denosumab†† vs. placebo 1 RCT18 4 yr</td>
<td>365 PM women with osteopenia or osteoporosis by BMD</td>
<td>SAE</td>
<td>RR=1.64 [0.69, 3.88]; ARR=7 [-3, 17]</td>
<td>Insufficient (RB, IN, h-IM)</td>
</tr>
<tr>
<td>Raloxifene vs. placebo 1 RCT with 1 CCT extension16, 17, 26-37 4 to 8 yr</td>
<td>6,828 PM women with osteoporosis by BMD or RVF</td>
<td>SAE</td>
<td>No difference: 8 yr: RR=0.93 [0.86, 1.00]**; ARR=-3 [-6, 0]</td>
<td>Low (RB, IM)</td>
</tr>
<tr>
<td>Raloxifene vs. no treatment 1 retrospective cohort observational study‡‡</td>
<td>19,324 adults (85% women) exposed to raloxifene or no osteoporosis drug general population controls from national database</td>
<td>ST or FS fracture DX codes without x-ray confirmation of AFF features (n=25 cases)</td>
<td>ST: HR=1.06 [0.34, 3.32]; ARR 0.04 [-0.06, 0.14] FS: HR=0.82 [0.21, 3.20]; ARR 0.01 [-0.07, 0.09]</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>Comparison # Studies by Design Treatment Duration</td>
<td>Participant Characteristics</td>
<td>Harms</td>
<td>Relative and Absolute Risk Differences (95% CI)</td>
<td>Strength of Evidence* (Justification)</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------------------------------</td>
<td>-------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>52 3.8 yr (mean)</td>
<td></td>
<td>ONJ DX codes without x-ray or pathology review (n=2 cases)</td>
<td>2 cases, only in control group</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
</tbody>
</table>

Abbreviations: AFF=atypical femoral fracture; ARR=absolute risk reduction; BMD=bone mineral density; CCT=controlled clinical trial; CI=confidence intervals; CO=consistent; DX=diagnosis; FS=femoral shaft; h-IM=highly imprecise; HR=hazard ratio; IM=imprecise; IN=indirect; LE=large effect; NA=not available (data not reported); ONJ=osteonecrosis of the jaw; OR=odds ratio; PM=postmenopausal; RB=medium risk of bias; RCT=randomized controlled trial; RR=risk ratio; RVF=radiographic vertebral fracture; SAE=serious adverse event; ST=subtrochanteric

*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.’
†Because the higher adjusted incidence rates in the alendronate group (0.15%) compared with the raloxifene-calcitonin group (0.08%) suggested a possibly increased risk, we manually recalculated the estimate of effect and found RR 1.20 (95% CI 0.59, 2.56). Authors were contacted for clarification, but did not reply.
‡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.
††Analyses pooled all participants initially assigned to denosumab, which included both those who received long-term and short-term denosumab.
Osteoporosis Drug Holidays

Effect

In postmenopausal women who previously received 3-5 years of bisphosphonate, two trials compared continued versus discontinued alendronate for 5 more years, and one compared continued versus discontinued zoledronate for 3 more years.\textsuperscript{55-57} None found a reduction in incident nonvertebral fractures (Table C).

However, these trials collectively suggested a reduction in incident vertebral fractures. One enrolled women who previously received 5 years of alendronate in the active treatment arm of a trial for osteopenia or osteoporosis and a subsequent extension, and reported that alendronate continuation for 5 years versus placebo (drug holiday) reduced incident clinical vertebral fractures (HR 0.45 [95% CI 0.24, 0.85]) (moderate SOE), but not incident radiographic vertebral fractures (HR 0.86 [95% CI 0.60, 1.22]) (moderate SOE).\textsuperscript{55} A second trial enrolled women who previously received 3 years of zoledronate in the active treatment arm of a trial for osteoporosis, and reported that zoledronate continuation for 3 years versus placebo (drug holiday) reduced incident radiographic vertebral fractures (HR 0.51 [95% CI 0.26, 0.95) (low SOE), but that evidence was insufficient about incident clinical vertebral fracture.\textsuperscript{58} In a third trial that enrolled women who previously received 5 years of alendronate in the active treatment arm of a trial for osteoporosis and a subsequent extension, and then nonrandomly assigned them to alendronate continuation for 2 years and 5 years versus placebo (drug holiday), evidence was insufficient to draw conclusions about differences in risk of incident clinical vertebral fractures or incident radiographic vertebral fractures.\textsuperscript{57-59} Similarly, we could not draw conclusions from a small, 4-year denosumab dose-finding trial, because fracture results were pooled between the denosumab continuation and discontinuation treatment arms.\textsuperscript{18}

In women who previously received 3-5 years of bisphosphonate treatment, continued bisphosphonate treatment for an additional 3-5 years was associated with stable or slightly decreased hip BMD, whereas women assigned to discontinue treatment (drug holiday) had significantly larger declines in hip BMD. A 4-year denosumab trial reported that compared to baseline, hip and spine BMD were most increased in women assigned denosumab for 4 years, back to pretreatment baseline in women assigned denosumab for 2 years followed by discontinuation for 2 years, and immediately increased in women assigned denosumab for 2 years, placebo for 1 year, then denosumab for 1 year.

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

In post hoc analyses, the effect of alendronate continuation versus discontinuation (drug holiday) on risk of incident clinical fractures, which did not differ overall, did not vary as a function of baseline BMD or radiographic vertebral fracture status.\textsuperscript{55,60} Further post hoc subgroup analyses suggested that in women without a prevalent radiographic vertebral fracture, continued alendronate versus discontinuation reduced risk of incident nonvertebral fractures in women with osteoporotic BMD but not in those with osteopenia. However, risk for incident nonvertebral fractures between alendronate continuation and discontinuation appeared not to differ in women with prevalent radiographic vertebral fractures. Further, risk of incident vertebral fractures appeared no different between alendronate continuation and discontinuation groups, regardless of baseline BMD or radiographic vertebral fracture status. The single positive
outcome may have been due to chance. We found no evidence about possible modifiers of the effect of continuing any other osteoporosis drug treatment versus discontinuation on risk of incident fracture.

Harms

Trials of alendronate and zoledronate continuation versus discontinuation reported no difference between treatment groups in risk of serious adverse events (Table D).55, 56, 58-62 Too few cases of AFF with confirmed radiologic features, subtrochanteric or femoral shaft fractures without confirmed AFF features, or ONJ, occurred in these trials to draw conclusions about differences in their risk between treatment continuation and discontinuation groups.50, 58 One retrospective cohort study reported that incidence of AFF was significantly higher in bisphosphonate users (99% alendronate) who continued versus discontinued use (0.15% vs. 0.03%; estimated OR 6.03 [95% CI 1.87, 19.42]).63 However, this analysis did not radiologically confirm AFF diagnoses and did not describe accounting for potentially confounding variables. Though atrial fibrillation appeared more frequently with zoledronate continuation versus discontinuation, the absolute number of events was low and possible differences between treatment groups were not statistically significant.58, 59 It was not possible to draw conclusions about differences in harms between the denosumab continuation and discontinuation arms in one trial reporting because harms results for these two groups were pooled.18

Variation in Harms 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 C. Evidence on effects of osteoporosis drug continuation versus discontinuation* on incident fractures

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Participant Characteristics</th>
<th>Incident Fracture Outcome</th>
<th>Relative and Absolute Risk Differences (95% CI)</th>
<th>Strength of Evidence†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alendronate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuation vs. discontinuation</td>
<td>1,099 PM women previously received alendronate 5 yr for osteopenia or osteoporosis (T-score ≤-1.6)</td>
<td>CF</td>
<td>No difference: RR=0.93 [0.71, 1.21]; ARR=-1 [-6, 4]</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td>(AL x 10 yr vs. AL x 5 yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVF</td>
<td></td>
<td></td>
<td>No difference: RR=1.00 [0.76, 1.32]; ARR=-0.1 [-5, 5]</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td></td>
<td></td>
<td>RR=1.02 [0.51, 2.10]; ARR=0 [-2, 2]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVF</td>
<td></td>
<td></td>
<td>Lower risk with continuation: RR=0.45 [0.24, 0.85]; ARR=-3 [-5, -0.5]</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RVF</td>
<td></td>
<td></td>
<td>No difference: RR=0.86 [0.60, 1.22]; ARR=-1 [-5, 2]</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zoledronate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuation vs. discontination</td>
<td>350 PM women previously received alendronate 5 yr for osteoporosis (T-score ≤-2.5) (n=350 for A7 vs. A5/P2; n=247 for A10 vs. A7/P3)</td>
<td>NVF</td>
<td>A7 vs. A5/P2: RR=0.87 [0.40, 1.91]; ARR=-1 [-7, 5]</td>
<td>A7 vs. A5/P2: Insufficient (h-IM)</td>
</tr>
<tr>
<td>(Z x 2 yr vs. Z x 1 yr)</td>
<td></td>
<td></td>
<td></td>
<td>A10 vs. A7/P3: Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td>1 RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>A10 vs. A7/P3: RR= 0.81 [0.38, 1.71]; ARR=-2 [-11, 6]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVF</td>
<td></td>
<td></td>
<td>A7 vs. A5/P2: RR= 0.92 [0.40, 2.10]; ARR=-1 [-6, 5]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RVF</td>
<td></td>
<td></td>
<td>AL10 vs. AL5/P5: RR=1.40 [0.52, 3.74]; ARR=2.6 [-4.6, 9.9]</td>
<td>Insufficient (RB, h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zoledronate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuation vs. discontination</td>
<td>379 PM women with osteopenia</td>
<td>CF</td>
<td>RR=1.37 [0.39, 4.78]; ARR=1 [-2, 4]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td>(Z x 2 yr vs. Z x 1 yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Zoledronate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continuation vs. discontinuation</td>
<td>1,233 PM women previously received zoledronic acid 3 yr for osteoporosis by BMD or RVF</td>
<td>CF</td>
<td>No difference: HR=1.04 [0.71, 1.54]; ARR NA</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td>(Z x 6 yr vs. Z x 3 yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 RCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVF</td>
<td></td>
<td></td>
<td>No difference: HR= 0.99 [0.7, 1.5]; ARR=-0.3 [-3, 3]</td>
<td>Moderate (IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip</td>
<td></td>
<td></td>
<td>HR= 0.90 [0.33, 2.49]; ARR=-0.2 [-1, 1]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVF</td>
<td></td>
<td></td>
<td>HR=1.81 [0.53, 6.2]; ARR NA</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RVF</td>
<td></td>
<td></td>
<td>Lower risk with continuation: OR=0.51 [0.26, 0.95]; ARR=-3 [-6, -1]</td>
<td>Low (h-IM)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF</td>
<td></td>
<td></td>
<td>HR=1.11 [0.45, 2.73]; ARR=1 [-7, 10]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td>Comparison # Studies by Design Treatment Duration</td>
<td>Participant Characteristics</td>
<td>Incident Fracture Outcome</td>
<td>Relative and Absolute Risk Differences (95% CI)</td>
<td>Strength of Evidence†</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Zoledronate continuation vs. discontinuation (Z x 9 yr vs. Z x 6 yr followed by PBO x 3 yr)</strong> 1 RCT&lt;sup&gt;19&lt;/sup&gt;</td>
<td>190 PM women previously received zoledronic acid 6 yr for osteoporosis by BMD or RVF</td>
<td>RVF</td>
<td>OR=0.58 [0.13, 2.55]; ARR= -2 [-8, 4]</td>
<td>Insufficient (h-IM)</td>
</tr>
<tr>
<td><strong>Denosumab continuation vs. discontinuation (D x 4 yr vs. D x 2 yr followed by PBO x 2 yr)</strong> 1 RCT&lt;sup&gt;18&lt;/sup&gt;</td>
<td>314 PM women with osteopenia or osteoporosis by BMD</td>
<td>CF</td>
<td>No numerical data</td>
<td>Insufficient (no data)</td>
</tr>
</tbody>
</table>

**Abbreviations:** AL=alendronate; ARR=absolute risk reduction; BMD=bone mineral density; CF=clinical fracture; CI=confidence intervals; CVF=clinical vertebral fracture; D=denosumab; h-IM=highly imprecise; HR=hazard ratio; IM=imprecise; NA=not available (no data reported); NVF=nonvertebral fracture; OR=odds ratio; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; RB=medium risk of bias; RR=risk ratio; RVF=radiographic vertebral fracture; Z=zoledronate

*Discontinuation ≥ 1 year after prior treatment ≥1 year.
†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.’

**Table D. Evidence on harms of osteoporosis drug continuation versus discontinuation**

<table>
<thead>
<tr>
<th>Comparison # Studies by Design Treatment Duration</th>
<th>Participant Characteristics</th>
<th>Harms</th>
<th>Relative and Absolute Risk Differences (95% CI)</th>
<th>Strength of Evidence† (Justification)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate continuation vs. discontinuation (AL x 10 yr vs. AL x 5 yr followed by PBO x 5 yr) 1 RCT&lt;sup&gt;15&lt;/sup&gt;</td>
<td>1,099 PM women previously received alendronate 5 yr for osteopenia or osteoporosis (T-score &lt;-1.6)</td>
<td>SAE</td>
<td>Stated as no difference, but no data provided</td>
<td>Insufficient (no data)</td>
</tr>
<tr>
<td>Subtrochanteric or femoral shaft fracture DX with rare x-ray review (n=3 cases)</td>
<td>HR=1.33 [0.12, 14.67]; ARR= -0.1 [-0.5, 0.7]</td>
<td>Insufficient (h-IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ONJ not defined (n=0 cases)</td>
<td>No cases in either group</td>
<td>Insufficient (h-IM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison # Studies by Design Treatment Duration</td>
<td>Participant Characteristics</td>
<td>Harms</td>
<td>Relative and Absolute Risk Differences (95% CI)</td>
<td>Strength of Evidence† (Justification)</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>------------------------------</td>
<td>-------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Alendronate continuation vs. discontinuation (AL x 7 yr [A7] vs. AL x 5 yr followed by PBO x 2 yr [A5/P2]; AL x 10 yr [A10] vs. AL x 7 yr + PBO x 3 yr [A7/P3]) 1 RCT</td>
<td>350 PM women previously received alendronate 5 yr for osteoporosis (T-score ≤-2.5)</td>
<td>SAE</td>
<td>A7 vs. A5/P2: RR= 1.05 [0.57, 1.96]; ARR=1 [-7, 8] A10 vs. A7/P3: RR= 1.21 [0.75, 1.96]; ARR=5 [-7, 16]</td>
<td>A7 vs. A5/P2: Insufficient (h-IM) A10 vs. A7/P3: Insufficient (RB, IM)</td>
</tr>
<tr>
<td>Bisphosphonate continuation vs. discontinuation (Continued BP 3.5 yr [mean] [persistent group] or 4.1 yr [mean] [nonpersistent group] vs. BP holiday 3.1 yr [mean]) 1 retrospective cohort observational study</td>
<td>39,502 women aged ≥45 yr with ≥3 yr of prior ≥50% adherent BP use (99% alendronate)</td>
<td>“AFF” (not defined) (n=47 cases)</td>
<td>Higher risk with bisphosphonate (alendronate) continuation: Pooled continuation groups 0.15% (44/28005) vs. discontinuation 0.03% (3/11497) OR=6.03 [1.87, 19.42]; ARR=0.13 [0.08, 0.19]</td>
<td>Low (RB, IM, LE)</td>
</tr>
<tr>
<td>Zoledronate continuation vs. discontinuation (Z x 2 yr vs. Z x 1 yr followed by PBO x 1 yr) 1 RCT</td>
<td>379 PM women with osteopenia</td>
<td>SAE</td>
<td>No difference: RR=0.91 [0.50, 1.67]; ARR= -1 [-7, 5]</td>
<td>Low (h-IM)</td>
</tr>
<tr>
<td>Zoledronate continuation vs. discontinuation (Z x 6 yr vs. Z x 3 yr followed by PBO x 3 yr) 1 RCT</td>
<td>1,233 PM women previously received zoledronic acid 3 yr for osteoporosis by BMD or RVF</td>
<td>SAE</td>
<td>No difference: RR=1.14 [0.96, 1.36]; ARR=4 [-1, 9]</td>
<td>Low (IM)</td>
</tr>
<tr>
<td>Zoledronate continuation vs. discontinuation (Z x 9 yr vs. Z x 6 yr followed by PBO x 3 yr) 1 RCT</td>
<td>190 PM women previously received zoledronic acid 6 yr for osteoporosis by BMD or RVF</td>
<td>SAE</td>
<td>No difference: RR=0.86 [0.54, 1.36]; ARR= -3 [-16, 9]</td>
<td>Low (IM)</td>
</tr>
<tr>
<td>Denosumab continuation vs. discontinuation (D x 4 yr vs. D x 2 yr followed by PBO x 2 yr) 1 RCT</td>
<td>314 PM women with osteopenia or osteoporosis by BMD</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; BP=bisphosphonate; CI=confidence intervals; D=denosumab; DX=diagnosis; h-IM=highly imprecise; HR=hazard ratio; IM=imprecise; LE=large effects; ONJ=osteonecrosis of the jaw; OR=odds ratio; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; RB=medium risk of bias; RR=risk ratio; SAE=serious adverse event; Z=zoledronate

*Discontinuation ≥1 year after prior treatment ≥1 year.
†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 long-term placebo-controlled trials, alendronate for 4 years reduced incident radiographic vertebral and nonvertebral fractures in women with osteoporosis, and zoledronate for 6 years reduced vertebral and nonvertebral fractures in women with osteopenia or osteoporosis. Observational studies suggested that long-term treatment with bisphosphonates as a class increased risk of AFF, ST/FSF, and ONJ, though these adverse events were rare. In women with osteoporosis, long-term raloxifene for 4 years reduced incident vertebral fractures, but not hip or nonvertebral fractures; long-term raloxifene also increased risk of deep vein thrombosis and pulmonary embolism. In women with unknown osteoporosis or osteopenia status, oral hormone therapy for 5-7 years reduced incident clinical and hip fractures compared with placebo, but increased risk of cardiovascular disease and cognitive impairment. Trials also showed that continuation of zoledronate or alendronate after 3-5 years of prior treatment versus discontinuation reduced some vertebral fracture outcomes but not others, did not reduce nonvertebral fractures, and observational data suggested that continuation of bisphosphonates as a class may increase risk of ST/FSF compared with discontinuation.

Whereas long-term treatment with alendronate reduced risk of incident clinical fractures compared with placebo in women with osteoporosis, it did not reduce fracture risk in women with osteopenia. Otherwise, risk of fracture with long-term alendronate versus placebo did not vary by history of prior fracture, World Health Organization Fracture Risk Assessment Tool (FRAX®) score, or pretreatment levels of bone turnover markers. Risk of incident fracture between long-term raloxifene and placebo did not vary as a function of age, baseline BMD, or history of prior fracture. Reduction in incident clinical fracture with oral hormone therapy compared with placebo appeared possibly greater in women aged 60-79 years than in younger women, though similar results were not found for hip fracture, suggesting possible chance findings. We found no information about possible modifiers of fracture risk with long-term zoledronate treatment.

Our findings have several clinical implications. In women with osteoporosis, indications for long-term raloxifene may be limited, because it only reduces vertebral fractures, while both long-term alendronate and zoledronate also reduced nonvertebral fractures. While the effects of long-term alendronate and zoledronate appear roughly similar in older women with osteoporosis, only long-term zoledronate also reduced nonvertebral fractures in women with osteopenia. It is unclear if these possibly discrepant findings are explained in part by differences in study populations (e.g., the zoledronate population was older). Unfortunately, there are no eligible long-term trials that directly compare alendronate and zoledronate in older women with osteopenia. While oral hormone therapies for 5-7 years lowered both clinical and hip fractures in women not selected to be at high fracture risk, and might be expected to have larger effects in those with osteoporosis, because fracture benefits were offset by risk of serious harms, these agents are not likely to be a viable option for long-term osteoporosis treatment. However, it is unknown whether a lower dose or different route of administration of hormone therapy would have a more favorable balance of fracture benefits to harms. In patients who have completed 3-5 years of bisphosphonate treatment, continued alendronate or zoledronate versus discontinuation each reduced one of two measures of incident vertebral fracture, but did not reduce nonvertebral fractures. Observational data suggested that long-term bisphosphonates increase risk of AFF, that risk likely increases with longer duration of treatment, and that these events are rare. Estimating the relative balance between benefits and harms, for every 1,000 women with osteoporosis
treated with alendronate for 4 years or with osteopenia or osteoporosis treated with zoledronate for 6 years compared with placebo, approximately 50 to 70 more will avoid an incident clinical fracture, while an additional 2 will experience a ST/FSF. Since only a minority of ST/FSF meet AFF criteria, the absolute number of additional AFF would be expected to be smaller. Analogously, for every 1,000 women previously treated for osteopenia or osteoporosis with 3 to 5 years of alendronate or zoledronate who continue bisphosphonate treatment another 3 to 5 years, compared with discontinuation, approximately 30 more will avoid an incident vertebral fracture, while an additional 1 will experience a ST/FSF. However, the inconsistency of the vertebral fracture results, the uncertainty around the outcome risk estimates, and the fact that relatively few ST/FSF meet AFF criteria, suggest that the ratio of these fracture benefits to AFF harms with bisphosphonate continuation could be either substantially larger or smaller. Evidence appeared less robust for ONJ, but suggested long-term bisphosphonate treatment also may increase risk of this outcome. Data from eligible studies did not identify clear patient, bone or drug characteristics that modify likelihood of fracture benefits or harms with long-term or continuing osteoporosis drug treatment.

Limitations

The available data limit this review in several ways. First, there were few unique trials of long-term osteoporosis drug treatment or of drug discontinuation, and only one trial that included a treatment arm involving osteoporosis drug discontinuation and subsequent osteoporosis drug resumption. We often identified only one trial for a given treatment comparison. Second, only two trials were designed with incident fracture as the primary outcome. Consequently, many studies had few incident clinical fractures, especially for hip fractures, and statistical power often was low to precisely estimate differences in their risk between treatment interventions. Third, all trials were conducted in generally health, usually white, postmenopausal women, limiting their generalizability. Further, most of the trials were conducted in populations selected for osteoporosis or osteopenia by BMD or radiographic vertebral fracture criteria. Generalizability of results to populations who have other reasons for heightened fracture risk is unknown. Fourth, observational studies investigating the association between treatment and risk of AFF or ONJ had marked methodologic differences that likely affected the specificity of these outcomes and the associated risk estimates. Major differences included the definitions of the cases (e.g., whether or not fractures were defined using American Society for Bone and Mineral Research [ASBMR] radiographic AFF features) and noncase controls, drug therapy exposure and control groups, and adjustment for possible confounding. Fifth, reporting on harms was sparse and inconsistent between studies, limiting confidence around harms risk estimates, and raising concerns about possible reporting bias. Sixth, few studies reported information about possible effect modifiers of drug treatment outcomes. These analyses were almost entirely post hoc, often did not test for interactions between potential effect modifiers, treatment assignment, and treatment outcomes, and did not test for multiple testing, raising the likelihood of type 1 errors. Seventh, there were no eligible long-term fracture trials for several U.S. Food and Drug Administration (FDA) approved osteoporosis drugs, including risedronate, ibandronate, teriparatide, and abaloparatide; long-term fracture data for denosumab came from only one small trial in which the pooling of fracture data from different intervention groups made interpretation impossible. There also were no eligible trials of sequential treatment, such as with an anabolic followed by an anti-resorptive, or denosumab followed by bisphosphonate. Finally, there were no usable data comparing different durations of osteoporosis drug holidays.
Research Needs

Future trials of long-term osteoporosis drug treatment and osteoporosis drug continuation versus discontinuation should be designed with adequate statistical power to assess risks of clinical fracture endpoints, including hip fractures, the fracture type with the greatest risk of morbidity and mortality. Broader trial samples that include men, nonwhite women, more adults with comorbidities, and adults aged 80 years and older are needed to improve generalizability. Future long-term trials should evaluate sequential osteoporosis drug treatment, including comparisons of anabolic therapy followed by antiresorptive therapy, and denosumab followed by bisphosphonate therapy, with both compared with continuous long-term antiresorptive therapy. Trials should compare continuous long-term osteoporosis drug treatment to different osteoporosis drug holiday durations, with or without restarting osteoporosis drug therapy, and possibly with repeating cycles of osteoporosis drug therapy alternating with drug holidays. Future studies should systematically collect, analyze and report harms data. Randomized trials will continue to have limited statistical power to estimate the risk of rare treatment harms such as AFF and ONJ, so observational studies will be essential for examining these outcomes. These observational studies should use consensus case definitions, standard non-case and exposure controls, cohort designs to estimate incidence rates, and adequate statistical adjustment to reduce the effects of confounding by indication and selection bias. Observational studies may provide insights about the benefits and harms of drug holidays of different durations and about patient and treatment characteristics that predict which patients are likely to benefit or be harmed by treatment continuation versus discontinuation. Future trials and observational studies should pre-specify analyses to investigate possible effect modifiers of benefits and harms of long-term osteoporosis drug treatment and drug holiday outcomes. Among other factors, these should include age, and BMD and bone markers before and during long-term treatment or drug holidays. Patient-level data from osteoporosis drug trials on the associations of early treatment changes in BMD and bone turnover markers with risk of incident fractures may improve understanding of the potential use and limitations of these measures as surrogates for incident fracture.

Conclusions

Only alendronate, zoledronate, and oral hormone therapy reduced nonvertebral fractures with long-term treatment. However, for all these agents, these fracture benefits were limited to mostly older, postmenopausal women. They were further limited to women with osteopenia or osteoporosis for zoledronate, and to women with osteoporosis for alendronate. Absolute reductions in clinical fractures with long-term bisphosphonates appeared far greater than absolute increases in risk of AFF and ONJ with these treatments. However, reductions in hip fracture with long-term oral hormone therapy appear offset by risk of serious harms. In patients with prior osteoporosis drug treatment, continued treatment appeared to reduce vertebral fractures but not nonvertebral fractures, and may increase risk of AFF. While fracture benefits of continued osteoporosis drug treatment versus drug holiday numerically appeared to outweigh these risks, the more limited morbidity prevented and greater uncertainty about the outcome measures and risk estimates require further investigation to better inform clinical decisions about continuing treatment. This research should include examination of how these benefits and risks vary as a function of patient, bone, and drug treatment characteristics (e.g., age, sex, comorbidity, pre-drug holiday BMD, duration of prior osteoporosis drug treatment). Future
modeling studies also may incorporate probabilities of experiencing fracture-related morbidity to help patients weigh trade-offs of treatment more easily.
References


70. 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]. American Society for Bone and Mineral Research Annual Meeting, September 28-October 1, Montréal, Québec, Canada. 2018.
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.\textsuperscript{1} 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.\textsuperscript{2} Considering BMD at either the femoral neck or lumbar spine, and extrapolating this definition to men and nonwhite women, osteoporosis affects more than 10 million U.S. adults aged 50 years or older.\textsuperscript{3} About 2 million U.S. adults experience an osteoporotic or other low- or no-trauma fracture each year, including about 300,000 with hip fractures.\textsuperscript{4} Clinically recognized fractures at many skeletal sites cause pain, disability, and impaired quality of life. Hip and clinical vertebral fractures, specifically, are associated with an increased risk of mortality.\textsuperscript{5, 6} Incident radiographic vertebral fractures, a commonly reported outcome in 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,\textsuperscript{7-12} usually because of back pain. Even in individuals whose new radiographic vertebral fracture is not diagnosed, new or worsened back pain is more common than in individuals without new radiographic vertebral fractures.\textsuperscript{12, 13} Because risk of most fractures rises steeply with age, and the population is aging, fracture burden is projected to increase in coming decades.

In short-term (18 to 36 months) randomized controlled trials for osteoporosis treatment, U.S. Food and Drug Administration approved (Table 1) oral and intravenous bisphosphonates (alendronate, zoledronate, risedronate, ibandronate), denosumab, teriparatide, and abaloparatide have lowered risk of both vertebral and nonvertebral fractures; and several bisphosphonates (alendronate, zoledronate, risedronate) and denosumab also have lowered risk of hip fractures.\textsuperscript{14, 15} Despite the evidence on the efficacy of short-term osteoporosis drug treatment for reducing fracture risk, there are limits to this evidence. Evidence on 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. Data are far more limited in postmenopausal women without osteoporosis, even in those with heightened fracture risk because of low bone mass (i.e., osteopenia) combined with other risk factors, such as falls or a high WHO Fracture Risk Assessment Tool [FRAX®]\textsuperscript{16} score.\textsuperscript{17} Short-term osteoporosis drug treatment also may be associated with side effects. Oral bisphosphonates increase risk of upper gastrointestinal symptoms and IV bisphosphonates may increase risk of atrial fibrillation; 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.\textsuperscript{14} Observational studies suggest that bisphosphonates and denosumab are associated with rare atypical femoral fractures (AFF) and osteonecrosis of the jaw.\textsuperscript{18}

Compared to what is known about the effects of short-term osteoporosis drug treatment, the benefits and harms of long-term (>3 years) osteoporosis drug treatment are far less clear, including among individuals with osteopenia, though heightened risk of AFF is a concern.\textsuperscript{19} 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 for fractures. Early treatment changes in BMD and, to a lesser extent, in bone turnover markers,\textsuperscript{20} may predict short-term risk of incident fractures. It is
unclear, however, whether changes in these measures will predict a high enough proportion of the anti-fracture efficacy of osteoporosis drug treatments to be considered adequate fracture 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.

Beyond any overall effect of long-term osteoporosis drug treatment, there may be important differences in both fracture efficacy and risk of harms based on patient, bone, and osteoporosis drug characteristics (i.e., possible effect modifiers). Better understanding of these factors may allow prescribers and patients to make more informed treatment decisions to maximize benefit (e.g., reduced fracture risk) while minimizing harms.

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 to periodically discontinue bisphosphonate therapy. 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. In contrast to bisphosphonates, drug holidays are not recommended after denosumab therapy; bone turnover and bone loss increase rapidly after denosumab discontinuation, and case series and post hoc data suggest a possible small post-treatment increase in risk of radiographic vertebral fractures.

Uncertainties about the most appropriate use of long-term osteoporosis drug treatment and of osteoporosis drug holidays spurred the scheduling of a National Institutes of Health Office of Disease Prevention Pathways to Prevention program workshop on this topic. The goal of the workshop was to present an evidence-based synthesis of the pertinent research base, identify research gaps, and suggest future research needs to assist patients, clinicians, and other healthcare decision makers. To inform this workshop and clinicians and osteoporosis researchers more generally, we conducted this systematic review to address four questions. First, what are the effects of long-term osteoporosis drug therapy (>3 years) versus control on risk for incident clinical fractures, incident radiographic vertebral fractures, change in BMD, and harms? Second, do the effects of long-term osteoporosis drug treatment on these outcomes vary as a function of patient, bone, or osteoporosis drug characteristics? Third, among individuals receiving osteoporosis drug treatment to prevent fracture, what are the effects of continuing versus at least temporarily discontinuing treatment on risks for incident fractures and harms? Fourth, do outcomes of osteoporosis drug treatment continuation versus discontinuation vary as a function of patient, bone, or osteoporosis drug characteristics? These questions were refined and formalized as the Key Questions listed below.
<table>
<thead>
<tr>
<th>Drug Class</th>
<th>Drug Name</th>
<th>Delivery Route</th>
<th>Dosing Frequency</th>
<th>Long-Term</th>
<th>Mechanism in Bone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphosphonate</td>
<td>Alendronate</td>
<td>Oral</td>
<td>Daily, weekly</td>
<td>&gt;3 years</td>
<td>Common mechanism for nitrogen bisphosphonates: preferentially adhere to bone surface at sites of high bone turnover, there inhibiting a specific osteoclast intracellular enzyme, causing osteoclast death and reduced bone resorption; bind to bone mineral and continue to be systemically released for years after drug discontinuation, continuing some suppression of bone turnover and reduction of bone resorption</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Ibandronate</td>
<td>Oral, IV</td>
<td>Daily, weekly, monthly</td>
<td>&gt;3 years</td>
<td>Common mechanism for nitrogen bisphosphonates</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Risedronate</td>
<td>Oral</td>
<td>Daily, weekly</td>
<td>&gt;3 years</td>
<td>Common mechanism for nitrogen bisphosphonates</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Zoledronate</td>
<td>IV</td>
<td>Every 12 to 18 months</td>
<td>&gt;3 years</td>
<td>Common mechanism for nitrogen bisphosphonates</td>
</tr>
<tr>
<td>Biologic</td>
<td>Denosumab</td>
<td>SC</td>
<td>6 months</td>
<td>&gt;3 years</td>
<td>A monoclonal antibody to RANKL that binds it and prevents its interaction with RANK on osteoclasts, inhibiting osteoclast formation, function, and survival, and thereby decreasing bone resorption; suppression of bone turnover and resorption is rapidly reversed after drug discontinuation</td>
</tr>
<tr>
<td>PTH related anabolic</td>
<td>Teriparatide</td>
<td>SC</td>
<td>Daily</td>
<td>Not used &gt;2 years</td>
<td>Binds specifically to PTH receptors. With intermittent (i.e., once daily) rather than continuous exposure, stimulates bone formation more than bone resorption, leading to net anabolic effects on bone</td>
</tr>
<tr>
<td>PTH related anabolic</td>
<td>Abaloparatide</td>
<td>SC</td>
<td>Daily</td>
<td>Not used &gt;2 years</td>
<td>In bone, acts as agonist at PTH receptors. Daily rather than continuous exposure stimulates bone formation more than bone resorption, leading to net anabolic effects on bone</td>
</tr>
<tr>
<td>SERM</td>
<td>Raloxifene</td>
<td>Oral</td>
<td>Daily</td>
<td>&gt;3 years</td>
<td>Binds to nuclear estrogen receptors in bone and other tissues, acting as agonist in some tissues and antagonist in others. Is agonist in bone, decreasing bone resorption.</td>
</tr>
<tr>
<td>Estrogen and Progestin combination products</td>
<td>Multiple*</td>
<td>Oral, transdermal, transvaginal</td>
<td>Daily</td>
<td>&gt;3 years</td>
<td>Estrogens bind estrogen receptors α and β, acting as agonists in estrogen-sensitive tissues. In bone, they decrease bone resorption. Progestins may decrease nuclear estrogen receptors and suppress epithelial DNA synthesis in endometrial tissue.</td>
</tr>
</tbody>
</table>
### Scope and Key Questions

**Key Questions**

Key Question 1. Among men and postmenopausal women aged ≥50 years with osteoporosis\(^a\) or osteopenia/low bone mass,\(^b\) 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\(^a\) or osteopenia/low bone mass,\(^b\) 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?\(^c\)

Key Question 3. Among men and postmenopausal women aged ≥50 years with osteoporosis\(^a\) or osteopenia/low bone mass,\(^b\) 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\(^a\) or osteopenia/low bone mass,\(^b\) 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?\(^c\)

Key Question 5. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy (≥1 year) started for osteoporosis\(^a\) or osteopenia/low bone mass\(^b\) 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\(^a\) or osteopenia/low bone mass\(^b\) 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?\(^c\)

Key Question 7. Among men and postmenopausal women aged ≥50 years currently receiving drug therapy (≥1 year) started for osteoporosis\(^a\) or osteopenia/low bone mass\(^b\) 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\(^a\) or osteopenia/low bone mass\(^b\) to prevent fracture, does risk of harms associated with osteoporosis drug treatment holidays vary as a function of patient, bone, or osteoporosis drug characteristics?\(^c\)

\(^a\)Osteoporosis defined by hip or lumbar spine dual-energy x-ray absorptiometry (DXA) BMD T-score < -2.5, past clinical hip or vertebral fracture, or prevalent radiographic vertebral fracture.

\(^b\)Osteopenia/low bone mass defined by hip or lumbar spine DXA BMD T-score < -1.0 and > -2.5.

\(^c\)Patient characteristics (age, sex, race, osteoporosis status\(^a\), fracture history [clinical fractures, radiographic vertebral fractures], calculated fracture risk [e.g. FRAX\(^b\)], comorbid conditions); Bone characteristics (BMD, biomarkers); Osteoporosis drug characteristics (dose, frequency, treatment duration, delivery route).


**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>
</table>
| **KQ 1: Long-term treatment efficacy** | Men and PM women aged >50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment. | Osteoporosis drug treatment (see Table 1)                                    | Placebo, active control | Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF)  
Intermediate:  
Primary: Incident radiographic vertebral fracture  
Secondary: DXA BMD change | >3 years | Any     |
| **KQ 2: Effect modifiers of long-term treatment efficacy** | Men and PM women aged ≥50 years with osteoporosis* or osteopenia/low bone mass† being studied for fracture prevention treatment. | 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 & femoral neck DXA BMD) and biochemical markers of bone turnover (CTX, NTX, P1NP, BSAP)  
*Osteoporosis drug characteristics:* dose, frequency, treatment duration, delivery route | Placebo, active control | Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF)  
Intermediate:  
Primary: Incident radiographic vertebral fracture | >3 years | Any     |
<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>
</table>
| **KQ 3:** Long-term treatment harms | 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. | Osteoporosis drug treatment (see Table 1) | Placebo, no treatment, active control | See Table 3 below for possible harms outcomes evaluated | >3 years | Any |
| **KQ 4:** Effect modifiers of long-term treatment harms | 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. | List of possible effect modifiers of harms with long-term treatment is the same as the possible effect modifiers of incident fractures with long-term treatment detailed above for KQ 2. | Placebo, no treatment, active control | See Table 3 below for possible harms outcomes evaluated | >3 years | Any |
<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 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 of prior osteoporosis drug treatment</td>
<td>Continued osteoporosis drug treatment after ≥1 year of 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 of prior osteoporosis drug treatment</td>
<td>Any</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KQ 6: Effect modifiers of 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>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 and biomarkers of bone turnover, including both measures prior to the drug holiday and during the drug holiday. Osteoporosis drug characteristics: pre-drug holiday agent/class, time between pre-holiday drug initiation and start of drug holiday, duration of drug holiday, post-drug holiday agent/class</td>
<td>Continued osteoporosis drug treatment after ≥1 year of prior osteoporosis drug treatment</td>
<td>Final: Incident clinical fracture (any, nonvertebral, hip, vertebral, nonhip nonvertebral, MOF) Intermediate: Primary: Incident radiographic vertebral fracture</td>
<td>≥1 year osteoporosis drug discontinuation after ≥1 year of prior osteoporosis drug treatment</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 7: Harms 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 of prior osteoporosis drug treatment</td>
<td>Continued osteoporosis drug treatment after ≥1 year of prior osteoporosis drug treatment</td>
<td>See Table 3 below for possible harms outcomes evaluated</td>
<td>&gt;1 year osteoporosis drug discontinuation after ≥1 year of prior osteoporosis drug treatment</td>
<td>Any</td>
</tr>
<tr>
<td>KQ 8: Effect modifiers of harms 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>List of possible effect modifiers of harms with osteoporosis drug treatment holidays is the same as the possible effect modifiers of incident fractures with drug holidays detailed above for KQ 6.</td>
<td>Continued osteoporosis drug treatment after ≥1 year prior osteoporosis drug treatment</td>
<td>See Table 3 below for possible harms outcomes evaluated</td>
<td>&gt;1 year osteoporosis drug discontinuation after ≥1 year prior osteoporosis drug treatment</td>
<td>Any</td>
</tr>
</tbody>
</table>

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.
Harms

Table 3 summarizes the harms that were evaluated by this review for possible association with long-term osteoporosis drug therapy.

**Table 3. Harms evaluated for possible association with long-term osteoporosis drug therapy**

<table>
<thead>
<tr>
<th>Drug Class</th>
<th>Harms Outcomes Evaluated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphosphonates</td>
<td>ONJ*, AFF*, atrial fibrillation*, heart attack, musculoskeletal pain, upper GI intolerance, esophageal cancer</td>
</tr>
<tr>
<td>Biologic (Denosumab)</td>
<td>ONJ*, AFF*, atrial fibrillation*, heart attack, musculoskeletal pain, upper GI intolerance, esophageal cancer, infection, fracture after stopping therapy</td>
</tr>
<tr>
<td>PTH related anabolic (recombinant PTH and PTH analog)</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), mild cognitive impairment, dementia, mortality</td>
</tr>
</tbody>
</table>

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

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

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

Figure 1. Analytic framework for long-term osteoporosis drug treatment and osteoporosis drug treatment holidays for fracture prevention

Abbreviations: BMD=bone mineral density; fx=fracture; KQ=key question; MOF=major osteoporotic fracture; PM=postmenopausal; yrs=years

*Osteoporosis defined by hip or lumbar spine dual-energy x-ray absorptiometry (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
Chapter 2. Methods

This comparative effectiveness review follows 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.26

Topic Refinement and Review Protocol

The National Institutes of Health (NIH) Office of Disease Prevention (ODP) Working Group, which included individuals from the ODP, National Institute on Aging and National Institute of Arthritis and Musculoskeletal and Skin Diseases, developed the first version of questions for this review. Questions were refined in collaboration with the NIH ODP Working Group, a NIH Content Area Expert Group, and a Technical Expert Panel. The resulting Key Questions (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 randomized controlled trials (RCTs), controlled clinical trials, 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 updated our electronic database search through October 2018 while the draft report was under 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>
<thead>
<tr>
<th>Category</th>
<th>Inclusion Criteria</th>
</tr>
</thead>
</table>
| **Study Population**          | *Included for all Key Questions:* Adults aged >50 years, including men and postmenopausal women  
Participants with osteoporosis (osteoporosis defined as hip or vertebral DXA BMD T-score ≤ -2.5, past hip or clinical 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.  
*Also included for Key Questions about rare harms (included in Key Questions 3, 4, 7 & 8):* Adults without osteoporosis/osteopenia or with unknown osteoporosis/osteopenia status being treated to prevent fractures.  
*Excluded for all Key Questions:* Studies focused on patients with cancer metastatic to bone or focused on drug effects on acute fracture healing.  
Studies focused on populations with known secondary causes of osteoporosis (e.g., solid or liquid organ transplant, spinal cord injury, exogenous glucocorticoids, hormone suppressive therapy, endogenous hypercortisolism, hyperparathyroidism, hyperthyroidism).  
Note that will include studies focused on populations with CKD, DM, or CVD. |
| **Study Objectives**          | To evaluate the efficacy and harms of long-term osteoporosis drug treatment (>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.                                                                                                           |
| **Study Design**              | *All Key Questions:* RCTs, CCTs  
*For assessment of harms (KQ 3, 4, 7 & 8):* Also included observational studies with contemporaneous human controls that employed methods to account for selection bias (adjusted for demographics and fracture risk [e.g., past fracture, BMD, or fracture risk calculator]).  
*For assessment of nonrare harms:* In addition to RCTs and CCTs, eligible observational studies were prospective cohort studies; sample size from these observational studies must have been >1000  
*For assessment of rare harms (i.e., AFF, ONJ, atrial fibrillation):* In addition to RCTs and CCTs, eligible observational studies were case-control, retrospective or prospective cohort, or administrative data studies; sample size from these observational studies must have been >100  
*Excluded:* Case reports, case series, and post-marketing reports were not eligible to address any Key Question.                                                                                                                                          |
| **Interventions**             | *Long-term treatment:* Drugs FDA approved for osteoporosis treatment or prevention (bisphosphonates, denosumab, teriparatide, abaloparatide, estrogen*, estrogen/progestin*, SERM, estrogen/SERM*)  
*Drug holiday:* Discontinuation of osteoporosis drug treatment for >1 year after >1 year of prior treatment                                                                                                                                                                         |
| **Comparisons**               | *For treatment efficacy and harms:* Placebo, active contemporaneous control  
*For osteoporosis drug treatment discontinuation (holiday)†:* Continued osteoporosis drug treatment                                                                                                                                                                                   |
| **Outcomes**                  | *Final health outcomes:* incident clinical fracture (e.g., any, nonvertebral, hip, clinical 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:* Possible harms outcomes are listed in Table 3.                                                                                                                                                                                                                       |
| **Possible effect modifiers (applicable only for KQs 2, 4, 6 & 8)** | *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., after 1 year of treatment) imaging (lumbar spine, total hip & femoral neck DXA BMD) and biochemical markers of bone turnover (CTX, NTX, P1NP, BSAP)  
*Osteoporosis drug characteristics:* For KQ 2 & 4: dose, frequency, treatment duration, delivery route; For KQ 6 & 8: pre-drug holiday agent/class, time between beginning pre-drug holiday drug and start of drug holiday, duration of drug holiday, post-drug holiday drug/class |
### Study Selection

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.

### Risk of Bias Assessment

Based on AHRQ guidance, we assessed risk of bias (ROB) of eligible studies in their design, analysis, and reporting. ROB was assessed by two independent investigators guided by a modified instrument we created for a previous AHRQ review (Appendix B). 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), (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 the reviewers’ collective informed opinion that the study estimate represented what the study would have found had it been perfectly designed, conducted, analyzed, and reported (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, atypical femoral fracture (AFF) with confirmed radiological features, subtrochanteric/femoral shaft fractures, osteonecrosis of the jaw (ONJ), and incident fracture after stopping osteoporosis drug treatment (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 Table 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 status (i.e., osteoporosis versus osteopenia/low bone mass versus normal, including how the study defined its eligible population), fracture history, calculator estimated fracture risk, bone mineral density (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). Studies that compared continuous long-term treatment with continuous control were included in the long-term osteoporosis drug treatment section of the report. Studies that compared osteoporosis drug continuation versus discontinuation were included in the drug holiday section of the report, even if the total duration of treatment in the continuation group was >3 years.

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 were 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, BSAP, drug dose, frequency, treatment duration and delivery route, 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 osteoporosis drug treatment (rebound fractures).

SOE ratings were based on five required domains: (1) study limitations; (2) directness; (3) consistency; (4) precision; and (5) reporting bias. Study limitations were rated as low, medium, or high, based on the collective ROB of the individual studies pertaining to an evidence base. Directness was rated as direct or indirect, based on the presence or absence of a single, direct link between intervention and outcome. Consistency was rated as consistent or inconsistent, based on the similarity of the direction and size of an effect between studies, or unknown when there was only a single study. Precision was rated as precise, imprecise, or highly imprecise, based on the degree of certainty around an estimate or a small number of fracture or harm events. An imprecise estimate was one for which the confidence interval was wide enough to include clinically distinct conclusions or there were few events, and a highly imprecise estimate was one for which confidence intervals were wide and there were few events, or the confidence intervals were wide enough to include both a large risk reduction and a large risk increase. 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 that were different than those described by population studies.
Peer Review and Public Commentary

AHRQ and NIH staff and an AHRQ associate editor reviewed the first draft of this report. After we revised the report based on these reviews, experts in osteoporosis, primary care, geriatrics, and systematic reviews were invited to provide external peer review. After we revised the report again based on these peer reviews, the next draft was posted on the AHRQ website for 4 weeks to elicit public comment. We then revised the report again based on these public comments, and documented all public comments, our responses and revisions in a disposition of comments report that we submitted with the final report for posting on the Effective Health Care website.
Chapter 3. Search Results

We identified 8,356 unique citations (Figure 2) from 1995 to October 2018 from bibliographic databases addressing osteoporosis and drug treatment. An initial title and abstract review excluded 7,026 publications that were not related to relevant drug treatments for patients with osteoporosis. Full texts of 1,330 publications from the electronic database searches and five publications identified by hand search were reviewed to determine final inclusion. Appendix E provides a list of publications excluded at full text review. We identified 61 unique references eligible for inclusion, which are summarized in Table 5. Risk of bias (ROB) assessments for these references can be found in Appendix C.

Of 48 publications with low or medium ROB, there were 35 randomized or controlled clinical trials (9 unique studies) and 13 controlled observational studies (11 unique studies) (Appendix C). Most publications were based on three randomized controlled trials 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 bone mineral density (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
8,356 references

Title and abstract review excluded
7,026 references

Hand search (reference lists of included studies)
5 references

Excluded
1,274 references

- Not available in English = 10
- Ineligible study design = 446
- Ineligible treatment duration = 355
- Ineligible population = 77
- No intervention of interest = 43
- No outcomes of interest = 343

Pulled for full text review
1,335 references

Eligible references = 61
(31 unique studies, including 20 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</td>
<td>Bisphosphonate</td>
<td>13 RCTs or CCTs(^{11-43}) (11 FIT/FLEX)</td>
<td>2 RCTs(^{48, 49}) (FIT) 4 observational(^{44-47})</td>
<td>23 publications: 15 RCTs or CCTs 8 observational</td>
</tr>
<tr>
<td>Zoledronic acid</td>
<td>Bisphosphonate</td>
<td>4 RCTs(^{54-57}) (2 HORIZON extensions)</td>
<td>1 RCT(^{58}) (HORIZON extension)</td>
<td>5 publications: 5 RCTs</td>
</tr>
<tr>
<td>Risedronate</td>
<td>Bisphosphonate</td>
<td>0</td>
<td>2 RCTs(^{59, 60})</td>
<td>2 publications: 2 RCTs</td>
</tr>
<tr>
<td>Ibandronate</td>
<td>Bisphosphonate</td>
<td>0</td>
<td>1 RCT(^61)</td>
<td>1 publication: 1 RCT</td>
</tr>
<tr>
<td>Any bisphosphonate (studies reported results as a class)</td>
<td></td>
<td>6 observational(^{62-67})</td>
<td>2 observational(^{68, 69})</td>
<td>8 publications: 8 observational</td>
</tr>
<tr>
<td>Denosumab</td>
<td>Biologic</td>
<td>1 RCT(^70)</td>
<td>1 RCT(^71)</td>
<td>2 publications: 2 RCTs</td>
</tr>
<tr>
<td>Teriparatide</td>
<td>(recombinant PTH anabolic)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Abaloparatide</td>
<td>(PTH analog anabolic)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Raloxifene</td>
<td>SERM</td>
<td>14 RCTs or CCTs(^{72-85}) (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></td>
<td>3 RCT(^86-88)</td>
<td>0</td>
<td>3 publications: 3 RCTs</td>
</tr>
<tr>
<td>Conjugated estrogens/ bazedoxifene* (Estrogen with SERM)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pooled osteoporosis drugs bisphosphonates, raloxifene</td>
<td></td>
<td>3 observational(^{89-91})</td>
<td>0</td>
<td>3 publications: 3 observational</td>
</tr>
<tr>
<td>Total number of publications</td>
<td></td>
<td>48 publications: 35 RCTs or CCTs 13 observational</td>
<td>13 publications: 7 RCTs 6 observational</td>
<td>61 publications: 42 RCTs or CCTs 19 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

*Approved by the Food and Drug Administration for osteoporosis prevention, but not for osteoporosis treatment.
Chapter 4. Efficacy of Long-Term Osteoporosis Drug Treatment

Chapter 4 reviews the evidence from eligible studies that address Key Questions 1 and 2. Key Question 1 examines the efficacy of long-term (>3 years) osteoporosis drug treatment versus inactive control (e.g., placebo, no treatment) or a different active treatment in reducing risk of incident fractures and on change in bone mineral density (BMD). Key Question 2 examines whether efficacy of long-term osteoporosis drug treatment versus control in reducing risk of incident fractures varies as a function of patient, bone, or drug characteristics.

We report the literature on the effect of continuing osteoporosis drug treatment versus discontinuing it on risk of incident fractures and on change in BMD in the drug holiday section of the report (Chapter 6, Key Question 5), regardless of the duration of drug treatment before allocation to continuation or discontinuation.

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 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 the report tables and appendixes.

Alendronate

Key Points

Key Question 1

- In postmenopausal women with either osteoporosis or osteopenia (femoral neck [FN] BMD T-score ≤ -1.6) but without a baseline vertebral fracture, between alendronate versus placebo for 4 years:
  - Alendronate was associated with no difference in risk of incident clinical fractures (low strength of evidence [SOE]), incident nonvertebral fractures (low SOE), or incident hip fractures (low SOE).
  - Alendronate was associated with a lower risk of incident radiographic vertebral fractures (high SOE).

Key Question 2

- In postmenopausal women with either osteoporosis or osteopenia and but without a baseline vertebral fracture, the effect of alendronate treatment versus placebo for 4 years varied by baseline BMD:
  - In women with FN-BMD T-score < -2.5 (osteoporosis), alendronate was associated with a lower risk of both incident clinical fractures (moderate SOE) and incident radiographic vertebral fractures (moderate SOE).
  - In women with FN-BMD T-score -2 to -2.5 (osteopenia), alendronate was not associated with a lower risk of incident clinical fractures (low SOE) and may not lower risk of incident radiographic vertebral fractures (low SOE).
  - In women with FN-BMD T-score -1.6 to -2 (osteopenia), alendronate was not associated with a lower 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 alendronate treatment versus placebo for 4 years:
  - On risk of incident nonvertebral fracture, was not modified by history of prior nonvertebral fracture.
  - On risk of incident clinical and radiographic vertebral fracture, was not modified by presence of baseline radiographic vertebral fracture.
  - On risks of incident clinical fracture, nonvertebral fracture, major osteoporotic fracture, and radiographic vertebral fracture, was not modified by baseline World Health Organization Fracture Risk Assessment Tool (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), there was no consistent modification of the effect of treatment with alendronate versus placebo for 4 years 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 risk of bias (ROB)\(^48-51\) and extracted only limited data (Appendix D). For the remaining five studies with low or medium ROB,\(^32, 35, 36, 39, 40\) 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 randomized controlled trials (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 U.S. and enrolled postmenopausal women aged 55-80 years with osteoporosis or osteopenia/low bone mass (FN-BMD ≤ 0.68g/cm\(^2\) [T score ≤ -1.6]).\(^35, 92\) 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.\(^92\) 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.\(^35\) 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.\(^32, 36, 39, 40\)
Baseline participant characteristics for the FIT-II RCT\textsuperscript{35} and the four related reports\textsuperscript{32, 36, 39, 40} are listed in Table 6 (below) and Appendix Table 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 \textless -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.\textsuperscript{32, 36, 39}

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\textsuperscript{35} FIT-II</th>
<th>Ryder 2008\textsuperscript{40} FIT-II subgroup</th>
<th>Donaldson 2012\textsuperscript{36} pooled FIT-I \ II &amp; II subset</th>
<th>Quarndt 2005\textsuperscript{39} pooled FIT-I &amp; II subset</th>
<th>Bauer 2006\textsuperscript{32} pooled FIT-I &amp; II</th>
<th>FIT-I\textsuperscript{02} (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>4432 (100)</td>
<td>2785 (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</td>
<td>4.2 yr</td>
<td>4.2 yr</td>
<td>4 yr\textsuperscript{*}</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>66</td>
<td>68</td>
<td>68</td>
<td>69</td>
<td>71</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>101</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 \textgeq 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>\textleq -1.6</td>
<td>-1.0 to -2.5</td>
<td>\textleq -1.6</td>
<td>-1.6 to -2.5</td>
<td>\textleq -1.6</td>
<td>-1.6</td>
</tr>
<tr>
<td>Comorbid conditions</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: FIT=Fracture Intervention Trial; NR=not reported for this study subsample; RVF=radiographic vertebral fracture; yr=year

\footnote{Calculated by Evidence-based Practice Center}

**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.\textsuperscript{35}

**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\% confidence intervals (CI) 0.73, 1.01]) (low SOE), incident nonvertebral fracture (11.8\% vs. 13.3\%; HR 0.88 [95\% CI 0.74, 1.04]) (low SOE), or incident hip fracture (0.9\% vs. 1.1\%, HR 0.79 [95\% CI 0.43, 1.44]) (low SOE).\textsuperscript{35}

**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 lumbar spine (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 treatment versus placebo or inactive control on risk of incident fracture varies as a function of patient, bone, 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 unmasked, the effect of long-term alendronate versus placebo on risks of any incident clinical fracture and incident radiographic vertebral fracture varied by baseline BMD.35 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 radiographic vertebral fractures. There appeared to be no difference between 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-scores of <-2.5 (HR 0.44 [95% CI 0.18, 0.97]), but not in those with baseline FN-BMD T-scores >-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).40

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).39
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.\textsuperscript{32} Participants were categorized at baseline as either osteoporotic (FN-BMD $\leq -2.5$ 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, major osteoporotic fracture (MOF), and radiographic vertebral fracture were not significantly different between individuals with higher versus lower FRAX scores (Appendix Table D3).\textsuperscript{36} 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 $\geq$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).\textsuperscript{32}

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 procollagen I intact N-terminal (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 bone-specific alkaline phosphatase (BSAP) or C-terminal telopeptide (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).
Zoledronic Acid (Zoledronate)

Key Points

Key Question 1
- In postmenopausal women with osteopenia in at least one hip (BMD T-score between -1.0 to -2.5) (a minority of women met at least one criterion for osteoporosis), between zoledronate for 6 years versus placebo:
  - Zoledronate was associated with a lower risk of incident clinical fractures (moderate SOE), incident nonvertebral fractures (high SOE), and incident clinical vertebral fractures (moderate SOE).
  - There appeared to be no difference in risk of incident hip fractures (low SOE).

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

Eligible Studies

One eligible publication of one RCT compared long-term zoledronate treatment versus placebo and reported on risk of incident fracture (Table 7).57 Details of this study with low ROB were extracted in evidence tables, summary SOE assessments (Appendix D), and ROB assessments (Appendix C).

This single-site RCT randomized 2000 postmenopausal women aged >65 years to zoledronate 5 mg intravenously every 18 months for 6 years versus injections with saline placebo. Women were eligible if they had osteopenia at either hip (T-score between -1.0 and -2.5), even if they had osteoporosis (T < -2.5) at the other hip or spine (8% of the participants) or had a baseline radiographic vertebral fracture (13% of participants). Women were excluded if their LS-BMD T-score was worse than -3.0.

Table 7. Baseline characteristics of the zoledronate trial versus placebo

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean or %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number randomized</td>
<td>2000</td>
</tr>
<tr>
<td>Age, mean years</td>
<td>71</td>
</tr>
<tr>
<td>Gender, woman</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white (European)</td>
<td>95%</td>
</tr>
<tr>
<td>Race, East Asian</td>
<td>2%</td>
</tr>
<tr>
<td>Women with a T-score &lt; -2.5 (considered osteoporotic)</td>
<td>8%</td>
</tr>
<tr>
<td>Mean FN-BMD T-score</td>
<td>-1.64</td>
</tr>
<tr>
<td>Mean LS-BMD T-score</td>
<td>-0.89</td>
</tr>
<tr>
<td>Mean hip BMD T-score*</td>
<td>-1.26</td>
</tr>
</tbody>
</table>
Outcomes

Incident Clinical Fractures

The primary trial endpoint was incident fragility fracture, which it defined as a clinical or radiographic vertebral fracture or any nonvertebral fracture excepting fractures of the toes, metatarsals, metacarpals, fingers, skull, facial bones, or mandible. Incident fractures of primary interest for this review were considered secondary endpoints.

Zoledronate for 6 years versus placebo significantly reduced the risk of incident clinical fracture (16.3% vs. 21.4%; HR 0.73 [95% CI 0.60, 0.90]) (moderate SOE), incident nonvertebral fracture (10.1% vs. 14.8%; HR 0.66 [95% CI 0.51, 0.85]) (high SOE), and incident clinical vertebral fracture (1.4% vs. 3.4%; HR 0.41 [95% CI 0.22, 0.75]) (moderate SOE). Zoledronate did not significantly reduce risk of incident hip fracture compared with placebo (0.8% vs. 1.2%, HR 0.66 [95% CI 0.27, 1.16]) (low SOE).

Incident Radiographic Vertebral Fractures

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

Change in BMD

Change in BMD was a secondary endpoint and results were displayed graphically. Compared with placebo, there were significant improvements in BMD with zoledronate over the 6-year treatment period. Mean change from baseline in LS-BMD increased approximately 7 percent with zoledronate versus a 1 percent decrease for placebo. Mean change in TH-BMD improved approximately 4 percent in the zoledronate group compared with a decrease of nearly 4 percent in the placebo group.

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

No eligible studies evaluated whether the efficacy of long-term zoledronate versus placebo or other inactive control on risk of incident clinical fracture varied as a function of patient, bone, or drug characteristics.
Denosumab

Key Points

Key Question 1
- In postmenopausal women with BMD T score < -1.8 (includes both osteopenia and osteoporosis), evidence was insufficient to draw conclusions about differences between 4 years of denosumab therapy versus placebo on risk of incident clinical fractures.
- There was no evidence from eligible trials about the efficacy of long-term denosumab versus placebo on risk of 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 other inactive 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 8). Details of this study with medium ROB were extracted in evidence tables, summary SOE assessments (Appendix D), and ROB assessments (Appendix C).

This multisite trial 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) versus one of seven intravenous denosumab dosing regimens (n=319) for 2 years. These dosing regimens included three that were lower and two that were higher than the current U.S. Food and Drug Administration (FDA) recommended denosumab dosing. After 2 years, individuals were nonrandomly assigned to treatment groups for the following 2 years. Those initially assigned placebo remained assigned to placebo. Five of the groups assigned denosumab, including three of the low dose groups, were assigned to denosumab 60 mg every 6 months for 2 years (n=231), one of the original high dose denosumab groups was switched to placebo for 2 years (n=47), and the last denosumab group was switched to placebo for 1 year before starting denosumab 60 mg every 6 months for 1 year (n=41).

Table 8. 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>
</tbody>
</table>
Incident Clinical Fractures

In this trial, clinical and osteoporotic fractures were considered neither primary or secondary outcomes, but rather were identified during routine adverse event reporting. It is unclear whether incident fracture reports were confirmed by radiographic reports. 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, incident clinical fractures occurred in 10.5% of 314 individuals originally assigned to denosumab (the discrepancy of this denominator from the 319 the study reported were initially assigned denosumab was not explained) and 10.9% of those originally assigned placebo (risk ratio [RR] 0.97 [95% CI 0.40, 2.35]). Incident osteoporotic fracture (any clinical fracture excluding those of the phalanges, face or those caused by severe trauma) occurred in 7.0% of the participants originally assigned denosumab vs. 8.7% in the original placebo group (RR 0.81 [95% CI 0.29, 2.23]). Because of inclusion in the denosumab group of some participants with low initial doses and others who were assigned only to short-term treatment, evidence is insufficient to draw conclusions about the efficacy of long-term use of FDA approved dosing of denosumab versus placebo for prevention of incident fractures.

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. Unlike for fracture outcomes, 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).

<table>
<thead>
<tr>
<th>Mean hip BMD T-score*</th>
<th>-1.42</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of clinical fracture</td>
<td>NR</td>
</tr>
<tr>
<td>Comorbid conditions</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: BMD=bone mineral density; LS-BMD=lumbar spine bone mineral density; NR=not reported

*Study did not specify whether reported baseline hip BMD was from total hip or femoral neck.
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 placebo or other inactive control on risk of incident fractures 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 significant difference between treatments in risk of incident nonvertebral fracture (high SOE) or incident hip fracture (moderate SOE).
  - Raloxifene was associated with 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 significant difference between treatments 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
- Associations between raloxifene and placebo with risk of incident fractures did not appear to 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 9). 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” [undefined] 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 initial 3-year 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 controlled clinical trial 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 Table D2. The four reports from MORE had low ROB,72, 76, 81, 85 and the single report from CORE had medium ROB for fracture outcomes but high ROB for BMD outcomes (available in <10% of CORE population).84

Table 9. Baseline participant characteristics from the MORE trial and CORE extension

<table>
<thead>
<tr>
<th>Characteristic, mean, or %</th>
<th>MORE</th>
<th>CORE*</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>Chronic kidney disease</td>
<td>NR</td>
<td>NR</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

*Characteristics available only from MORE baseline

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]) (moderate SOE). However, raloxifene 60 mg/day was associated with a reduced 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 <0.10.84

Hormone Therapy

Key Points

Key Question 1

- In postmenopausal women with unknown osteoporosis or osteopenia status, and with prior hysterectomy, 7 years of unopposed, high dose oral estrogen versus placebo lowered risk of both incident clinical fractures (high SOE) and incident hip fractures (moderate SOE).
- In postmenopausal women with unknown osteoporosis or osteopenia status, and with an intact uterus, 5.6 years of high dose oral estrogen plus progestin versus placebo lowered risk of both incident clinical fractures (high SOE) and incident hip fractures (moderate SOE).
- In postmenopausal women with past clinical fracture, but unknown BMD or vertebral fracture status, and with prior hysterectomy, 7 years of unopposed, high dose oral estrogen versus placebo lowered risk of both incident clinical fractures (moderate SOE) and incident hip fractures (low SOE).
- In postmenopausal women with past clinical fracture, but unknown BMD or vertebral fracture status, and with an intact uterus, 5.6 years of high dose oral estrogen plus progestin lowered risk of incident clinical fractures (low SOE), but was not associated with lower risk of incident hip fractures (low SOE).

Key Question 2

- Evidence was insufficient to draw conclusions about whether differences in risk of incident fractures between long-term, high dose unopposed oral estrogen or estrogen plus progestin versus placebo or other inactive treatment vary as a function of patient, bone or drug characteristics.

Eligible Studies

We identified 3 eligible RCTs that compared hormone therapy versus placebo86, 87 or nonplacebo control88 and reported long-term results for incident fracture (Table 10). 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 (oral conjugated estrogen 0.625 mg/day plus medroxyprogesterone 2.5 mg/day) versus placebo for a mean treatment duration of 5.6 years.86 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 6 percent of participants (1024 women enrolled at 3 of 40 clinical sites). The second WHI trial randomized 10,739 postmenopausal women with a history of hysterectomy to unopposed oral estrogen
(conjugated 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 9 percent of participants (938 women enrolled 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 ≤-2.0. Of the 72 women enrolled in this study, 36 were randomized to etidronate alone or to a combination of etidronate plus conjugated estrogen/progestin. Because etidronate is not FDA approved, these treatment groups were not analyzed. The remaining 36 women were randomized to an oral conjugated 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 10. Baseline characteristics of hormone therapy versus placebo/control RCTs

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment intervention</td>
<td>Oral Estrogen + Progestin vs. PBO</td>
<td>Oral Estrogen vs. PBO</td>
<td>Oral 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 6% [n=1024])</td>
<td>10,739 (BMD measured in 9% [n=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>&lt;10 yr = 36%</td>
<td>Mean 19 yr</td>
<td>Median 15 yr</td>
</tr>
<tr>
<td></td>
<td>10-19 yr = 40%</td>
<td>&lt;10 yr = 18%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>≥20 yr = 24%</td>
<td>10-19 yr = 32%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;20 yr = 50%</td>
<td></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 baseline RVF per participant</td>
<td>NR</td>
<td>NR</td>
<td>2.3</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>MI = 2%</td>
<td>MI = 3%</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Stroke = 0.8%</td>
<td>Stroke = 1.6%</td>
<td></td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Diabetes</td>
<td>4%*</td>
<td>8%*</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: BMD=bone mineral density; FN=femoral neck; LS=lumbar spine; MI=myocardial infarction; 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 etidronate plus estrogen/progestin.
Outcomes

Incident Clinical Fractures

In the first WHI trial, in women with unknown osteoporosis or osteopenia status and an intact uterus, those randomized to oral estrogen/progestin for 5.6 years versus placebo had a lower risk of both incident clinical fracture (8.6% vs. 11.1%; HR=0.76 [95% CI 0.69, 0.83]) (high SOE) and incident hip fracture (0.61% vs. 0.90%; HR=0.67 [95% CI 0.47, 0.96]) (moderate SOE). Among the subset of women with a history of prior clinical fracture, those randomized to estrogen/progestin versus placebo 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). In the subset of women who completed baseline BMD testing and were found to have osteoporosis by baseline DXA BMD T-score <-2.5 at any hip or vertebral site, evidence was insufficient to draw conclusions about a difference in risk of any incident clinical fracture between the estrogen/progestin and placebo groups (HR=0.53 [95% CI 0.25, 1.10]).

In the second WHI trial, in women with unknown osteoporosis or osteopenia status and a past hysterectomy, those randomized to oral estrogen for 7.1 years versus placebo had a lower risk of both incident clinical fracture (10% vs. 14%; HR=0.71 [95% CI 0.64, 0.80]) (high SOE) and incident hip fracture (0.87% vs. 1.3%; HR=0.65 [95% CI 0.45, 0.94]) (moderate SOE). Among the subset of women with a history of prior clinical fracture (36% of participants), 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). In the subset of women who completed baseline BMD testing and were found to have osteoporosis (n=53) or osteopenia (n=363), evidence was insufficient to draw conclusions about differences in risk of any incident clinical fracture between treatment groups (HR=0.83 [95% CI 0.17, 3.91] and HR=0.83 [95% CI 0.49, 1.40], respectively).

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, available only in the small minority of women with BMD measurements (<10%). 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 TH-BMD and 4-5% greater increase in LS-BMD between baseline and follow-up at 3 years.
In the small, non-WHI trial, the primary study endpoints were mean percentage change from baseline in DXA LS-BMD and TH-BMD. Among the 36 participants randomized to estrogen/progestin versus placebo, 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]). Results also favored the estrogen/progestin group for change in TH-BMD (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**

In both WHI trials, the effect of treatment assignment between hormone therapy and placebo on incident fracture risk in the women with unknown osteoporosis or osteopenia status was evaluated as a function of different patient characteristics in more than 40 tests for interaction.

In both trials, the effects of treatment assignment on risk of incident clinical fracture and incident hip fracture were evaluated as a function of age, years since menopause, body mass index, smoking status, fall history, total calcium intake, parental fracture history, history of prior hormone therapy, race/ethnicity, fracture history, and summary fracture risk score. In the trial that compared estrogen/progestin versus placebo, results also were examined as a function of osteoporotic BMD in the small subset of participants with available baseline BMD measures. In the trial that compared unopposed estrogen versus placebo, results also were examined as a function of history of oophorectomy.

Nearly all tests of interaction were not statistically significant (p>0.05), with the following exceptions: the reduction in risk of incident hip fracture with estrogen/progestin versus placebo appeared greater in women with total calcium intakes of more than 1200 mg/day compared to lower intakes (p=0.02); the reduction in risk of incident hip fracture with unopposed estrogen versus placebo appeared greater in women at least 20 years since menopause compared to women with fewer years since menopause (p=0.05); the reduction in risk of incident clinical fracture with unopposed estrogen versus placebo appeared greater in women aged 60-69 and 70-79 compared with women 50-59 years (p=0.03), and greater in women with high or moderate versus low baseline summary fracture risk scores (p=0.04). In addition, reduction in risk of incident clinical fracture with unopposed estrogen versus placebo, though not statistically significantly different by years since menopause (p=0.06), showed a trend similar to that for hip fractures, with risk appearing numerically lower in women 60-69 and 70-79 compared to that in women aged 50-59 years.
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.\textsuperscript{59, 60} These studies were rated high ROB and only limited data were extracted (Appendix Table 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 Treatment

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 treatment versus discontinuing it on risk of harms in the drug holiday section of the report (Chapter 7, Key Question 7), regardless of the duration of drug treatment before allocation to continuation or discontinuation.

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 randomized controlled trials (RCT):
  - In postmenopausal women with femoral neck (FN) bone mineral density (BMD) T-score ≤-1.6 and without baseline vertebral fracture, between alendronate and placebo for 4 years:
    - There was no difference in risk of upper gastrointestinal (GI) events, mortality (low strength of evidence [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:
    - Evidence was insufficient to draw conclusions about differences in risk of atypical femoral fracture (AFF) with confirmed radiologic features or subtrochanteric or femoral shaft fractures without confirmed radiologic AFF features.
- In observational studies:
  - Between alendronate and no osteoporosis medication use for up to 11 years:
    - 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 no treatment in a second cohort study.
    - Risk of osteonecrosis of the jaw (ONJ) appeared higher with alendronate versus no osteoporosis drug prescription in one cohort study (low SOE).
    - 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 \( \leq -1.6 \), between alendronate and placebo for 3.2 to 4 years:
  - Risk of upper gastrointestinal 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 ROB\(^47, 50-52\) and extracted only limited data (Appendix Table D7). For the remaining ten studies with low or medium ROB, there were four eligible publications from one unique RCT with low\(^35, 43\) and medium\(^31, 38\) ROB, respectively, and six observational studies with medium ROB\(^44-46, 89-91\) We extracted additional information from these studies in evidence tables, SOE summary tables (Appendix D) and ROB summary tables (Appendix C). Summary data of rare harms from controlled observational studies can be found in Table 11.

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.\(^35\) In a 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.\(^31\) The third FIT report pooled participants from FIT-I and FIT-II and estimated risk of incident subtrochanteric or femoral shaft fracture.\(^43\) 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.\(^38\) None of the reports provided information about the risk of ONJ or atrial fibrillation.

Four observational studies used Danish national data to examine risk of subtrochanteric and femoral shaft fracture,\(^44, 89\) inflammatory jaw events possibly attributable to ONJ,\(^91\) or atrial fibrillation\(^90\) with long-term alendronate. In the earliest of these studies, among men and women...
≥51 years of age who experienced a hospital treated nonhip fracture, 178 individuals adherent with alendronate for >6 years (i.e., medication possession ratio >80%) were matched 2:1 with age, sex, and same index fracture site controls who were not exposed to bisphosphonate treatment. Subtrochanteric and femoral shaft fractures were identified by International Classification of Diseases (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 of these studies 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. 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. For the third report, the primary outcome was atrial fibrillation or flutter defined as a recorded incident leading to hospitalization or an outpatient contact. 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 that in 1869 osteoporotic women using raloxifene. 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 sequestrum 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.

A second 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 clinical vertebral or hip fracture. 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%), more often having had a hip fracture as their pre-treatment fracture (29% vs. 15%), and more often 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 Table D1.

Outcomes

**Serious Adverse Events**

In the FIT-II RCT, although serious adverse events as defined by the U.S. Food and Drug Administration (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%; hazard ratio [HR] 0.92 [95% confidence interval [CI] 0.59, 1.45]) (low SOE) or adverse events leading to hospitalization (29% vs. 27%; relative hazard [RH] 1.09 [95% CI 0.98, 1.22]) (low SOE).35

**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).35 Risk of esophagitis was 0.9% in the alendronate group versus 0.5% in the placebo group (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]).31 There also was no between group difference in risk of 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 Table 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.43, 44, 89 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)).43 One retrospective cohort study reported five cases of subtrochanteric or femoral shaft fracture in the combined groups of patients highly adherent with alendronate >6 years (n=178) and a no treatment group (n=356), and provided insufficient strength evidence to draw conclusions about a difference in risk between groups (HR 1.37 [95% CI 0.22, 8.62]).44 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).89

**Osteonecrosis of the Jaw–Defined by Diagnostic Codes Plus Clinical Confirmation**

One retrospective cohort study reported higher risk with alendronate versus raloxifene for ONJ defined by diagnostic codes for an inflammatory jaw condition, plus radiologic and pathologic confirmation (HR 7.42; 95% CI 1.02, 54.09) (low SOE).45 A second study, that defined ONJ based on recurrent and persistent plus receipt of appropriate antibiotics reported
that risk for alendronate versus raloxifene or calcitonin was HR 0.86 (95% CI 0.44, 1.69) (insufficient SOE).46. Because the higher adjusted incidence rates in the alendronate group (0.15%) compared with the raloxifene-calcitonin group (0.08%) suggested a risk estimated higher than 1.0, we manually recalculated the estimate of effect and found RR 1.20 (95% CI 0.59, 2.56). Authors were contacted for clarification, but did not reply.

**Osteonecrosis of the Jaw–Defined by Diagnostic Codes Only**

One retrospective cohort study reported an increased risk of any inflammatory jaw event defined only by diagnostic codes in participants exposed to approximately 3.8 years of alendronate exposure versus no osteoporosis drug prescription (0.03% vs. 0.01%; HR 3.15 [95% CI 1.44, 6.87]) (low SOE).91

**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]).90
Table 11. Estimated risk of ST/FS fracture,* ONJ, or atrial fibrillation with alendronate use from controlled observational studies

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Author, Year</th>
<th>Country</th>
<th>Risk of bias</th>
<th>Study Design (n=# Cases; N=# Evaluated)</th>
<th>Population Age (mean) Gender (%) Comorbid Conditions (%)</th>
<th>ST/FS fracture, ONJ, or Atrial Fibrillation Case Definitions</th>
<th>Alendronate Treatment Duration</th>
<th>Treatment Control Group</th>
<th>Results for Alendronate vs. Control (95% CI) Model Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST/FS fracture</td>
<td>Abrahamsen 200944</td>
<td>Denmark</td>
<td>Medium</td>
<td>Retrospective cohort n=5 N=534</td>
<td>Age 73 Female 90% Comorbid conditions NR</td>
<td>ST/FS cases: ST/FS 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></td>
<td>Vestergaard 201189</td>
<td>Denmark</td>
<td>Medium</td>
<td>Retrospective cohort n=309 N=220,360</td>
<td>Age 71 Female 85% CHF 2% DM 2% HTN 34%</td>
<td>ST/FS cases: ST/FS 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: 0.17%alendronate vs. 0.06% control; HR 2.41 (1.78, 3.27); ARR 0.11 (0.08, 0.15) Femoral shaft: 0.12% alendronate vs. 0.03% control; HR 2.90 (1.97, 4.26); ARR 0.09 (0.06, 0.12) ST/FS: 0.29% alendronate vs. 0.09% control; ARR 0.20 (0.15, 0.25) History of past fracture, systemic hormone use, systemic corticosteroid use, alcoholism</td>
</tr>
<tr>
<td>ONJ</td>
<td>Vestergaard 201221</td>
<td>Denmark</td>
<td>Medium</td>
<td>Retrospective cohort n=28 N=220,360</td>
<td>Age 71 Female 85% CHF 2% DM 2% HTN 34%</td>
<td>ONJ case: Diagnosis codes for inflammatory jaw event including osteomyelitis, osteitis, osteoradionecrosis, periostitis and sequestrum. Radiology and pathology records not reviewed.</td>
<td>Range 0-11 years Mean duration ~3.8 years</td>
<td>No prescription for osteoporosis drug</td>
<td>0.03% alendronate vs. 0.01% control; HR 3.15 (1.44, 6.87) Diabetes, Sjogren’s syndrome, chemotherapy, irradiation, alcoholism, use of systemic corticosteroids, prior jaw events</td>
</tr>
<tr>
<td></td>
<td>Chiu 201445</td>
<td>Taiwan</td>
<td>Medium</td>
<td>Retrospective cohort n=40 N=8,354</td>
<td>Age 72 Female 88% CKD 2% DM 16% HTN 44%</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 are unclear.</td>
</tr>
<tr>
<td>Outcome</td>
<td>Author, Year</td>
<td>Study Design (n=# Cases; N=# Evaluated)</td>
<td>Population Age (mean) Gender (%) Comorbid Conditions (%)</td>
<td>ST/FS fracture, ONJ, or Atrial Fibrillation Case Definitions</td>
<td>Alendronate Treatment Duration</td>
<td>Treatment Control Group</td>
<td>Results for Alendronate vs. Control (95% CI) Model Covariates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------------------------------------------</td>
<td>--------------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin 2014&lt;sup&gt;46&lt;/sup&gt; Taiwan Medium</td>
<td>Retrospective cohort Unmatched cohort: n=46 N=43,645 Propensity matched cohort: n=37 N=32,006</td>
<td>Age 74 Female 84% CKD 3% DM 24% ONJ case: ICD9 diagnostic codes from at least 3 consecutive visits over at least 8 weeks, and receipt of appropriate broad-spectrum antibiotics.</td>
<td>Up to 6 years (mean and median not reported)</td>
<td>Non-bisphosphonate osteoporosis medication (e.g. raloxifene or calcitonin)</td>
<td>Non-bisphosphonate osteoporosis medication (e.g. raloxifene or calcitonin)</td>
<td>Propensity matched cohort: 0.15% alendronate vs. 0.08% control; HR 0.86 (0.44, 1.69)&lt;sup&gt;†&lt;/sup&gt; Age, gender, year, fracture type, fracture history, fall history, comorbidities (DM, hyperlipidemia, pancreatitis, gingival and periodontal diseases, other diseases and conditions of teeth and supporting structures, dentoalveolar surgery, rheumatoid arthritis, systemic lupus erythematosus, renal disease, HTN, Alzheimer's disease), and comedinations (antiepileptics, β-blockers, benzodiazepines, glucocorticoids, hormone therapy, COX-2 agents, SSRI, thyroid drugs, and sleep/hypnotic agents).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation Vestergaard 2010&lt;sup&gt;50&lt;/sup&gt; Denmark Medium</td>
<td>Retrospective cohort n=2,364 N=220,360</td>
<td>Age 71 Female 85% CHF 2% DM 2% HTN 34% Atrial fibrillation case: recorded incident leading to hospitalization or an outpatient contact</td>
<td>Range 0-11 years Mean duration ~3.8 years</td>
<td>No prescription for osteoporosis drug</td>
<td>No prescription for osteoporosis drug</td>
<td>1.3% alendronate vs. 1.0% control; 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</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ARR=absolute risk reduction; CHF=chronic heart failure; CI=confidence interval; CKD=chronic kidney disease; COPD=chronic obstructive pulmonary disease; COX=cyclooxygenase; DM=diabetes mellitus; HR=adjusted hazard ratio; HTN=hypertension; ICD=International Classification of Diseases; N=number; ONJ=osteonecrosis of the jaw; SSRI=selective serotonin reuptake inhibitors; ST/FS=Subtrochanteric/femoral shaft

*No studies reported results for atypical femoral fractures meeting American Society for Bone and Mineral Research radiographic criteria.
†Because the higher adjusted incidence rates in the alendronate group compared with the raloxifene-calcitonin group (0.15% vs. 0.08%) suggested a possibly increased risk, we manually recalculated the estimate of effect and found risk ratio 1.20 (95% CI 0.59, 2.56). Authors were contacted for clarification, but did not reply.
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 control (placebo, no treatment, or active 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:31, 38

Upper Gastrointestinal Tract Adverse Events

A post hoc pooled analysis of FIT-I and FIT-II data31 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 Table 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.38
Zoledronic Acid (Zoledronate)

Key Points

Key Question 3
- All data on long-term zoledronate treatment harms were from a single RCT. There were no eligible observational studies that compared long-term zoledronate versus no treatment or another control, let alone any that reported on risk of harms.
- In mostly osteopenic postmenopausal women (one hip BMD T-score -1 to -2.5, but including a minority of women with osteoporosis by BMD or baseline vertebral fracture), between zoledronate and placebo for 6 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 (no cases of either AFF or ONJ occurred during the trial).

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

Eligible Studies

One eligible RCT reported on risk of harms of long-term zoledronate treatment versus placebo.57 No eligible observational studies addressed risk of harms of long-term zoledronate treatment.

Described in detail previously, the single eligible trial enrolled 2000 postmenopausal women, aged >65 years, most of whom had osteopenia (all had a T-score -1.0 to -2.5 in at least one hip). However, some participants met criteria for osteoporosis (8% had a BMD T-score < -2.5 at their other hip or the spine, and 13% had a baseline radiographic vertebral fracture). Participants were randomized to intravenous zoledronate 5 mg every 18 months (n=1000) versus placebo (n=1000). Treatment duration was 6 years. The trial was rated with low ROB. Study details were extracted in evidence tables, a ROB assessment table (Appendix C), and a SOE assessments summary table (Appendix D).

Reports of eligible studies that compared zoledronate continuation versus discontinuation (placebo drug holiday) and reported on risk of harms are reviewed in Chapter 7 of this report.

Outcomes

Serious Adverse Events
Risk of any serious adverse event was 40.0 percent with zoledronate treatment versus 44.3 percent with placebo, a result that was not statistically significant (odds ratio [OR] 0.84 [95% CI 0.70, 1.00]; p=0.052) (low SOE).
Cardiovascular Events
Risk of a cardiovascular event, including sudden death, myocardial infarction, coronary-artery revascularization, or stroke, was not significantly different between women allocated to zoledronate versus placebo (5.3% vs. 6.9%; OR 0.76 [95% CI 0.52, 1.09]). Individually, risk of sudden death, myocardial infarction, or stroke was not statistically significantly different between groups.

Mortality
Risk of all-cause mortality was not statistically significantly different between the zoledronate and placebo groups (2.7% vs. 4.1%; OR 0.65 [95% CI 0.40, 1.05]). Sudden death was reported for 3 women in the zoledronate and one in the placebo group.

Atypical Femoral Fracture
The trial reported that no cases of AFF occurred in either treatment group.

Osteonecrosis of the Jaw
The trial reported that no cases of ONJ occurred in either treatment group.

Variation in Long-Term Treatment Harms as a Function of Patient, Bone, or Osteoporosis Drug Characteristics
We identified no eligible study that compared long-term zoledronate 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.
Any Bisphosphonate

Key Points

Key Question 3

- All data on harms associated with unspecified bisphosphonate treatment were from observational studies. There were no randomized long-term trials that compared any bisphosphonate versus placebo, let alone any that reported on risk of harms.
- There was higher risk of AFF with confirmed radiologic features with long-term use of any bisphosphonate versus no bisphosphonate use (low SOE) or versus 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.
- There was no evidence from eligible studies about risk of ONJ between 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

Eight eligible publications of eight unique observational studies 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 two studies as high ROB68, 69 and extracted limited data (Appendix Table D7). The remaining six studies had medium ROB.62-67 Appendix D provides detailed evidence tables and summary SOE assessments; ROB assessments are provided in Appendix C.

Three studies compared the effect of long-term use of any bisphosphonate versus no osteoporosis drug use on risk of harms. The first of these studies, conducted in Sweden, performed both retrospective cohort and case-control analyses.66 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. All had their radiographs reviewed for American Society for Bone and Mineral Research (ASBMR) criteria.19 Those whose fractures met ASBMR criteria were categorized as having AFF and those whose fractures did not meet ASBMR criteria 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 subtrochanteric or femoral shaft fractures 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. 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. The third of these studies was a multi-center case-control study conducted in South Korea. Participants were women aged >50 years with a diagnosis code for either subtrochanteric or femoral shaft fracture during a four-year period, and no associated femoral neck, intertrochanteric, distal femoral, high-energy, prosthetic or pathologic fracture. All participants had radiographs reviewed; those whose met ASBMR criteria were categorized as having AFF and those whose subtrochanteric or femoral shaft fractures were judged not to have AFF radiologic features were categorized as non-AFF controls. Risk of AFF was estimated for those with a history of bisphosphonate use (per medical record chart review; mean duration 5.2 years) compared with those without history of bisphosphonate use. Osteoporosis or osteopenia rates were higher in the AFF group (85%) than the control group (18%).

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 ≥5 years versus that with past use for <100 days. 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 postmenopausal women in Korea. It compared AFF risk between women with current use of any bisphosphonate for ≥5 years versus that in women with past use for ≥1 year that was stopped between 6 months and 5 years prior to the study start date. Participants were identified through electronic hospital records over an 8-year study period. AFF cases were defined using diagnostic codes and radiologic confirmation of AFF features. Cases were excluded for excessive trauma or cancer. Mean participant age was 68 years, mean pre-fracture (or most recent) dual-energy x-ray absorptiometry (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. 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**

All data on harms associated with unspecified bisphosphonate treatment were from observational studies.

**Atypical Femoral Fractures**

Three controlled observational studies each provided low strength evidence that long-term bisphosphonate use was associated with a significantly increased risk of radiologically confirmed AFF.

Two studies compared risk of long-term bisphosphonate use versus no bisphosphonate use. In one of these studies, long-term (>3 years) bisphosphonate use was associated with increased risk of AFF in both retrospective cohort (RR 126 [95% CI 55, 288]) and case-control analyses (range of OR 40 [95% CI 17, 91] to 116 [95% CI 58, 234]). In the second of these studies, bisphosphonate use for 5.2 years was associated with increased risk of AFF (OR 25.65 [95% CI 10.74, 61.28]). In this latter study, 85 percent of the bisphosphonate group had osteoporosis or osteopenia (defined as FN-BMD T-score <-2.0) compared to 18 percent in the control group. This difference in population may be evidence of confounding by indication.

One study reported on risk of radiologically confirmed AFF for current versus past bisphosphonate use, and risk was significantly increased in all reported models (HR 3.36 to 5.17) (Table 12).

**Subtrochanteric/Femoral Shaft 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 between three studies reporting these results (Table 12). Compared to no use of bisphosphonates, risk of subtrochanteric or femoral shaft 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 subtrochanteric or femoral shaft fractures was not significantly increased with 3-5 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]). Compared to raloxifene or calcitonin use, risk of subtrochanteric or femoral shaft fractures was not increased with either 3-5 years or >5 years of bisphosphonate use.

Risks for AFF and subtrochanteric or femoral shaft fractures varied substantially between studies. This variation may have been attributable to several differences in study design. First, studies differed in case definition, in particular whether radiological confirmation for AFF features was required. Studies that didn’t require confirmation of these radiologic features necessarily included many typical osteoporotic femur fractures as cases. This would have had the effect of overestimating the absolute risk of AFF, and could have underestimated the relative risk of AFF if long-term bisphosphonate treatment lowered risk of typical osteoporotic femur fractures in these study populations. Studies also used different non-AFF control groups, with some using no hip fracture and others using subtrochanteric or femoral shaft fractures with
radiographs that did not have AFF features. If long-term bisphosphonates lowered risk of these
typical osteoporotic femoral fractures, use of this latter control group could have increased the
estimated risk of AFF independently of any direct effect on AFF risk. Studies also differed in
terms of the duration of bisphosphonate exposure, the nonbisphosphonate control group (no
bisphosphonate use, past bisphosphonate use, other osteoporosis drug use), case control versus
retrospective cohort design, and which covariables were included in adjusted statistical models.

The impact of these factors in isolation was difficult to discern because they generally were
not compared within single studies and studies differed in multiple ways. Nevertheless, the two
studies of any long-term bisphosphonate use versus no use that defined AFF using ASBMR task
force criteria, excluded pathological fractures, and compared risk of being a radiologically
confirmed AFF case to that of being a subtrochanteric or femoral shaft fracture without
radiologic AFF features had the highest estimated risks. These ranged from OR 40 to 116 and
RR 126 in one study,66 to OR 25.65 in the second study.67 The single study that compared risk of
radiologically confirmed AFF to no history of AFF, while still excluding pathologic fractures,
reported smaller, but still significantly increased risks, with HR ranging from 3.36 to 5.17.64 For
the two studies that didn’t evaluate radiographs for AFF features, but excluded cases for trauma
or cancer, risk estimates ranged from OR 1.59 to 2.7465 and OR 9.46.62 By comparison, the
single study that neither evaluated radiographs for AFF features, nor excluded cases for trauma
or cancer, found no increase in risk.63 Further, the three 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 9.46,62 to RR 40 and 126 (range of estimates in this
study due to different models and follow-up times),66 and RR 25.65,67 By comparison, the
studies that compared long-term bisphosphonate use to past use64, 65 reported intermediately
increased risk of AFF, and 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.63

**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>
<thead>
<tr>
<th>Outcome</th>
<th>Author, Year Country Risk of Bias</th>
<th>Study Design</th>
<th>Population Age (mean) Gender (%) Comorbid Conditions (%)</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>AFF*</td>
<td>Schilcher, 2015t Sweden Medium</td>
<td>Retrospective cohort (n=98,200 BP users, n=2.8 million non-BP users)</td>
<td>Age NR Female 87% Comorbid conditions NR</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, 288) (women only) Age</td>
</tr>
<tr>
<td>Case-control</td>
<td>Age 81 Female 83% CKD 39% CVD 58% DM 14%</td>
<td>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>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>See above</td>
<td>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>
<td></td>
<td></td>
</tr>
<tr>
<td>Case-control</td>
<td>See above</td>
<td>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>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>Author, Year</td>
<td>Country</td>
<td>Risk of Bias</td>
<td>Study Design</td>
<td>Population Age (mean) Gender (%) Comorbid Conditions (%)</td>
<td>Case and Control Definitions</td>
<td>BP Treatment Duration</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>---------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td></td>
<td>Koh, 2017&lt;sup&gt;4&lt;/sup&gt; Korea Medium</td>
<td>Case-control</td>
<td>Age 68 Female 100% CKD NR CVD NR DM 13%</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 ≥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></td>
<td>Lim, 2018&lt;sup&gt;9&lt;/sup&gt; South Korea Medium</td>
<td>Case-control</td>
<td>Age 72 Female 100% DM 20% CKD NR CVD 10%</td>
<td>AFF case (n=196): Diagnosis codes for subtrochanteric and femoral shaft fracture, exclusions for cancer/pathologic fracture, met ASBMR radiographic criteria for AFF. Non-AFF control (n=96): 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>Mean 5.2 years (range 1-17 years)</td>
<td>Non-use of BP</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>Outcome</td>
<td>Author, Year Country Risk of Bias</td>
<td>Study Design</td>
<td>Population Age (mean) Gender (%) Comorbid Conditions (%)</td>
<td>Case and Control Definitions</td>
<td>BP Treatment Duration</td>
<td>Treatment Control Group</td>
<td>RR or OR (95% CI) Model Covariates</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>ST/FS fracture†</td>
<td>Erviti, 2013 Spain Medium</td>
<td>Case-control</td>
<td>Age 82 Female 100% CKD 6% CVD NR DM 21%</td>
<td><strong>ST/FS cases</strong> (n=44): Diagnosis codes for subtrochanteric and femoral shaft fracture with exclusions for excessive trauma and cancer, but no review of radiographs. <strong>Non-ST/FS fracture control</strong> (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) 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></td>
<td>Park-Wyllie, 2011 Canada Medium</td>
<td>Case-control</td>
<td>Age 83 Female 100% CHF 6% CKD NR DM 7%</td>
<td><strong>ST/FS cases</strong> (n=204): Hospitalized with diagnosis codes for first subtrochanteric or femoral shaft fracture, with exclusions including for excessive trauma, cancer/pathologic fracture, and current use of non-BP osteoporosis medications, but no review of radiographs. <strong>Non-ST/FS fracture control</strong> (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) Socioeconomic status, comedications, drug count, comorbidities, recent medical visits, prior fall, prior osteoporotic fracture, BMD test past 5 years</td>
</tr>
<tr>
<td></td>
<td>Kim, 2011 US Medium</td>
<td>Case-control</td>
<td>See above</td>
<td><strong>ST/FS cases</strong> (n=121): Same as above. <strong>Non-ST/FS fracture control</strong> (n=460): Same as above.</td>
<td>≥5 years</td>
<td>Same as above</td>
<td>OR 2.74 (1.25, 6.02) Socioeconomic status, comedications, drug count, comorbidities, recent medical visits, prior fall, prior osteoporotic fracture, BMD test past 5 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Retrospective cohort</td>
<td>Age 80 Female 97% CHF 22% CKD 3% DM 26% HTN 67%</td>
<td><strong>ST/FS cases</strong> (n=26): Diagnosis codes for subtrochanteric and femoral shaft fracture, but no exclusions for trauma or cancer/pathologic fracture, and no review of radiographs. <strong>Non-ST/FS fracture control</strong>: Not specified.</td>
<td>3-5 years</td>
<td>Raloxifene or calcitonin</td>
<td>HR 1.20 (0.55, 2.61) Propensity score–matched by demographics, healthcare utilization, comorbidities, and other medications</td>
</tr>
<tr>
<td>Outcome</td>
<td>Author, Year Country</td>
<td>Risk of Bias</td>
<td>Study Design</td>
<td>Population Age (mean) Gender (%) Comorbid Conditions (%)</td>
<td>Case and Control Definitions</td>
<td>BP Treatment Duration</td>
<td>Treatment Control Group</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------------------------------------------------</td>
<td>----------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Retrospective cohort (n=2,371 BP users, n=1,726 raloxifene or calcitonin users)</td>
<td>See above</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>
</tr>
</tbody>
</table>

Abbreviations: AFF=atypical femoral fracture; ASBMR=American Society for Bone and Mineral Research; BMI=body mass index; BP=bisphosphonates; CHF=chronic heart failure; CI=confidence interval; CKD=chronic kidney disease; CVD=cardiovascular disease; DM=diabetes mellitus; HR=adjusted hazard ratio; HTN=hypertension; N=number; OR=odds ratio; RR=risk ratio; ST/FS=subtrochanteric/femoral shaft

*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.66 The age-adjusted risk of AFF for bisphosphonate treatment compared with no use of osteoporosis drugs appeared possibly greater in women ≥80 years of age (OR 163 [95% CI 39, 687]) compared to that in women <80 years of age (OR 100 [95% CI 40, 253]). However, 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 in populations with >3 years of bisphosphonate use, risk of these fractures may be higher with longer treatment duration.

One study reported on risk of AFF.66 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 or 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]).65 The second study reported risk for subtrochanteric or femoral shaft fractures with long-term bisphosphonate use versus raloxifene or calcitonin use.63 Risk appeared possibly 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]), though neither result was statistically significant. Confidence intervals overlapped in all studies and none tested for an interaction as a function of treatment duration.
Denosumab

Key Points

Key Question 3
- In postmenopausal women with osteoporosis, between long-term denosumab versus placebo, trial evidence was insufficient to draw conclusions about differences in risk of on-treatment serious adverse events, AFF, or ONJ, or post-treatment fracture risk.
- No controlled observational studies compared long-term treatment with denosumab versus no treatment or another control, let alone reported on risk of harms such as AFF, ONJ, or post-treatment fracture risk.

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

Eligible Studies

Two eligible publications of two trials 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 ROB71 and only limited data were extracted (Appendix Table D7). The remaining publication,70 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 intravenous denosumab dosing regimens (n=319) for 2 years. These dosing regimens included three that were lower and two that were higher than the current FDA recommended denosumab osteoporosis dosing. After 2 years, individuals were nonrandomly assigned to treatment groups for the following 2 years. Those initially assigned placebo remained assigned to placebo. Five of the groups assigned denosumab, including three of the low dose groups, were assigned to denosumab 60 mg every 6 months for 2 years (n=231), one of the original high dose denosumab groups was switched to placebo for 2 years (n=47), and the last denosumab group was switched to placebo for 1 year before starting denosumab 60 mg every 6 months for 1 year (n=41).

Outcomes

Serious Adverse Events

Through 4 years, risk of any serious adverse event was 17.8 percent in the women initially assigned to one of the denosumab groups versus 10.9 percent in the women originally assigned to placebo (RR 1.64 [95% CI 0.69, 3.88]). However, because harms results were reported collectively for all women initially assigned to denosumab, regardless of starting dose or whether they continued denosumab for 4 years, 2 years, or stopped it for 1 year before restarting it for 1
year, no direct comparison between 4 years of continuous denosumab and placebo groups was possible.

**Cardiovascular Events**

One fatal cardiovascular event (stroke) was reported in the collective denosumab group.

**Mortality**

There were four deaths (1.3%) in the collective 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.

**Post-Treatment Fractures**

Authors 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.

**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.
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 deep venous thrombosis at 4 years and of pulmonary embolism at 8 years.
- In postmenopausal women with osteoporosis, 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,72-80, 82, 83, 85 and three eligible publications from one observational study,89-91 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 Multiple Outcomes of Raloxifene Evaluation (MORE) RCT with low ROB compared raloxifene to placebo through 4 years. Two compared raloxifene 60 mg/day to placebo (n=5114 to 5133)77, 85, three compared a pooled raloxifene group (60 mg/day and 120 mg/day) to placebo (n=7617 to 7705)72, 73, 80, and three separately compared raloxifene 60 mg/day and 120 mg/day to placebo (n=6828 to 7705).74-76 Four medium ROB studies from the 4 year Continuing Outcomes Relevant to Evista (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 (doses not specified) to no osteoporosis treatment.89-91 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 followup 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). 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 stroke or 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 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 percentage of participants with pulmonary embolism 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 pulmonary embolism 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]).

Subtrochanteric or Femoral Shaft 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-Defined by Diagnostic Codes Only

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 as defined only by diagnostic codes (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 multivariable-adjusted 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 0.82 [95% CI 0.60, 1.13]), 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 venous thromboembolism, mostly comprised of DVT or pulmonary embolism, was not statistically significantly different between raloxifene 60 mg/day and placebo in the overall study population (HR 1.86 [95% CI 0.97, 3.56]) or in a subgroup at 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, pulmonary embolism, 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 pulmonary embolism between the raloxifene 60 mg/day and 120 mg/day groups, and that risks of venous thromboembolism and DVT (but not of pulmonary embolism) were higher in the pooled raloxifene groups versus placebo through 2 years of treatment but not in later years.
Hormone Therapy

Key Points

Key Question 3

- In postmenopausal women with unknown osteoporosis or osteopenia status:
  - In those with a prior hysterectomy, compared with placebo, 7 years of oral, high dose conjugated estrogen was associated with an increased risk for stroke, deep venous thrombosis, the composite outcome of cardiovascular disease, and the composite outcome of mild cognitive impairment or probable dementia.
  - In those with an intact uterus, compared with placebo, 5.6 years of oral, high dose estrogen plus progestin was associated with an increased risk for stroke, coronary heart disease, deep venous thrombosis, pulmonary embolism, invasive breast cancer, probable dementia, the composite outcome of cardiovascular disease, and the composite outcome of mild cognitive impairment or probable dementia.
- In postmenopausal women with past clinical fracture or osteoporosis by BMD, evidence was insufficient to draw conclusions about any differences in risk for long-term harms between estrogen or estrogen/progestin therapy and control.

Key Question 4

- The association of estrogen plus progestin versus placebo on risk of breast cancer appeared greater in women with increased duration of prior postmenopausal hormone use.
- All other reported tests of interaction for whether the risk of harms associated with either estrogen alone or estrogen plus progestin versus placebo differed by participant characteristic were not statistically significant.

Eligible Studies

As detailed earlier, two large Women’s Health Initiative (WHI) trials randomized postmenopausal women aged 50 to 79 years with unknown osteoporosis or osteopenia status to long-term hormone therapy versus placebo and reported harms. One trial assigned 16,608 women with an intact uterus to oral estrogen (conjugated estrogen 0.625 mg/day) plus progestin (medroxyprogesterone 2.5 mg/day) versus placebo for 5.6 years.94 The other allocated 10,739 women with a prior hysterectomy to oral estrogen (conjugated estrogen 0.625 mg/day) versus placebo for 7 years.95 Two additional WHI reports provided details on risk of cognitive diagnoses with hormone therapy versus placebo.96,97 An additional, 4-year, non-WHI trial, described in greater detail earlier, randomized 36 postmenopausal women with osteoporosis or osteopenia to estrogen/progestin versus placebo.88
Outcomes

Serious Adverse Events

Incidence of serious adverse events was not reported. Only the small non-WHI trial reported on withdrawals due to adverse events (or to other medical problems), which it stated were not different at 4 years between the estrogen/progestin and control groups (17% vs. 17%, p=1.0).

The WHI trial comparing oral estrogen plus progestin versus placebo reported that during an average 5.6 year follow-up period, women with unknown osteoporosis or osteopenia status who were randomized to estrogen plus progestin had a higher risk of stroke (HR 1.41 [95% CI 1.07, 1.85]), coronary heart disease (HR 1.29 [95% CI 1.02, 1.63]), deep venous thrombosis (HR 2.07 [95% CI 1.49, 2.87]), pulmonary embolism (HR 2.13 [95% CI 1.39, 3.25]), invasive breast cancer (HR 1.26 [95% CI 1.00, 1.59]), probable dementia (HR 2.05 [95% CI 1.21, 3.48]), the composite outcome of cardiovascular disease (HR 1.22 [95% CI 1.09, 1.36]), and the composite outcome of mild cognitive impairment or probable dementia (HR 1.44 [95% CI 1.04, 1.99]). However, there was a lower risk of colorectal cancer (HR 0.63 [95% CI 0.43, 0.92]).

The WHI trial comparing oral estrogen versus placebo reported that during an average 7 year follow-up period, women with unknown osteoporosis or osteopenia status who were randomized to estrogen had a higher risk of stroke (HR 1.39 [95% CI 1.10, 1.77]), deep venous thrombosis (HR 1.47 [95% CI 1.04, 2.08]), the composite outcome of cardiovascular disease (HR 1.12 [95% CI 1.01, 1.24]), and the composite outcome of mild cognitive impairment or probable dementia (HR 1.38 [95% CI 1.01, 1.89]). However, there was no significant difference between treatment groups in risk of coronary heart disease (HR 0.91 [95% CI 0.75, 1.12]), pulmonary embolism (HR 1.34 [95% CI 0.87, 2.06]), invasive breast cancer (HR 0.77 [95% CI 0.59, 1.01]), colorectal cancer (HR 1.08 [95% CI 0.75, 1.55]), or probable dementia (HR 1.49 [95% CI 0.83, 2.66]).

To weigh the risk of serious harms against reduction in hip fractures, both WHI trials reported results for a global index, defined as first occurrence of coronary heart disease, invasive breast cancer, stroke, pulmonary embolism, endometrial cancer, colorectal cancer, hip fracture, or death due to other causes. These studies then reported the hazard ratio for the global index outcome in tertiles of study participants stratified by their baseline risk of incident fracture. For estrogen plus progestin compared with placebo, the risk for the global index outcome in women at the lowest, intermediate, and highest fracture risk was HR 1.20 (95% CI 0.93, 1.58), HR 1.23 (95% CI, 1.04, 1.46), and HR 1.03 (95% CI, 0.88, 1.24), respectively. For estrogen compared with placebo, the risk for the global index outcome in women at the lowest, intermediate, and highest fracture risk was HR 0.81 (95% CI 0.62, 1.05), HR 1.09 (95% CI, 0.92, 1.30), and HR 1.04 (95% CI, 0.88, 1.23), respectively.

Mortality

In the WHI trial comparing oral estrogen plus progestin versus placebo, during an average 5.6-year follow-up period, hormone therapy was not associated with an increased risk of all-cause mortality (HR 0.98 [95% CI 0.82, 1.18]). Similarly, in the WHI trial comparing oral estrogen versus placebo, during an average 7-year follow-up period, hormone therapy was not associated with an increased risk of all-cause mortality (HR 1.04 [95% CI 0.88, 1.22]). In the small non-WHI trial conducted in women with osteoporosis or osteopenia, 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

In the WHI trial that compared oral estrogen plus progestin versus placebo, the association of treatment assignment with risk of breast cancer appeared greater with increased duration of prior postmenopausal hormone use. However, the association of treatment with invasive breast cancer did not differ by age, race/ethnicity, family history, parity, age at first birth, body mass index, or Gail-model risk score. The association of treatment group with risk of coronary heart disease did not differ based on history of prior coronary heart disease. The association of treatment group with risk of coronary heart disease, stroke, or venous thromboembolism did not significantly differ by age, race/ethnicity, body mass index, prior hormone use, smoking status, blood pressure, diabetes, aspirin use, or statin use. Last, the associations of estrogen alone and estrogen plus progestin versus placebo on risk for dementia did not significantly differ by any of 13 dementia risk factors evaluated (p for all tests for interaction >0.05).

In the WHI trial that compared oral estrogen versus placebo, the association of treatment assignment with risk of coronary heart disease, stroke, venous thromboembolism, invasive breast cancer, and mortality did not significantly differ by age, race/ethnicity, or body mass index.

We identified no eligible studies in women selected for osteoporosis or osteopenia 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, no treatment, or another active treatment.\(^61\) This study was rated high ROB and only limited data were extracted (Appendix Table 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 placebo, no treatment, or a different active treatment reported on harms.\(^{59,60}\) These studies were rated high ROB and only limited data were extracted (Appendix Table 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 Osteoporosis Drug Holidays

Chapter 6 discusses Key Questions 5 and 6. Key Question 5 addresses the effects of osteoporosis drug treatment 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 osteopenia or osteoporosis (femoral neck [FN] bone mineral density [BMD] T-score ≤ -1.6), alendronate continuation for 5 more years versus alendronate discontinuation was associated with:
  o Lower risk of incident clinical vertebral fractures (moderate strength of evidence [SOE]).
  o No difference in risk of incident clinical fractures (moderate strength of evidence [SOE]), incident nonvertebral fractures (moderate SOE), or incident radiographic vertebral fractures (low SOE).
  o Insufficient evidence about incident hip fractures.

• In postmenopausal women who had received 5 years of alendronate for osteoporosis (lumbar spine [LS]-BMD T-score ≤ -2.5):
  o Between alendronate continuation for 2 more years versus alendronate discontinuation, evidence was insufficient to draw conclusions about differences in risk of incident nonvertebral fractures or incident clinical vertebral fractures.
  o Between alendronate continuation for 5 more years versus alendronate continuation for 2 more years followed by 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 ≤ -1.6 (osteopenia or osteoporosis):
  o 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.
  o 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.
  o 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 risk of bias (ROB)\(^{33, 41, 42}\) and one as having medium ROB.\(^{34}\) 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 Fracture Intervention Trial (FIT) study.\(^{33, 41}\) Eligibility in FLEX was limited to women assigned to the alendronate treatment arm 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 blinded follow-up in FIT (mean 3.8 years) and a subsequent open-label period (mean 1.9 years). Further, they must have had a total hip (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) (Table 13). 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).\(^{33}\) 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\(^{34, 42}\) were extensions of a pair of nearly identical dose-ranging randomized controlled trials (RCT), one conducted in the U.S. and the other in several non-U.S. countries.\(^{98}\) 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 64 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 a total of 5 years from the beginning of the study, 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.\(^{42}\) 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.\(^{34}\) 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.
Table 13. Trial participant characteristics at the time of assignment to alendronate continuation versus discontinuation

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>FLEX 32,40</th>
<th>Main Study* (Excluded)52</th>
<th>Extension 2* (Included)41</th>
<th>Extension 3* (Included)33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number enrolled and randomized</td>
<td>N=1099</td>
<td>N=994</td>
<td>N=350</td>
<td>N=247</td>
</tr>
<tr>
<td>Age, mean years</td>
<td>73</td>
<td>64</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>Gender, women</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Race, white</td>
<td>97%</td>
<td>87%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mean FN-BMD (g/cm²)</td>
<td>0.61</td>
<td>Hologic 0.60</td>
<td>Lunar 0.73</td>
<td>Norland 0.66</td>
</tr>
<tr>
<td>Mean LS-BMD (g/cm²)</td>
<td>0.90</td>
<td>Hologic 0.71</td>
<td>Lunar 0.81</td>
<td>Norland 0.67</td>
</tr>
<tr>
<td>Mean TH-BMD (g/cm²)</td>
<td>0.73</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>FN-BMD T-score ≤-2.5</td>
<td>29%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>LS-BMD T-score ≤-2.5</td>
<td>NR</td>
<td>100%</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>TH-BMD T-score ≥-3.5</td>
<td>100%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Prior fracture ≥45 years</td>
<td>60%</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Prevalent radiographic vertebral fracture</td>
<td>34%</td>
<td>21%</td>
<td>21%</td>
<td>25%</td>
</tr>
<tr>
<td>Comorbid conditions</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: BMD=bone mineral density; FLEX=Fracture Intervention Trial Long-Term Extension; FN=femoral neck; LS=lumbar spine; NR=not reported; TH=total hip

*Dose ranging study

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 drug holiday) had reduced risk of incident clinical vertebral fracture (2.4% vs. 5.3%; risk ratio [RR]0.45 [95% confidence interval (CI) 0.24, 0.85]) (moderate SOE). However, there was no difference between continuation and drug holiday in risk of any incident clinical fracture (19.9% vs. 21.3%; RR 0.93 [95% CI 0.71, 1.21]) (moderate SOE), or nonvertebral fracture (18.9% vs. 19.0%; RR 1.00 [95% CI 0.76, 1.32]) (moderate SOE). Evidence was insufficient to draw conclusions about differences in risk of hip fracture (3.0% vs. 3.0%; RR 1.02 [95% CI 0.51, 2.10]).

In the dose-ranging extension study, in results pooling both alendronate continuation arms versus the discontinuation arm, for the year 6-7 extension, evidence was insufficient for differences between treatment groups for both incident nonvertebral fracture (RR 0.87 [95% CI
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). In the dose-ranging extension study, continued alendronate versus discontinuation did not reduce risk of incident radiographic vertebral fracture between years 6-10, whether considering the 5 mg/day and 10 mg/day alendronate continuation groups together (RR 1.40 [95% CI 0.52, 3.74]) or separately (calculated by Evidence-based Practice Center).

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%; mean difference [MD] 2.36 [95% CI 1.81, 2.90]). 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 studies 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, during this period neither TH-BMD or FN-BMD significantly changed while LS-BMD increased in both continuing alendronate dose groups.

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. By comparison, 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 in women without a prevalent radiographic vertebral fracture at baseline, 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, neither the effect of continued versus discontinued alendronate on risk of incident clinical vertebral fracture nor incident radiographic vertebral fracture appeared 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).
Zoledronic Acid (Zoledronate)

Key Points

Key Question 5

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

• In postmenopausal women with osteoporosis, between zoledronate continued for 6 years and zoledronate stopped after 3 years and switched to placebo for 3 years:
  o There was a lower risk of incident radiographic vertebral fractures (low SOE).
  o There was no difference in risk of incident clinical fractures (moderate SOE) or incident nonvertebral fractures (moderate SOE).
  o 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 zoledronate continued for 9 years and zoledronate stopped after 6 years and switched to placebo for 3 years:
  o 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 zoledronate 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 zoledronate versus discontinuation of zoledronate (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 Table 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, enrolled postmenopausal women with osteopenia (LS-BMD < -1 and > -2.5, FN-BMD > -2.5, no grade two or three baseline radiographic vertebral fracture, and zero to one grade one baseline radiographic vertebral fracture). Participants were randomized to zoledronate 5 mg once annually for 2 years (Z2) versus zoledronate 5 mg once annually for 1 year followed by 1 year of placebo (Z1/P1). 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). This trial randomized participants to zoledronate 5 mg once annually for 3 years versus placebo. Following the initial 3-year trial period, women initially randomized to zoledronate were
randomized to continue zoledronate 3 more years for a total of 6 years (Z6) versus placebo for three more years (Z3/P3). A second extension was conducted in which women randomized to zoledronate for 6 years were then randomized to remain on zoledronate 3 more years for a total of 9 years (Z9) or switch to placebo for 3 years (Z6/P3). Table 14 summarizes baseline participant characteristics of the zoledronate continuation versus discontinuation trials.

Table 14. Participant characteristics at the baseline of the zoledronate continuation versus discontinuation trials

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>HORIZON, First Extension54</th>
<th>HORIZON, Second Extension55</th>
<th>McClung56</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*</td>
<td>-1.98*</td>
<td>NR</td>
</tr>
<tr>
<td>FN-BMD T-score &lt; -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 Evidence-based Practice Center.

Outcomes

Incident Clinical Fractures

In the McClung trial, incident fractures were neither a primary nor secondary endpoint, but were reported as adverse events. 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 zoledronate 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]) (moderate SOE) or incident nonvertebral fractures (7.6% vs. 8.2%; HR 0.99 [0.7, 1.5]) (moderate 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 differences between the Z9 and Z6/P3 treatment groups in risk of incident clinical fractures (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]) (low SOE). During the second HORIZON extension (years 7-9), there were few incident radiographic vertebral fractures and evidence was insufficient about whether risk differed between the Z9 and Z6/P3 treatment groups (3.2% vs. 5.3%; OR 0.58 [95% CI 0.13, 2.55]).

Change in BMD

The primary study endpoints in the McClung trial and the first and second 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 zoledronate continuation versus discontinuation (placebo drug holiday) and assessed whether risk of incident fracture varied as a function of 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).
- Evidence was insufficient to draw conclusions about whether placebo following prior denosumab treatment was associated with rapid bone loss compared to continued denosumab treatment.

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, were 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. These dosing regimens included three that were lower and two that were higher than the current U.S. Food and Drug Administration recommended denosumab osteoporosis dosing. After 2 years, individuals were nonrandomly assigned to treatment groups for the following 2 years. Those individuals initially assigned placebo remained assigned to placebo. Five of the groups assigned denosumab, including three of the low dose groups, were assigned to denosumab 60 mg every 6 months for 2 years (n=231), one of the original high dose denosumab groups was switched to placebo for 2 years (n=47), and the last denosumab group was switched to placebo for 1 year before starting denosumab 60 mg every 6 months 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, among women originally randomized to denosumab, 10.5 percent had an incident clinical fracture and 7.0 percent had an incident osteoporotic fracture (any clinical fracture excluding those of the phalanges, face or those caused by severe trauma). Authors 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, the increase in LS-BMD and TH-BMD between baseline and year 4 were 9.0 percent and 3.9 percent, respectively. 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. Authors did not report changes in BMD during the placebo periods following prior denosumab.

**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 eligible 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 eligible 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 eligible 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 eligible 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 eligible 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 eligible 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 femoral neck (FN)-bone mineral density (BMD) T-score ≤ -1.6 (osteopenia or osteoporosis):
  - 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 osteonecrosis of the jaw (ONJ), atypical femoral fracture (AFF), or subtrochanteric or femoral shaft fracture.
- In postmenopausal women who had received 5 years of alendronate for lumbar spine (LS)-BMD T-score ≤ -2.5:
  - Between alendronate continuation for 2 years versus discontinuation, there was no difference in risk for serious adverse events (low strength of evidence [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 risk of bias (ROB)\(^3^3, 3^7, 4^3\) and three as having medium ROB.\(^3^4, 4^2, 5^3\) 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\(^3^4, 4^2\) or a mix of osteoporosis and osteopenia/low bone mass\(^3^3, 4^1, 4^3\) when they had been randomly assigned to the alendronate arm of a placebo controlled randomized controlled
trial (RCT) 5 years earlier. After 5 years of alendronate, the Fracture Intervention Trial Long-Term Extension (FLEX) study randomized 1099 women to continuation versus discontinuation\textsuperscript{33, 41, 43} and the other study nonrandomly assigned two of three alendronate dose ranging groups to continued alendronate and the third alendronate group to placebo\textsuperscript{34, 42}.

**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{33}. 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{37}. 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\%; risk ratio [RR] 1.05 [95\% confidence interval [CI] 0.57, 1.96]) (low SOE)\textsuperscript{42}. 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{34}

**Upper Gastrointestinal Adverse Events**

Participants randomized to alendronate continuation for 5 years versus discontinuation in FLEX did not appear to differ in risk of upper gastrointestinal (GI) tract adverse events, but no numerical data were provided\textsuperscript{33}. 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{37}. 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{42} 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{34}

**Subtrochanteric or 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\%; hazard ratio [HR] 1.33 [95\% CI 0.12, 14.67] (insufficient SOE)\textsuperscript{43}.

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{53} Alendronate accounted for 99\% 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 International Classification of Disease (ICD)\textsuperscript{9th} Edition diagnosis codes only, with no documentation that radiographs were
reviewed for features of atypical femoral fractures. Incidence of subtrochanteric or femoral shaft fractures was rare, with 44 (0.15%) in the combined persistent and non-persistent continual use groups combined and three (0.03%) in the drug holiday group (odds ratio [OR] 6.03 [95% CI 1.87, 19.42]) (low SOE).

**Osteonecrosis of the Jaw**

In FLEX, authors reported that there were no cases of ONJ.33

**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 that assessed whether the risk of harms between alendronate continuation and discontinuation vary as a function of patient, bone or drug characteristics.
Zoledronic Acid (Zoledronate)

Key Points

Key Question 7
- In postmenopausal women with osteopenia:
  - Between zoledronate for 2 years versus zoledronate 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 zoledronate for 6 years versus zoledronate for 3 years followed by placebo for 3 years, there was no difference in risk of serious adverse events (low SOE).
  - Between zoledronate for 9 years versus zoledronate 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 zoledronate continuation and discontinuation.

Key Question 8
- We identified no eligible studies that assessed whether differences in risk of harms between zoledronate 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 zoledronate 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.

In the McClung trial, 379 postmenopausal women with osteopenia were randomized to zoledronate for two years versus zoledronate for one year followed by placebo for one year. In the first Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly-Pivotal Fracture Trial (HORIZON) trial extension, 1233 postmenopausal women who previously had been randomized to receive zoledronate for 3 years for osteoporosis were randomized to receive zoledronate for another 3 years versus placebo. In the second HORIZON trial extension, 190 women who had been randomized to zoledronate in the first extension were randomized to receive zoledronate for another 3 years versus placebo.

Outcomes

Serious Adverse Events

In the McClung trial, risk of harms was not separately reported for 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 zoledronate continuation versus discontinuation (9.4% vs. 10.6%; RR 0.91 [95% CI 0.50, 1.67]) (moderate SOE).\textsuperscript{56} 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 zoledronate continuation and discontinuation groups, neither during years 4 to 6 (31% vs. 27%; RR 1.14 [95% CI 0.96, 1.36]) (low SOE) nor during years 7 to 9 (26% vs. 30%; RR 0.86 [95% CI 0.54, 1.36]) (low SOE).\textsuperscript{54, 55}

**Atypical Femoral Fractures**

Both HORIZON extension studies reported that no cases of AFF occurred in either the zoledronate continuation or discontinuation group during follow-up,\textsuperscript{54, 55} with documentation that radiographs were reviewed for possible AFF features in just one of these reports\textsuperscript{55} (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{56} In the HORIZON extension studies, one case of ONJ was reported in the zoledronate continuation group and none in the zoledronate discontinuation group during years 4 to 6 (insufficient SOE),\textsuperscript{54} and no cases of ONJ occurred in either group during years 7 to 9 (insufficient SOE).\textsuperscript{55}

**Atrial Fibrillation**

The McClung trial reported that no cases of atrial fibrillation occurred in either the zoledronate continuation or discontinuation 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{54} or between years 7 to 9 (5.4% vs. 1.1%, p=0.11).\textsuperscript{55}

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

We identified no eligible studies that compared zoledronate 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, including for risk of AFF, ONJ, and fractures after denosumab discontinuation.

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. These dosing regimens included three that were lower and two that were higher than the current U.S. Food and Drug Administration recommended denosumab dosing. After 2 years, individuals initially assigned placebo remained assigned to placebo. Five of the groups assigned denosumab, including three of the low dose groups, were assigned to denosumab 60 mg every 6 months for 2 years (n=231), one of the original high dose denosumab groups was switched to placebo for 2 years (n=47), and the last denosumab group was switched to placebo for 1 year before starting denosumab 60 mg every 6 months for 1 year (n=41).

Outcomes

Serious Adverse Events

Through 4 years, risk of any serious adverse event was 17.8 percent in the women initially assigned to one of the denosumab groups. However, because harms results were reported collectively for all women initially assigned to denosumab, regardless of whether they continued denosumab for 4 years, 2 years, or stopped it for 1 year before restarting it for 1 year, no direct comparison between denosumab continuation and discontinuation groups was possible.

Atypical Femoral Fracture

No information was reported about incidence of AFF.

Osteonecrosis of the Jaw

No information was reported about incidence of ONJ.
Post-Treatment Fractures
Authors 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.

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 eligible 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 eligible 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 eligible 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 eligible 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 eligible 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 eligible 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

Prior systematic reviews have addressed the well-established randomized controlled trial (RCT) evidence on the efficacy of numerous U.S. Food and Drug Administration (FDA) approved medications for short-term reduction in risk of vertebral and nonvertebral fractures, and of bisphosphonates and denosumab for short-term reduction in risk of hip fractures.\textsuperscript{15,18,100} This literature was not the focus of the current report.

Rather, in the current systematic review, we examined evidence pertaining to the clinical question of how long to treat with osteoporosis drugs in patients without prior treatment and what to do after completing a course of short-term osteoporosis drug treatment. More specifically, we evaluated the evidence on the benefits and harms of longer-term osteoporosis drug treatment (>3 years) and the benefits and harms of osteoporosis drug holidays. We defined osteoporosis drug holidays as discontinuation of treatment for at least 1 year after at least 1 year of prior treatment. We also examined whether the effects of these interventions vary as a function of patient, bone, or osteoporosis drug characteristics.

We limited the review of efficacy to RCTs and controlled clinical trials (CCT). For harms, we also evaluated observational studies, requiring contemporaneous control groups to limit bias. We analyzed data only from studies judged to have low or medium risk of bias (ROB). We graded strength of evidence (SOE) for the most clinically important, patient-centered outcomes of incident clinical fractures, incident radiographic vertebral fractures, serious adverse events (SAEs), atypical femoral fractures (AFF), osteonecrosis of the jaw (ONJ), and post-treatment fractures.

Key Findings and Strength of Evidence

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

There were few eligible trials of long-term (>3 years) osteoporosis drug treatment versus placebo. All participants in these studies were postmenopausal women.

In women with osteoporosis, we found high strength evidence that both alendronate and raloxifene for 4 years reduced the risk of radiographic vertebral fractures (~35-45% relative risk reduction [RRR] and 2-5% absolute risk reduction [ARR]), and that raloxifene reduced risk of clinical vertebral fractures (~40% RRR and 2% ARR). However, neither significantly reduced risk of any type of nonvertebral fracture. In women with osteopenia or osteoporosis, we found high strength evidence that zoledronate for 6 years versus placebo reduced the risk of incident nonvertebral fracture (~35% RRR and 5% ARR) and incident clinical vertebral fracture (~60% RRR and 2% ARR), and moderate strength evidence that zoledronate reduced risk of incident clinical fracture (~25% RRR and 5% ARR). In women with unknown osteoporosis or osteopenia status, we found that both oral estrogen/progestin (5.6 years in those with an intact uterus) and unopposed oral estrogen (7 years in those with prior hysterectomy) versus placebo reduced the risk of both incident clinical fracture (~25-30% RRR, high SOE) and incident hip fracture (~35% RRR, moderate SOE). Evidence was insufficient about efficacy of other treatments for long-term fracture prevention.
Key Question 2: Variation in Efficacy of Long-Term Osteoporosis Drug Treatment as a Function of Patient, Bone, or Osteoporosis Drug Treatment Characteristics

Efficacy of long-term alendronate appears to vary as a function of baseline bone mineral density (BMD). Between long-term alendronate and placebo, there was a significant 35 percent RRR and 6-7% ARR in incident clinical fractures among women with osteoporosis, but not in women with osteopenia. Similarly, women with osteoporosis had a 55 percent RRR and 1 percent ARR in 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 different. Women with osteoporosis had a significant 50 percent RRR and 3 percent ARR in incident radiographic vertebral fractures. By comparison, women with an osteopenic femoral neck (FN)BMD T-score of -2.5 to -2 had a similar 45 percent RRR in incident radiographic vertebral fractures with long-term alendronate versus placebo, but experienced fewer events, confidence intervals were wider, and results were not statistically significant. Neither past nonvertebral fracture, World Health Organization Fracture Assessment Tool (FRAX) 10-year major osteoporotic fracture 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. Few of several tests for interaction suggested that the effect of oral estrogen/progestin or unopposed oral estrogen versus placebo on risk of incident clinical fracture or incident hip fracture varied by patient characteristics, though results suggested a possibly greater benefit in women aged 60-79 years or more than 20 years since menopause compared to younger women or those with more recent menopause. Strength of evidence was not assessed for effect modifiers. We found no other evidence from eligible studies about whether efficacy of long-term treatment with any other osteoporosis drugs varied by patient, bone or drug characteristics.

Key Question 3: Harms of Long-Term Osteoporosis Drug Treatment

RCT data provided insufficient evidence about whether long-term alendronate, zoledronate, or bisphosphonates as a class increase risk of radiologically confirmed AFF, subtrochanteric or femoral shaft fractures without confirmed AFF features, or ONJ; too few events were observed in eligible trials.

Data from controlled, long-term observational studies suggested that alendronate and bisphosphonates as a class increase risk of both radiologically confirmed AFF (low SOE), and subtrochanteric and femoral shaft fractures that have not been radiologically confirmed to have AFF features (low SOE). Relative risk estimates for these outcomes varied widely across studies, likely related to heterogeneity in study designs. Magnitudes of relative risk appeared higher in studies that defined cases as radiologically confirmed AFF than in those that relied only on diagnostic codes without more specific radiologic confirmation for AFF features. However, some of these AFF risk estimates, including from one recently published study, may have been biased because they used a control group of patients with subtrochanteric or femoral shaft fractures without AFF features. Results from these studies reflect the probability that a subtrochanteric or femoral shaft fracture, if it occurs, will have AFF features, not the relative risk of sustaining an AFF per se. Further, because there were no cohort studies that reported cases of radiologically confirmed AFF compared to nonfracture controls, it was not possible from eligible studies to estimate population incidence of AFF attributable to alendronate or
Several long-term observational studies of alendronate collectively suggested an association with increased risk of ONJ (insufficient to low SOE). However, as with AFF, the magnitude of relative risk estimates varied widely between studies, likely due to heterogeneity in case definitions, treatment control groups, and covariate modeling. ONJ incidence rates from eligible cohort studies also varied widely. We found insufficient evidence about whether long-term zoledronate increases risk of AFF or ONJ. The single trial of long-term zoledronate versus placebo reported no cases of AFF or ONJ in either treatment group, and we identified no eligible observational studies that evaluate the risk of these harms with long-term zoledronate versus any control group.

Due to authors’ pooling of short- and long-term results, it was not possible to conclude anything about the risk of harms between long-term denosumab and placebo from the one study that met eligibility for this review. The long-term trials of estrogen/progestin and estrogen versus placebo in postmenopausal women with unknown osteoporosis or osteopenia status identified a host of increased harms risks, including stroke, deep venous thrombosis, cardiovascular disease, and the composite outcome of mild cognitive impairment or probable dementia. In addition, estrogen/progestin was associated with an increased risk of coronary heart disease, pulmonary embolism, and invasive breast cancer. Long-term raloxifene versus placebo significantly increased the risk of deep vein thrombosis (DVT) and pulmonary embolism (PE).

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 risks of harms with long-term osteoporosis drug treatment. 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. Trials of long-term hormone therapy evaluated whether risk of harms varied by a long list of patient characteristics. Though results suggested that risk of breast cancer with estrogen/progestin compared to placebo may be greater with increased duration of prior postmenopausal hormone use, this was the only statistically significant result out of many tests of interaction, and may have been attributable to chance. Strength of evidence was not assessed for effect modifiers.

Key Question 5: Effects of Osteoporosis Drug Treatment Holidays

RCT data showed that continuation of alendronate for 10 years versus 5 years and continuation of zoledronate for 6 years versus 3 years each lowered risk of one of two measures of incident vertebral fracture, but did not lower risk of any type of nonvertebral fracture. Data were insufficient about whether there is a difference in risk of fractures between zoledronate for 9 years versus 6 years. All these findings were reported in prior systematic reviews. Data from two additional trials, neither recently published, were insufficient to draw conclusions about differences in fracture risk between continuation versus discontinuation of zoledronate for 2 years versus 1 year, or for denosumab 4 years versus either 2 years followed only by placebo or
2 years followed by placebo and later denosumab resumption. We did not consider observational studies for evidence about risk of fractures with osteoporosis drug treatment holidays due to concerns about confounding by indication.

**Key Question 6: Variation in Effect of Osteoporosis Drug Treatment Holidays as a Function of Patient, Bone, or Osteoporosis Drug Treatment Characteristics,**

The little published data about whether the effect of osteoporosis drug continuation versus discontinuation on risk of incident fracture varied by patient, bone or drug characteristics was available only for alendronate. A post hoc subgroup analysis suggested that while alendronate continuation versus discontinuation did not reduce risk of any nonvertebral fracture overall, in the subset of women without a prevalent radiographic vertebral fracture, those with osteoporotic FN-BMD may be more likely than those with osteopenic FN-BMD to have a reduced risk of incident nonvertebral fracture. However, risk of other incident fractures with continued versus discontinued treatment in this patient subgroup did not differ by BMD. In the absence of consistent findings, it is possible this outcome was due to chance. 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 zoledronate 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). For denosumab, one trial assigned postmenopausal women to continuous denosumab treatment for 4 years versus treatment for 2 years followed by placebo versus treatment for 2 years followed by 1 year of placebo and retreatment for 1 year. Authors stated that no increase in fracture incidence was observed among the few women who discontinued denosumab treatment, but reported no numerical data comparing post-treatment fractures risk with fracture risk in women assigned to continue denosumab, those assigned continuous placebo, and those assigned to restart denosumab after a placebo interval.

**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 oldest old, and patients with multiple comorbid conditions common in primary care settings is unknown. The evidence on the efficacy of long-term alendronate, denosumab, raloxifene, and hormone therapy versus placebo on risk of incident fracture came almost entirely from trials conducted entirely in the U.S., whereas the long-term zoledronate trial was conducted at a single site in New Zealand. So, the applicability of results to other geographic settings is uncertain. Trial populations often were selected to be low-risk for specific treatment harms (e.g., upper gastrointestinal 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 randomized study phase, their participants may be even 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 Agency for Healthcare Research and Quality (AHRQ) review, a related 2017 American College of Physicians Clinical Practice Guideline, and a subsequently published trial of abaloparatide, there is well-established evidence from short-term RCTs (18 to 36 months) in postmenopausal women with osteoporosis that alendronate, risedronate, ibandronate, zoledronate, denosumab, teriparatide, abaloparatide, and raloxifene lower risk of incident vertebral fractures. All these agents except ibandronate and raloxifene also lower short-term risk of incident nonvertebral fractures, and alendronate, zoledronate, and denosumab also lower short-term risk of incident hip fractures. RCT evidence on short-term fracture protection in men comes primarily from a single 2-year trial of zoledronate in men with osteoporosis that showed a reduction in risk of incident radiologic vertebral fractures.

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

Efficacy of Long-Term Osteoporosis Drug Treatment

Long-term placebo-controlled trials of osteoporosis drugs versus placebo address the question of whether a long-term course of osteoporosis drug treatments lowers risk of fractures compared to a long-term placebo. They do not directly address the question of whether patients who have completed short-term osteoporosis drug treatment will lower their risk of fractures by continuing that treatment.
The few long-term (>3 year duration), placebo-controlled trials that were included in the prior AHRQ review on osteoporosis drug treatments, and were eligible for the current review and reported interpretable data evaluated alendronate, raloxifene, estrogen/progestin, and unopposed estrogen, respectively. A 4-year alendronate trial in postmenopausal women with osteopenia or osteoporosis, and without a baseline radiographic vertebral fracture, showed a reduction in risk of incident radiographic vertebral fractures and, in the subgroup of women with osteoporosis, also showed a reduction in risk of incident clinical fractures. A 4-year raloxifene trial in postmenopausal women with osteoporosis showed a reduction in clinical vertebral fractures and radiographic vertebral fractures but no reduction in nonvertebral fractures. Two long-term trials in postmenopausal women with unknown osteoporosis or osteopenia status that compared unopposed oral estrogen for 7 years and oral estrogen/progestin for 5.6 years, respectively, versus placebo, reported reduced risks in both incident hip and incident clinical fractures.

Findings from the recently published 6-year RCT comparing zoledronate versus placebo in older women with osteopenia or osteoporosis are a substantial addition to this prior evidence base on the efficacy of long-term osteoporosis drug treatment versus long-term placebo. Zoledronate treatment was associated with significant reductions in risk of both nonvertebral fractures and clinical vertebral fractures. Where reported, the magnitudes of the relative reductions in fracture risk with zoledronate versus placebo look similar between the overall study population and in the osteopenic subset. In contrast, our review found no eligible studies with interpretable trial data on the long-term efficacy of denosumab versus placebo for fracture reduction, and no placebo-controlled trial data on the efficacy of long-term treatment with any sequential osteoporosis drug treatment regimen (e.g., anabolic followed by antiresorptive, denosumab followed by bisphosphonate). Further, we found no eligible long-term trials directly comparing different active osteoporosis drug treatment regimens.

Harms of Long-Term Osteoporosis Drug Treatment

While the prior AHRQ review reported a significant 10 percent relative increase in risk of mild upper gastrointestinal events with bisphosphonates versus placebo, the current review of long-term treatment found no difference in risk. 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 Multiple Outcomes of Raloxifene Evaluation (MORE) trial and follow-up Continuing Outcomes Relevant to Evista (CORE) study found a large magnitude, but borderline statistically 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 and femoral shaft fractures and AFF considered as a pooled outcome (risk ratio [RR] 1.70 [95% CI 1.22, 2.37]). In stratified analysis, that meta-analysis reported that risk in studies that used the American Society for Bone and Mineral Research (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 included new studies that examined the association of long-term bisphosphonate treatment with AFF as defined by ASBMR criteria. However, study methodology still varies in other ways that likely affected their risk estimates (e.g., selection of exposure control, selection of fracture control, model covariables). In addition, whereas prior reviews have estimated an incidence rate of 0.8 AFFs per 10,000 women treated with 3 years of
bisphosphonate treatment, we found that it was not possible from the eligible studies in the current review to estimate a population incidence rate of AFF with long-term treatment. 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 found these same challenges when limiting our review to studies reporting long-term osteoporosis drug treatment. For example, long-term studies differed regarding whether they relied only on diagnostic codes, whether they required that clinical findings were persistent, whether they required pathologic or radiologic diagnostic confirmation, and whether they excluded cases that may have been caused by other factors, such as local radiation.

**Effect of Osteoporosis Drug Treatment Holidays**

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 of how long patients should be treated with drugs to prevent fracture.

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 zoledronate 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 vertebral fracture or of any nonvertebral 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 Fracture Intervention Trial Long Term Extension (FLEX) trial and in risk of incident radiographic vertebral fracture with continued zoledronate in the Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly (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 somewhat limited. The current review did not identify additional trials that addressed this issue. More studies are needed to clarify the effects of bisphosphonate continuation versus discontinuation 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 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 effects of long-term or continuing osteoporosis drug treatment on fracture risk and harms, 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. By comparison, we found one post hoc, subgroup analysis that suggested that in women without a prevalent radiographic vertebral fracture, those with lower FN-BMD may have a greater reduction in risk of incident nonvertebral fracture with continued versus discontinued alendronate. These results were not corroborated by differences in risk of any other fracture type examined, or in women with prevalent radiographic vertebral fractures, and this may have been a chance finding. Though preliminary data from the Foundation for the National Institutes of Health (FNIH) Bone Quality project suggest that early treatment changes in BMD and, to a lesser extent, bone formation markers, may predict the anti-fracture efficacy of antiresorptive osteoporosis drugs, the applicability of these data to long-term treatment is unknown. 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 Comparative Effectiveness Review**

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 National Institute of Health (NIH) Office of Disease Prevention 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, most of which were designed with BMD change rather than incident fracture as the primary outcome. There were no long-term trials of denosumab that provided usable data on the risk of incident fractures. There also were no long-term trials of either teriparatide or abaloparatide, both of which have product labeling recommending against lifetime use for more than 2 years. There 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 Fracture Intervention Trial (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. Studies that compared risk of AFF to that of subtrochanteric or femoral shaft fractures may have provided an exaggerated estimate of the risk of AFF with treatment, instead estimating the risk that patients who have subtrochanteric or femoral shaft fractures have AFF features. 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 also could have affected risk estimates. Still other differences in study design could have affected risk estimates. Examples include differences in treatment control group (e.g., no bisphosphonate use, past bisphosphonate use, raloxifene or calcitonin use), and in the covariables 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. For example, studies that examined the risk of ONJ with long-term treatment differed regarding whether they defined cases only based on diagnostic codes, whether they required that clinical findings were persistent, whether they required pathologic or radiologic diagnostic confirmation, and whether they excluded cases that may have been caused by other factors, such as local radiation.
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 were not always 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 15 summarizes the areas needing future research based on the gaps identified in this review.

Table 15. Future research recommendations

<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</td>
<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 generalizable to many of the patients with osteoporosis in clinical settings, due to their higher prevalence of comorbidities.</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>
</tr>
<tr>
<td>In spite of the availability of these effective short-term treatment options to reduce fracture risk, few men and women with osteoporosis receive these treatments and these numbers are declining. Among older U.S. adults with hip fractures, not treating with osteoporosis drugs has been estimated to increase risk of future nonvertebral fractures. Even among patients who start osteoporosis drug treatment, adherence and persistence with treatment are poor.</td>
<td>Though potential undertreatment of appropriate patients with osteoporosis drugs was outside the scope of this review, future research should examine how patients and clinicians make decisions about these treatments.</td>
<td></td>
</tr>
</tbody>
</table>
**Key Question**

KQ1: 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 for reducing risk of incident fracture and on change in BMD?

**Research Gap**

- 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.
- There were no eligible long-term placebo-controlled trials with adequate power to assess risk of clinical fracture endpoints that compared long-term ibandronate, risedronate, or denosumab with placebo.
- There were no eligible long-term trials of sequential treatment with adequate power to assess risk of clinical fracture endpoints.
- There were no eligible long-term placebo-controlled trials with adequate power to assess risk of clinical fracture endpoints that compared long-term teriparatide or abaloparatide with placebo, likely related to associations of these treatments with osteosarcoma in animals.

**Future Research Recommendations**

- Future RCTs of long-term osteoporosis drug treatment should be designed with adequate power to compare risk of clinical fracture endpoints.
- Future RCTs with adequate power to compare risk of clinical fracture endpoints should be conducted for osteoporosis drugs that don’t yet have established long-term fracture prevention efficacy, including ibandronate, risedronate, and denosumab.
- Future RCTs with adequate power to compare risk of clinical fracture endpoints should be conducted to compare sequential therapy, including denosumab followed by bisphosphonate therapy, and anabolic therapy (e.g. teriparatide or abaloparatide). Studies should include a continuous long-term antiresorptive treatment control group.
- Future RCTs with adequate power to compare risk of clinical fracture endpoints should be conducted to compare anabolic therapy (e.g. teriparatide or abaloparatide) followed by antiresorptive therapy versus antiresorptive therapy alone.
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gap</th>
<th>Future Research Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future long-term placebo-controlled osteoporosis drug trials and comparative effectiveness trials will be expensive and time consuming. Adequate surrogates for incident fracture 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 suggest that early changes in BMD explain a larger percentage of the treatment effect than do changes in bone markers, and do so for the effect of treatment on both vertebral and nonvertebral fractures.</td>
<td>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>
</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>
</tbody>
</table>
### Key Question

**KQ2:** Among men and postmenopausal women aged ≥50 years with osteoporosis* or osteopenia/low bone mass,† does efficacy of long-term (>3 years) osteoporosis drug therapy for reducing risk of incident fracture vary as a function of patient, bone, or osteoporosis drug characteristics?

**Research Gap**

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.

Guidelines are recommending pharmacologic fracture prevention therapy for individuals with elevated estimated fracture risk even in the absence of trials showing fracture reduction with treatment of this population.

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.

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.

**Future Research Recommendations**

Future trials of long-term osteoporosis drug treatment should include populations who have not been included in prior trials, such as men, individuals who are frail and have multiple comorbidities, nonwhite women, and the oldest old. This would be 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.

Future studies should be conducted in populations who have high fracture risk in the absence of osteoporosis by either BMD, past hip fracture, or prevalent vertebral fracture criteria.

### KQ3: 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?

**Research Gap**

Harms reporting in published long-term osteoporosis drug trials have not always 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.

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.

**Future Research Recommendations**

Future long-term osteoporosis drug trials should systematically collect, analyze and report 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.

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.
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Research Gap</th>
<th>Future Research Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<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 ≥8 weeks after recognition by a healthcare provider, and no history of craniofacial radiation.</td>
<td></td>
</tr>
<tr>
<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 retrospective and prospective cohort studies should be conducted to estimate 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>
<td></td>
</tr>
<tr>
<td>Accelerated bone turnover and rapid bone loss occur after denosumab discontinuation. Case series and a large prospective study using a virtual placebo in lieu of a control group suggest that risk of multiple radiographic vertebral fractures is increased after denosumab discontinuation. It is uncertain whether this risk of post-denosumab vertebral fractures appears increased only because of removal of an effective treatment or it truly is increased compared to if the patients never had received denosumab. The magnitude and duration of risk of post-denosumab vertebral and nonvertebral fractures are unclear. The efficacy of different post-denosumab antiresorptive treatment regimens for fracture prevention are unknown.</td>
<td>Future research should: (1) better estimate the magnitude and duration of risk for post-denosumab vertebral and nonvertebral fractures; (2) clarify whether post-denosumab fracture risk only appears increased because of removal of an effective treatment or truly is increased compared to if patients had not received denosumab at all; and (3) evaluate the efficacy of different post-denosumab antiresorptive regimens for fracture prevention.</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>KQ4:</strong> 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 are 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. Studies should use common methods to enable pooling of results and/or validation of findings. Future trials and prospective cohort studies that compare long-term osteoporosis drug treatment versus placebo and systematically collect harms outcomes, should plan a priori to examine possible effect modifiers of harms outcomes. Future research should investigate whether findings of extended DXA femur scans may independently predict risk of atypical femoral fracture in patients currently receiving bisphosphonate or denosumab treatment for osteoporosis.</td>
</tr>
<tr>
<td><strong>KQ5:</strong> Among men and postmenopausal women aged ≥50 years currently receiving drug therapy started for osteoporosis* or osteopenia/low bone mass† to prevent fracture, what is the effect of osteoporosis drug treatment holidays on incident fracture risk and improving BMD?</td>
<td>Past trials of osteoporosis drug treatment continuation versus discontinuation have been too small to be adequately powered to evaluate the effects on incident clinical fracture outcomes.</td>
<td>Larger trials are needed that compare osteoporosis drug treatment continuation versus discontinuation and are powered to look at incident clinical fracture outcomes. 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>Key Question</td>
<td>Research Gap</td>
<td>Future Research Recommendations</td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>KQ6: Among men and postmenopausal women aged &gt;50 years currently receiving</td>
<td>Past trials of osteoporosis drug treatment continuation versus discontinuation have been</td>
<td>Larger trials are needed that compare osteoporosis drug treatment continuation versus discontinuation and are powered to look at 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>osteoporosis* or osteopenia/low bone mass† to prevent fracture, does the</td>
<td>inadequately powered to evaluate for possible effect modification of treatment effects on risk of</td>
<td></td>
</tr>
<tr>
<td>effect of osteoporosis drug treatment holidays on incident fracture risk vary</td>
<td>incident clinical fracture outcomes.</td>
<td></td>
</tr>
<tr>
<td>as a function of patient, bone, or osteoporosis drug characteristics?</td>
<td>No published trials have directly compared different durations of osteoporosis drug treatment</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></td>
<td>holidays with or without restarting osteoporosis drug treatment and reported on differences in</td>
<td></td>
</tr>
<tr>
<td></td>
<td>risk of harms between these groups.</td>
<td></td>
</tr>
<tr>
<td>KQ7: Among men and postmenopausal women aged &gt;50 years currently receiving</td>
<td>No published trials have directly compared different durations of osteoporosis drug treatment</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. Studies even could alternate multiple cycles of active treatment and discontinuation. These trials should systematically collect, analyze and report risk of harms between treatment groups, including both common and rare harms. The trials also should include an arm of continuous osteoporosis drug treatment (i.e., drug holiday duration of zero).</td>
</tr>
<tr>
<td>osteoporosis* or osteopenia/low bone mass† to prevent fracture, what is the</td>
<td>holidays with or without restarting osteoporosis drug treatment and reported on differences in</td>
<td></td>
</tr>
<tr>
<td>risk of harms of osteoporosis drug treatment holidays?</td>
<td>risk of harms between these groups.</td>
<td></td>
</tr>
<tr>
<td>KQ8: Among men and postmenopausal women aged &gt;50 years currently receiving</td>
<td>We found limited evidence from eligible studies about whether patient, bone or osteoporosis</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>osteoporosis* or osteopenia/low bone mass† to prevent fracture, does the</td>
<td>drug characteristics modify risk for harms associated with osteoporosis drug continuation versus</td>
<td></td>
</tr>
<tr>
<td>effect of osteoporosis drug treatment holidays vary as a function of patient,</td>
<td>discontinuation.</td>
<td></td>
</tr>
<tr>
<td>bone, or osteoporosis drug characteristics?</td>
<td>Ongoing or completed trials of osteoporosis drug continuation versus discontinuation that are in</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>
<tr>
<td></td>
<td>the process of collecting or already have collected harms did not plan a priori to evaluate</td>
<td></td>
</tr>
<tr>
<td></td>
<td>possible effect modification of treatment harms.</td>
<td></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.

*Osteoporosis defined by hip or lumbar spine dual-energy x-ray absorptiometry (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
Conclusions

Only alendronate, zoledronate, and oral hormone therapy reduced nonvertebral fractures with long-term treatment. However, for all these agents, these fracture benefits were limited to mostly older, postmenopausal women. They were further limited to women with osteopenia or osteoporosis for zoledronate, and to women with osteoporosis for alendronate. Absolute reductions in clinical fractures with long-term bisphosphonates appeared far greater than absolute increases in risk of AFF and ONJ with these treatments. However, reductions in hip fracture with long-term oral hormone therapy appear offset by risk of serious harms. In patients with prior osteoporosis drug treatment, continued treatment appeared to reduce vertebral fractures but not nonvertebral fractures, and may increase risk of AFF. While fracture benefits of continued osteoporosis drug treatment versus drug holiday numerically appeared to outweigh these risks, the more limited morbidity prevented and greater uncertainty about the outcome measures and risk estimates require further investigation to better inform clinical decisions about continuing treatment. This research should include examination of how these benefits and risks vary as a function of patient, bone, and drug treatment characteristics (e.g., age, sex, comorbidity, pre-drug holiday BMD, duration of prior osteoporosis drug treatment). Further, incorporating fracture morbidity in a decision model possibly would help patients weigh their trade-offs more easily.
References


105. 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]. American Society for Bone and Mineral Research Annual Meeting, September 28-October 1, Montréal, Québec, Canada. 2018; Abstract #1070.


Abbreviations

AE  Adverse event
AFF  Atypical femoral fracture
AFib  Atrial fibrillation
aHR  Adjusted hazard ratio
AHRQ  Agency for Healthcare Research and Quality
ALP  Alkaline phosphatase
aOR  Adjusted odds ratio
ASBMR  American Society for Bone and Mineral Research
BSAP  Bone specific alkaline phosphatase
BMD  Bone mineral density
BMI  Body mass index (kg/m²)
CCT  Controlled clinical trial
CENTRAL  Cochrane Central Register of Controlled Trials
CER  Comparative Effectiveness Review
CFx  Clinical fracture
CHF  Congestive heart failure
CI  Confidence interval
CKD  Chronic kidney disease
CORE  Continuing Outcomes Relevant to Evista
COX  Cyclooxygenase
CTX  C-terminal telopeptide of type 1 collagen
CVD  Cardiovascular disease
CVFx  Clinical vertebral fracture
DM  Diabetes mellitus
DVT  Deep vein thrombosis
DXA  Dual-energy x-ray absorptiometry
eGFR  Estimated glomerular filtration rate
EHR  Electronic health record
EPC  Evidence-Based Practice Center
FDA  U.S. Food and Drug Administration
FIT  Fracture Intervention Trial
FLEX  Fracture Intervention Trial Long-Term Extension
FN  Femoral neck
FN-BMD  Femoral neck bone mineral density
FRAX  World Health Organization Fracture Risk Assessment Tool
Fx  Fracture
GRADE  Grading of Recommendations Assessment, Development, and Evaluation
GI  Gastrointestinal
g/cm²  Grams per centimeter squared
HORIZON  Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly Trial
HR  Hazard ratio
ICD  International Classification of Diseases
ITT  Intention to treat
IV  Intravenous
KQ  Key Question
<table>
<thead>
<tr>
<th>WHO</th>
<th>World Health Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>yr</td>
<td>Year</td>
</tr>
</tbody>
</table>
Appendix A. Search Strategy

Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) Search Strategy

1 exp Osteoporosis/ (56601)
2 osteoporosis.ti. (24986)
3 Bone Density/ (53169)
4 exp Fractures, Bone/ (178905)
5 or/1-4 (249251)
6 Bone Density Conservation Agents/ (13618)
7 exp Diphosphonates/ (26164)
8 bisphosphonate*.ti. (6667)
9 alendronate.ti. (2190)
10 ibandronate.ti. (497)
11 risedronate.ti. (732)
12 zoledronic acid.ti. (1876)
13 or/7-12 (27852)
14 denosumab.ti. (998)
15 exp Anabolic Agents/ (15267)
16 teriparatide.ti. (835)
17 abaloparatide.ti. (22)
18 or/15-17 (16105)
19 exp Selective Estrogen Receptor Modulators/ (28675)
20 raloxifene.ti. (1433)
21 or/19-20 (28784)
22 Hormone Replacement Therapy/ (9611)
23 Estrogen Replacement Therapy/ (15549)
24 Estrogens, Conjugated/ (3783)
25 (conjugated adj2 estrogens).ti. (442)
26 (conjugated adj2 oestrogens).ti. (52)
27 bazedoxifene.ti. (203)
28 parathyroid.ti. (21206)
29 pth.ti. (2443)
30 24 or 25 or 26 or 27 or 28 or 29 (27047)
31 Romosozumab.ti. (37)
32 6 or 13 or 14 or 18 or 21 or 22 or 23 or 30 or 31 (124263)
33 5 and 32 (19473)
34 meta analysis as topic/ (17374)
35 meta-analy$.tw. (132466)
36 metaanaly$.tw. (1964)
A-2

meta-analysis/ (94826)
(systematic adj (review$1 or overview$1)).tw. (121503)
exp Review Literature as Topic/ (10321)
or/34-39 (237928)
cochrane.ab. (61695)
embase.ab. (65904)
(psychlit or psyclit).ab. (957)
(psychinfor or psycinfo).ab. (18590)
or/41-44 (100777)
reference list$.ab. (15891)
bibliograph$.ab. (16312)
hand search.ab. (1444)
relevant journals.ab. (1105)
manual search$.ab. (3843)
or/46-50 (36058)
selection criteria.ab. (28294)
data extraction.ab. (16712)
52 or 53 (42825)
review/ (2480829)
54 54 and 55 (28679)
comment/ (735829)
letter/ (1035382)
editorial/ (470175)
animal/ (6598636)
human/ (18057963)
or/57-59,62 (6336940)
or/40 or 45 or 51 or 56 (281525)
or/64 not 63 (266871)
55 randomized controlled trials as topic/ (123612)
randomized controlled trial/ (505454)
random allocation/ (101086)
double blind method/ (159463)
single blind method/ (27137)
clinical trial/ (553719)
clinical trial, phase i.pt. (20676)
clinical trial, phase ii.pt. (33325)
clinical trial, phase iii.pt. (15646)
clinical trial, phase iv.pt. (1673)
controlled clinical trial.pt. (100423)
randomized controlled trial.pt. (505454)
multicenter study.pt. (254739)
clinical trial.pt. (553719)
exp Clinical trials as topic/ (335805)
or/66-80 (1338507)
(clinical adj trial$).tw. (327584)
((singl$ or doubl$ or treb$ or tripl$) adj (blind$3 or mask$3)).tw. (169807)
placebos/.tw. (211894)
randomly allocated.tw. (25617)
(allocated adj2 random$).tw. (28819)
82 or 83 or 84 or 85 or 86 or 87 (592975)
81 or 88 (1569425)
case report.tw. (279795)
case report.tw. (279795)
letter/.tw. (1035382)
historical article/.tw. (358172)
90 or 91 or 92 or 93 (1658574)
89 not 94 (1533984)
exp cohort studies/.tw. (1866525)
cohort$.tw. (483043)
controlled clinical trial.pt. (100423)
edemiologic methods/.tw. (32559)
limit 99 to yr=1971-1983 (5524)
96 or 97 or 98 or 100 (2167906)
exp case-control study/.tw. (974951)
(case$ and control$).tw. (459124)
102 or 103 (1309725)
65 or 95 or 101 or 104 (3860160)
33 and 65 (723)
33 and 95 (6177)
33 and 101 (3725)
33 and 104 (1519)
106 or 107 or 108 or 109 (8692)
limit 110 to animals (636)
limit 111 to humans (415)
113 not 111 (8056)
114 or 112 (8471)
limit 114 to "all child (0 to 18 years)" (541)
limit 115 to "all adult (19 plus years)" (276)
117 not 115 (7930)
118 or 116 (8206)
limit 118 to (addresses or autobiography or bibliography or biography or case reports or comment or dataset or dictionary or directory or editorial or interactive tutorial or interview or legal cases or legislation or letter or news or newspaper article or patient education handout or periodical index or personal narratives or portraits or validation studies or video-audio media or webcasts) (389)
120 not 119 (7817)
limit 120 to ("young adult (19 to 24 years)" or "adult (19 to 44 years)") (1684)
limit 121 to ("middle age (45 to 64 years)" or "middle aged (45 plus years)" or "all aged (65 and over)" or "aged (80 and over)") (1411)
123 not 121 (6133)
124 123 or 122 (7544)  
125 limit 124 to english language (6901)  
126 limit 125 to yr="1995-Current" (6527)  

***************************  
Database: Embase Classic+Embase Search Strategy  
--------------------------------------------------------------------------------

1 exp *Osteoporosis/ (58317)  
2 osteoporosis.ti. (33878)  
3 *Bone Density/ (22135)  
4 exp *Fractures, Bone/ (160114)  
5 or/1-4 (227622)  
6 *Bone Density Conservation Agents/ (1590)  
7 exp *bisphosphonic acid derivative/ (23518)  
8 bisphosphonate*.ti. (8042)  
9 alendronate.ti. (2785)  
10 ibandronate.ti. (705)  
11 risedronate.ti. (921)  
12 zoledronic acid.ti. (2886)  
13 or/7-12 (24459)  
14 denosumab.ti. (1831)  
15 teriparatide.ti. (1212)  
16 abaloparatide.ti. (46)  
17 14 or 15 or 16 (3048)  
18 exp *Selective Estrogen Receptor Modulators/ (1852)  
19 raloxifene.ti. (1741)  
20 or/18-19 (3449)  
21 *hormone substitution/ (9570)  
22 *estrogen therapy/ (12439)  
23 exp *conjugated estrogen/ (4654)  
24 (conjugated adj2 estrogens).ti. (586)  
25 (conjugated adj2 oestrogens).ti. (68)  
26 bazedoxifene.ti. (327)  
27 parathyroid.ti. (24501)  
28 pth.ti. (3560)  
29 23 or 24 or 25 or 26 or 27 or 28 (32182)  
30 Romosozumab.ti. (63)  
31 6 or 13 or 14 or 17 or 20 or 21 or 22 or 29 or 30 (82327)  
32 5 and 31 (12350)  
33 Clinical trial/ (967720)  
34 Randomized controlled trial/ (482705)  
35 Randomization/ (76524)  
36 Single blind procedure/ (30258)  
37 Double blind procedure/ (147602)  
38 Crossover procedure/ (54478)
Placebo/ (321912)
Randomized controlled trial$.tw. (171939)
Rct.tw. (26582)
Random allocation.tw. (1803)
Randomly allocated.tw. (29202)
Allocated randomly.tw. (2326)
(allocated adj2 random).tw. (951)
33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 (1492328)
Case study/ (60326)
Case report.tw. (370340)
Abstract report/ or letter/ (1042296)
47 or 48 or 49 (1464628)
46 not 50 (1452601)
Clinical study/ (169001)
exp case control study/ (139905)
family study/ (27182)
longitudinal study/ (108891)
retrospective study/ (601258)
prospective study/ (419321)
cohort analysis/ (336075)
(cohort adj stud*).mp. (217474)
(observational adj stud*).mp. (172157)
(case control adj stud*).mp. (177221)
(follow up adj stud*).mp. (64043)
(epidemiologic* adj stud*).mp. (100966)
(cross sectional adj stud*).mp. (282797)
or/52-64 (2122395)
51 or 65 (3372681)
32 and 66 (5210)
limit 67 to (embryo <first trimester> or infant <to one year> or child <unspecified age> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>) (112)
limit 68 to (adult <18 to 64 years> or aged <65+ years>) (55)
67 not 68 (5098)
70 or 69 (5153)
limit 71 to (amphibia or ape or bird or cat or cattle or chicken or dog or "ducks and geese" or fish or "frogs and toads" or goat or guinea pig or "hamsters and gerbils" or horse or monkey or mouse or "pigeons and doves" or "rabbits and hares" or rat or reptile or sheep or swine) (104)
73 not 72 (5049)
limit 73 to (abstract report or books or "book review" or chapter or conference abstract or "conference review" or editorial or letter or note or patent or short survey or tombstone) (1313)
74 not 73 (3736)
limit 75 to yr="1995 -Current" (3564)
limit 76 to English language (3174)
Appendix B. Risk of Bias Assessment Decision Aid

Selection Bias

Systematic differences between baseline characteristics of the groups that arise from self-selection of treatments, physician-directed selection of treatments, or association of treatment assignments with demographic, clinical, or social characteristics.

Good randomization produces study groups that are likely comparable for known and unknown risk factors, removes investigator bias in allocation, and allow the most valid statistical inference in comparing outcomes between groups. In randomized studies, whether there is bias in allocation of study participants to treatment groups is a function both of whether the methods of randomization are good AND whether randomization successfully achieved a balance between treatment groups in risk factors or prognostic covariates.

- **Randomized: The study reports that it was randomized**
  - **Good Methodology of Randomization:** The study used a randomization method such as random numbers table, computer-generated random number producing algorithm, blocked randomization, stratified randomization, adaptive randomization (e.g., minimization).
  - **Unclear Methodology of Randomization:** Study reports that allocation/assignment was randomized, but gives no further detail.

- **Not Randomized:**
  - **Systematic allocation of treatment by investigator:** Systematic and predictable investigator allocation of treatment assignment (e.g., alternation, based on day of week, based on the month of birthday)
  - **Observational study:** Treatment allocation not assigned by investigator (i.e. noninterventional study)

- **Balance in prognostic variables.** A study is assumed to have achieved a balance between known/measured risk factors or prognostic variables when participant characteristics are balanced between treatment groups. Balance between unmeasured risk factors or prognostic variables cannot be determined.
Figure B1. Selection bias decision aid

- **Good Methodology**
  - Unclear/no imbalance in prognostic variables: Low
  - Significant imbalance in prognostic variables: Medium
  - No imbalance in prognostic variables: Medium

- **Unclear Methodology**
  - Unclear imbalance in prognostic variables: Unclear
  - Significant imbalance in prognostic variables: High

- **Not Randomized**
  - Good adjustment (propensity score, instrumental variable, multivariate) without evidence that any significant imbalance in prognostic variables was unaddressed: Medium
  - Inadequate/no adjustment: High
Performance Bias

Systematic differences in the care provided to participants and protocol deviation. Examples include contamination of the control group with the exposure or intervention, unbalanced provision of additional interventions or co-interventions, difference in co-interventions, and inadequate blinding of providers and participants.

- **Intention-to-Treat Principle (ITT)** is when the study counts events in all randomized participants according to their treatment assignment, regardless of whether they received assigned treatment. It does not exclude participants from analysis for nonadherence, protocol deviations, withdrawal, or anything else that happens after randomization. To exclude such participants undercuts the benefit of randomization in minimizing selection bias.

- **Modified ITT (mITT)** is where analyses exclude randomized participants who did not receive any of their assigned treatment. This is not strictly ITT, but is accepted as such by the FDA in evaluating drug trials for approval.

- **Adjustment for known potential confounders.**
  - **Adequate adjustment** includes adjustment for at least the following three categories: [1] age, [2] fracture risk (e.g., history of past fracture, BMD, FRAX or other risk calculator) and [3] some estimate of comorbidity.
  - **Partially adequate adjustment** adjusts for 1 or 2 of these potential confounder categories.
  - **Inadequate adjustment** does not adjust for any of these potential confounder categories.

Table B1. Performance bias decision aid

<table>
<thead>
<tr>
<th>Domain</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. ITT/modified ITT-applicable only for RCTs</td>
<td>Yes</td>
</tr>
<tr>
<td>1b. Adjustment for known confounders-applicable for CCTs, observational studies, and RCTs with imbalance in baseline prognostic variables)</td>
<td>Adequate</td>
</tr>
<tr>
<td>2. Participant Blinding-applicable for RCTs and CCTs</td>
<td>Yes</td>
</tr>
<tr>
<td>Overall Performance Rating</td>
<td>Low = All best categories</td>
</tr>
</tbody>
</table>
Reporting Bias

Systematic differences between reported and unreported findings (e.g., differential reporting of outcomes or harms, incomplete reporting of study findings).

- Were all outcomes reported in the methods section reported in the result section and vice versa?

Table B2. Reporting bias decision aid

<table>
<thead>
<tr>
<th>Domain</th>
<th>Options</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>All outcomes reported</td>
<td>Yes</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Medium</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

Notes to Reviewers for Using Decision Aid: Reviewers should first rate reporting bias based on the information in the article. In the ideal world, reviewers will have the protocol available for all the studies and will go back to look for whether all major outcomes (those for which we are assessing ROB and SOE) reported in the protocol methods were reported in the paper results. This may allow us to find reporting bias that wasn’t apparent from the paper, which could result in a downgrading of the overall ROB rating.

To save work, we may decide that if a study already has an overall high ROB without checking the protocol, there is no good reason to check the protocol. If the study has an overall low or medium ROB, we can check the protocol for evidence of reporting bias.
Attrition

Loss of participants from the study, potential systematic differences in that loss to follow-up, and how losses were accounted for in the results (e.g., incomplete follow-up, differential attrition). Those who drop out of the study or who are lost to follow-up may be systematically different from those who remain in the study. Attrition bias can potentially change the collective (group) characteristics of the relevant groups and their observed outcomes in ways that affect study results by confounding and spurious associations.

- Are reasons for incomplete/missing data adequately explained?
- Do the authors attempt to address attrition in the analysis?
- Analysis should be done with appropriate method (i.e. sensitivity analysis with various scenarios), noting that this may help explain the size and direction of the potential bias, but they don’t eliminate the bias.
- **Overall attrition** refers to attrition in all treatment groups combined for a given outcome comparison and timepoint.
- **Differential attrition** refers to the absolute difference between treatment groups in attrition for a given outcome comparison and timepoint.

**Figure B2. Attrition bias decision aid**

![Attrition Bias Decision Aid Diagram](image-url)
Detection Bias

Systematic differences in outcomes assessment among groups being compared, including systematic misclassification of the exposure or intervention, covariates, or outcomes because of variable definitions and timings, diagnostic thresholds, recall from memory, inadequate assessor blinding, and faulty measurement techniques. Erroneous statistical analysis might also affect the validity of effect estimates.

- Were the outcome assessors blinded to the treatment assignment?
- Were outcome measures validated, reliable, and were the groups assessed using comparable outcome measures? Were outcomes confirmed/adjudicated, either by direct measurement in clinic, review of study records (e.g. study x-rays) or review of clinical records?

Table B3. Detection bias decision aid

<table>
<thead>
<tr>
<th>Domain</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome assessor blinded to treatment assignment</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
</tr>
<tr>
<td>High quality Instrument/measurement or outcome confirmed/adjudicated</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No (unconfirmed self-report outcomes)</td>
</tr>
<tr>
<td></td>
<td>Intermediate (outcomes based on administrative data)</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Overall Performance Rating</th>
<th>Low = 2 Yes OR 1 Yes, 1 Unclear</th>
<th>Medium = all unclear or intermediate</th>
<th>High = at least 1 No</th>
</tr>
</thead>
</table>
Overall Risk of Bias Guide

ASSESSMENT GUIDANCE: Overall risk of bias is determined by reviewer, or team, consensus. Figure B3 provides a guide for how to rate overall risk of bias, based on the assessment of each individual domain. Reviewers should use this guide when making judgements about overall risk of bias. However, there may be cases where deviation from this guide is necessary and appropriate. For clarification and transparency, reviewers should provide a brief written justification for these deviations.
Figure B3. Overall risk of bias decision aid
## Appendix C. Included References Risk of Bias Assessment

### Table C1. Included studies with summary risk of bias assessments

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Study</th>
<th>Selection Bias</th>
<th>Attrition</th>
<th>Performance Bias</th>
<th>Detection Bias</th>
<th>Reporting Bias</th>
<th>Overall Rating</th>
<th>Justification (if deviated from decision aid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphosphonate Alendronate</td>
<td>Abrahamsen 2009</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Abrahamsen 2010</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Abrahamsen 2016</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Bauer 2000</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Pooled data from long-term RCT and extension(s) with differing attrition</td>
</tr>
<tr>
<td></td>
<td>Bauer 2006</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Pooled data from long-term RCT and extension(s) with differing attrition</td>
</tr>
<tr>
<td></td>
<td>Black 2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Black 2010</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Bone 2004</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low overall attrition, high differential attrition</td>
</tr>
<tr>
<td></td>
<td>Chapurlat 2005</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Chiu 2014</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Chiu 2018</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Cummings 1998</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Donaldson 2012</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Ensrud 2004</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Jamal 2007</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Lin 2014</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study</td>
<td>Selection Bias</td>
<td>Attrition</td>
<td>Performance Bias</td>
<td>Detection Bias</td>
<td>Reporting Bias</td>
<td>Overall Rating</td>
<td>Justification (if deviated from decision aid)</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Nevitt 1999</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Quandt 2005</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Ryder 2008</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Schwartz 2010</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Tonino 2000</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>High attrition overall, unable to determine attrition for extension study only</td>
</tr>
<tr>
<td></td>
<td>Wang, 2016</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Black 2012</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Black 2015</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Cosman 2014</td>
<td>Medium</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>McClung 2009</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Reid 2018</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Sorensen 2003</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Ste 2004</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Miller 2012</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Partial randomization not described clearly, missing data not clearly defined</td>
</tr>
<tr>
<td>Bisphosphonate</td>
<td>Adams 2018</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Erviti 2013</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Kim 2011</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study</td>
<td>Selection Bias</td>
<td>Attrition</td>
<td>Performance Bias</td>
<td>Detection Bias</td>
<td>Reporting Bias</td>
<td>Overall Rating</td>
<td>Justification (if deviated from decision aid)</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------</td>
<td>----------------</td>
<td>-----------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------------------------------------------------------------</td>
</tr>
<tr>
<td>Denosumab</td>
<td>Koh 2017</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Lee 2018</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Lim 2018</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Meier 2012</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Park-Wylie 2011</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Schilcher 2014</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td>Denosumab</td>
<td>Leder 2015</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Miller 2008</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td>Raloxifene</td>
<td>Barrett 2002</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition well explained and addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Barrett 2004</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition well explained and addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Barrett 2009</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Cauley 2001</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Delmas 2002</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition well explained and addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Duvernoy 2005</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Ensrud 2006</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Grady 2004</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study</td>
<td>Selection Bias</td>
<td>Attrition</td>
<td>Performance Bias</td>
<td>Detection Bias</td>
<td>Reporting Bias</td>
<td>Overall Rating</td>
<td>Justification (if deviated from decision aid)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------</td>
<td>----------------</td>
<td>-----------</td>
<td>------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Grady 2010</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Data from long-term RCT and extension(s) with attrition addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Johnell 2004</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td></td>
<td>Martino 2004</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Martino 2005</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Siris 2005</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Sontag 2010</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Data from long-term RCT with attrition addressed in analysis</td>
</tr>
<tr>
<td>Hormone therapy</td>
<td>Wimalawansa (Estrogen/Progestin) 1998</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Cauley (Estrogen/Progestin) 2003</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Jackson (Estrogen) 2006</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>NA</td>
</tr>
<tr>
<td>Osteoporosis drugs</td>
<td>Vestergaard 2010</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Vestergaard 2011</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Vestergaard 2012</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>NA</td>
</tr>
</tbody>
</table>
# Appendix D. Evidence Tables

## Table D1. Characteristics of observational studies with low or medium risk of bias

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Intervention</th>
<th>Overall sample</th>
<th>Comparison sample</th>
<th>Followup Duration</th>
<th>Outcomes Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrahamsen, 2009 Database cohort</td>
<td>Alendronate</td>
<td>N=15561</td>
<td>Alendronate n=5187</td>
<td>0-9 yr (longterm outcomes reported for 6-9 years)</td>
<td>Fracture: NR Predictors: NR Harms: AFF Rare harms: AFF</td>
</tr>
<tr>
<td>Location: Denmark, Funder: Foundation, Risk of Bias: Medium</td>
<td>Dose NR, Delivery route NR, Duration &gt; 6 yr</td>
<td>Born in 1945 or earlier, sustained hospital treated Fx except hip Fx between 1997 and 2005 Age 73 Female 90% Comorbidities NR</td>
<td>Age 73 Female 90% Comorbidities NR Baseline BMD NR Baseline fracture Spine 8% Humerus 12% Forearm 35% Other 45% Fx risk NR</td>
<td>No treatment n=10374</td>
<td>Fracture:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adams, 2018 Database cohort</td>
<td>Bisphosphonates (mostly alendronate (99%), risedronate, ibandronate, etidronate, pamidronate, zoledronic acid) ≥3 yr BP drug holiday (no use ≥1 yr)</td>
<td>N=39502</td>
<td>BP use ≥3 yr n=17123 Age 71 Female 100% Quan-Charlson Comorbidity Index 1.3 FN T-score -2.3 FRAX (any osteoporotic fx) 14.9 Prior fx 12.9% Vetebral T-score: -1.8 T-score ≤ -2.5: 48%</td>
<td>BP drug holiday ≥1 yr n=11497 Age 69 Female 100% Quan-Charlson Comorbidity Index 1.1 FN T-score -1.6 FRAX (any osteoporotic fx) 13.4 Prior fx 11.2% Vetebral T-score: -1.6 T-score ≤ -2.5: 37%</td>
<td>Mean followup 4 yr (4.9 yr for drug holiday group; 3.5 for comparison group)</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Location</td>
<td>Funder</td>
<td>Risk of Bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>--------</td>
<td>--------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiu, 2014</td>
<td>Taiwan</td>
<td>Government</td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrospective cohort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Taiwan</td>
<td>Government</td>
<td>Medium</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Intervention

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Intervention</th>
<th>Comparison(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Alendronate</td>
<td></td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Alendronate</td>
<td>Nonbisphosphonate osteoporosis medications (raloxifene or calcitonin)</td>
</tr>
</tbody>
</table>

### Dosage

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Dosage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Dose NR</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Dose NR</td>
</tr>
</tbody>
</table>

### Delivery route

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Delivery route</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Delivery route NR</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Delivery route NR</td>
</tr>
</tbody>
</table>

### Duration

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Mean duration 4 yr (0-11 yr)</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Duration 6 yr (range NR)</td>
</tr>
</tbody>
</table>

### Overall sample

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>N=</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>7332 alendronate users</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>32006</td>
</tr>
</tbody>
</table>

### Inclusion criteria

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Women ≥50 and men ≥60 yr old prescribed alendronate via hospital pharmacy. Excluded adults with head or neck cancer before/during treatment.</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Women and men aged ≥50 with newly diagnosed hip or VFx, new to alendronate within 30 days ofFx compared with propensity-matched raloxifene or calcitonin users.</td>
</tr>
</tbody>
</table>

### Intervention sample

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Age 72 Female 88% DM 16% CKD 2% Hypertension 44%</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Age 74 Female 84% DM 24% Renal disease 3%</td>
</tr>
</tbody>
</table>

### Comparison sample

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Mean age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Did not have ONJ n=7292</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>Raloxifene or calcitonin n=16003</td>
</tr>
</tbody>
</table>

### Follow up Duration

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>Mean drug duration 4 yr (0-11 yr)</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>6 yr</td>
</tr>
</tbody>
</table>

### Outcomes Reported

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Fracture:</th>
<th>Predictors:</th>
<th>Harms:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>NR</td>
<td>NR</td>
<td>ONJ</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>NR</td>
<td>NR</td>
<td>ONJ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Rare harms:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chiu, 2014</td>
<td>ONJ</td>
</tr>
<tr>
<td>Lin, 2014</td>
<td>ONJ</td>
</tr>
</tbody>
</table>

D-2
<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Location</th>
<th>Funder</th>
<th>Risk of Bias</th>
<th>Intervention</th>
<th>Comparison(s)</th>
<th>Overall sample</th>
<th>Intervention sample</th>
<th>Comparison sample</th>
<th>Followup Duration</th>
<th>Outcomes Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erviti, 2013</td>
<td>Case control</td>
<td>Spain</td>
<td>Spanish Ministry of Health</td>
<td>Medium</td>
<td>Bisphosphonates</td>
<td>Dose NR</td>
<td>Delivery route NR &gt; 3 yr</td>
<td>N=264</td>
<td>Women aged &gt;65</td>
<td>&gt;3 yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age 82</td>
<td>Female 100%</td>
<td>Fracture: NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Comorbidities</td>
<td>DM 21%</td>
<td>Predictors: NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CKD 6%</td>
<td>Harms: NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CVD NR</td>
<td>Rare harms: AFF</td>
</tr>
<tr>
<td>Kim, 2011</td>
<td>Database cohort</td>
<td>US</td>
<td>No funding</td>
<td>Medium</td>
<td>Bisphosphonates (alendronate, risedronate, etidronate)</td>
<td>Raloxifene or calcitonin nasal spray</td>
<td>Dose NR Oral (except calcitonin) &gt;90 days</td>
<td>N=33815</td>
<td>Medicare beneficiaries with at least one osteoporosis treatment prescription and one medical claim in previous 3 consecutive 6 month periods; Jan 1996-Dec 2006</td>
<td>Up to 9.5 yr (est) (mean 2 yr)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age 80</td>
<td>Female 97%</td>
<td>Fracture: NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>White 95%</td>
<td>Predictors: none</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DM 26%</td>
<td>Harms: NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CKD 3%</td>
<td>Rare harms: subtrochanteric and diaphyseal femur fx (proxy for AFF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hypertension 67%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CHF 22%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BMD NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hip Fx 4%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VFx 11%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fx risk NR</td>
<td></td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Overall sample</td>
<td>Intervention sample</td>
<td>Comparison sample</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Koh, 2017, Case control</td>
<td>Bisphosphonates (alendronate, risedronate, ibandronate, or zoledronate)</td>
<td>N=172 out of 35,104 patients prescribed bisphosphonates ≥1 yr, Cancer history, hyperparathyroidism, renal osteodystrophy, and osteomalacia excluded. Cases experienced AFF. Controls received bisphosphonates without AFF matched 1:3 from EHR.</td>
<td>Cases: AFF n=43 Mean drug duration: 7.3 yr Age 68 (NR) Female 100% Race NR DM 19% BMD NR Previous osteoporotic Fx 19% Fx risk NR Had drug holiday 16% Glucocorticoid ≥ 1yr 37%</td>
<td>Controls: No AFF n=129 Mean drug duration: 5.2 yr Age 68 (NR) Female 100% Race NR DM 11% BMD NR Previous osteoporotic Fx 12% Fx risk NR Discontinued drug 36% Glucocorticoid ≥ 1yr 9%</td>
<td>Mean 5.7 yr (range NR)</td>
<td>Fracture: NR Predictors: NR Harms: NR Rare harms: AFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lim, 2018, Case control</td>
<td>Bisphosphonates (alendronate, risedronate, ibandronate, pamidronate or zoledronate)</td>
<td>N= 290 from a cohort of 6644 patients with hip and femoral fractures identified via medical record diagnosis codes Age 72 Female 100% Comorbidities DM 16% CKD NR CVD 13% T-score &lt;-2.0 85%</td>
<td>Cases: AFF n=196 Age 72 Female 100% Comorbidities DM 16% CKD NR CVD 13% T-score &lt;-2.0 85%</td>
<td>Controls: Typical femoral fracture n=94 Age 71 Female 100% Comorbidities DM 28% CKD NR CVD 5% T-score &lt;-2.0 18%</td>
<td>Mean 5.2 yr (range 1-17 yr)</td>
<td>Fracture: NR Predictors: NR Harms: NR Rare harms: AFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Overall sample N=</td>
<td>Intervention sample n=</td>
<td>Comparison sample n=</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Comparison(s)</td>
<td>N=</td>
<td>Mean age</td>
<td>Mean age</td>
<td>Mean age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funder</td>
<td>Dosage</td>
<td>Inclusion criteria</td>
<td>% Female</td>
<td>% Female</td>
<td>% Female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of Bias</td>
<td>Delivery route</td>
<td>Mean age</td>
<td>Comorbidities</td>
<td>Baseline BMD</td>
<td>Comorbidities</td>
<td>Baseline BMD</td>
<td>Fracture history</td>
<td>Fracture history</td>
<td>Fracture risk</td>
<td>Fracture risk</td>
</tr>
<tr>
<td></td>
<td>Duration</td>
<td>% with DM, CKD, CVD</td>
<td>Fracture history</td>
<td>Fracture risk</td>
<td>Fracture history</td>
<td>Fracture risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Park-Wyllie, 2011</td>
<td>Bisphosphonates (alendronate, risedronate or etidronate)</td>
<td>4296</td>
<td>Postmenopausal women aged ≥50. Excluded women with 10 yr history of cancer, conditions with altered bone integrity, gastric bypass, used other osteoporosis drugs, high energy trauma, andFx treated with hip replacement. Cases matched with up to 5 controls.</td>
<td>Cases: Bisphosphonates ≥5 yr n=716</td>
<td>Controls: Bisphosphonates 3-5 yr n=3580</td>
<td>4-5 yr</td>
<td>Fracture: NR Predictors: drug duration Harms: NR Rare harms: AFF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case control</td>
<td>Dose NR</td>
<td>Age 83 (median) Female 100% Race NR CHF 13% DM 12% BMD NR Osteoporotic Fx in prior 5 yr 70% Falls prior 5 yr: 67%</td>
<td>Age 83 (median) Female 100% Race NR CHF 13% DM 12% BMD NR Osteoporotic Fx in prior 5 yr 70% Falls prior 5 yr: 67%</td>
<td>Age 83 (median) Female 100% Race NR CHF 5% DM 6% BMD NR Osteoporotic Fx in prior 5 yr 24% Falls prior 5 yr: 6%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>&lt;100 days (reference) to &lt;5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schilcher, 2015</td>
<td>Bisphosphonates</td>
<td>2.9 million women and men aged ≥ 55 yr over the study period of 2008-2010, 97,900 with a history of bisphosphonate use (87% women)</td>
<td>Cases: AFF n=172</td>
<td>Controls: Ordinary shaft fx n=952</td>
<td>4-5 yr</td>
<td>Fracture: NR Predictors: age Harms: NR Rare harms: AFF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cohort/Case control</td>
<td>Dose NR</td>
<td>Age 76 Female 93% Comorbidities DM 4% CKD 30% CVD 46%</td>
<td>Age 82 Female 81% Comorbidities DM 16% CKD 40% CVD 60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>Delivery route NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schilcher Research Council</td>
<td>4-5 yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Overall sample</td>
<td>Intervention sample</td>
<td>Comparison sample</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>----------------</td>
<td>--------------------</td>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location Funder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vestergaard Series</td>
<td>Osteoporosis drugs (bisphosphonates including alendronate, risedronate, zoledronate, ibandronate), raloxifene, teriparatide</td>
<td>N= varied</td>
<td>n= varied</td>
<td>n= varied</td>
<td>3.8 yr (mean, after date of exposure)</td>
<td>Fracture: NR Predictors: Cumulative dose Harms: NR Rare Harms: Inflammatory jaw events (proxy for ONJ) (Vestergaard 2012), Femoral shaft and subtrochanteric fractures (proxy for AFF) (Vestergaard 2011), AFib or atrial flutter (Vestergaard 2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vestergaard 2010, Vestergaard 2011, Vestergaard 2012 Database cohort</td>
<td>Dose varied Delivery route varied 3.8 yr observation (mean)</td>
<td>Danish people who filled a prescription for a drug used to treat or prevent osteoporosis Jan 1, 1996-Dec 31, 2006; matched (1:3) by age and gender to Danish people who had not filled such a prescription</td>
<td>Age 71 Female 85% White NR DM 2% CKD NR Hypertension 37% CHF 3%</td>
<td>Age 71 Female 85% White NR DM 2% CKD NR Hypertension 34% CHF 2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark Industry, foundation Med</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** Afib=atrial fibrillation; AFF=atypical femoral fracture; ALP=alkaline phosphatase; BMD=bone mineral density; BP=bisphosphonates; BSALP=bone specific alkaline phosphatase; CFx=clinical fracture; CHF=congestive heart failure; CKD=chronic kidney disease; CVD=cardiovascular disease; DM=diabetes mellitus; EHR=electronic health record; FIT=Fracture Intervention Trial; FN=femoral neck; FRAX=World Health Organization (WHO) Fracture Risk Assessment Tool; Fx=fracture; g/cm²=grams per centimeter squared; LS=lumbar spine; mg=milligrams; N=number; NR=not reported; NTX=N-terminal telopeptide of type 1 collagen; ONJ=osteonecrosis of the jaw; QD=per day; RCT=randomized controlled trial; TH=total hip; UGI=upper Gastrointestinal; VFx=vertebral fracture; vs=versus; yr=years
<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Intervention, Dosage, Delivery Route, Duration</th>
<th>Comparison, Dosage, Delivery Route, Duration</th>
<th>Sample, # Randomized, Mean age, % Female, % White Race, Comorbidities (% with DM, CKD, CVD)</th>
<th>Inclusion criteria, Baseline BMD</th>
<th>Followup Duration</th>
<th>Outcomes Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bauer, 2000</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral Fit-I: mean 2.9 yr Fit-II: mean 4.3 yr</td>
<td>Placebo QD Oral Fit-I: mean 2.9 yr Fit-II: mean 4.3 yr</td>
<td>N=6459 Age 69 Female 100% Race NR Comorbidities NR Prior Frx NR</td>
<td>Postmenopausal women with or without baseline VFx. FN BMD ≤ 0.68g/cm² (T score ≤ -1.6)</td>
<td>Mean 3.8 yr (median NR); pooled study arms with mean 2.9 yr and 4.3 yr followup</td>
<td>Fracture: NR Predictors: age, UGI disease, NSAID use. Harms: UGI only (any, gastric/duodenal, esophageal) Rare harms: NR</td>
</tr>
<tr>
<td>Bauer, 2006</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral Fit-I: 3 yr Fit-II: 4 yr Calcium and Vitamin D supplements</td>
<td>Placebo QD Oral Fit-I: 3 yr Fit-II: 4 yr Calcium and Vitamin D supplements</td>
<td>N=6186 (completed trial, complete data) Age 69 yr Female 100% Race NR Comorbidities NR Prior fracture: NR Baseline VFx: 31%</td>
<td>Postmenopausal women with or without baseline VFx and FN-BMD ≤ 0.68g/cm² (T score ≤ -1.6)</td>
<td>“Mean 3.2 yr” (median NR)</td>
<td>Fracture: clinical (pathologic or severe trauma excluded), radiographic vertebral Predictors: pretreatment bone turnover; osteoporosis Harms: NR Rare harms: NR</td>
</tr>
<tr>
<td>Ensrud, 2004</td>
<td>Alendronate 5 mg QD Oral 5 yr (5-10 yr total) Alendronate 10 mg Daily Oral 5 yr (5-10 yr total) 1.6% on raloxifene or HRT. Calcium 500mg and 250 IU Vitamin D offered.</td>
<td>Placebo (discontinue alendronate after 5 yr) QD Oral 5 yr 2.7% on raloxifene or HRT. Calcium 500mg and 250 IU Vitamin D offered.</td>
<td>N=1099 (39% of 2852 ALN users from FIT) with mean of 5 yr. prior ALN use</td>
<td>Prior enrollment in ALN arm of FIT (I or II), took ≥3 yr of ALN, with TH BMD &gt; -3.5 at FLEX baseline</td>
<td>5 yr (5-10 yr FIT extension)</td>
<td>Fractures: CVFx, R VFx, all nonVFx, forearm fx, hip fx; pathologic, skull, or excess trauma fxs excluded. Other outcomes: Change in TH BMD, new R VFx Predictors: FN BMD, prevalent VFx Harms: UGI, death Rare harms: ONJ</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Sample</td>
<td>Inclusion criteria</td>
<td>Followup</td>
<td>Outcomes Reported</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>------------</td>
<td>--------</td>
<td>-------------------</td>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Black, 2010 RCTs (pooled Fit-I and Fit-II, FLEX) USA Industry Low</td>
<td>Alendronate Pooled FIT-I and FIT-II: Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral FIT-I: mean 2.9 yr FIT-II: mean 4.3 yr Calcium and Vitamin D: NR FLEX: Alendronate 5 mg QD Oral 5 yr (5-10 yr total) Alendronate 10 mg Daily Oral 5 yr (5-10 yr total)</td>
<td>Pooled FIT-I and FIT-II: Placebo QD Oral FIT-I: mean 2.9 yr FIT-II: mean 4.3 yr Calcium and Vitamin D: NR</td>
<td>Pooled FIT-I and FIT-II: N=6459 For baseline characteristics, see Donaldson, 2012 FLEX: N=1099 For baseline characteristics, see Schwartz, 2010</td>
<td>Pooled FIT-I and FIT-II: see Donaldson, 2012 FLEX: see Schwartz, 2010</td>
<td>Pooled FIT-I and FIT-II: “Median 1457 days” (4 yr) [31.4% median 3 yr.; 68.6% median 4 yr.] FLEX: 5 yr (5-10 yr FIT extension)</td>
<td>Fracture: NR Predictors: NR Harms: NR Rare harms: AFF</td>
</tr>
<tr>
<td>Bone, 2004 Prospective CCT (2 pooled partial RCT extension with discontinuation; subset of Tonino, 2000) Multicenter Industry Medium</td>
<td>Alendronate Continue ALN 5 mg QD oral</td>
<td>placebo/discontinue QD Oral 500mg calcium, no vitamin D</td>
<td>N=247</td>
<td>See Tonino, 2000</td>
<td>8-10 yr</td>
<td>Fracture: RVFx, (6-10 yr) CFx, nonVFx (traumatic fx included) Other: change in LS BMD, FN BMD, trochanteric BMD, TH BMD, total body BMD, forearm BMD; NTX, BSALP, ALP. Predictors: NR Harms: death, UGI Rare harms: NR</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention Dosage</td>
<td>Comparison Dosage</td>
<td>Sample Size</td>
<td>Inclusion criteria Baseline BMD</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Cummings, 1998 FIT-II (CFx arm) RCT</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD x 2 yr Oral 4 yr Calcium and Vitamin D supplements</td>
<td>Placebo QD Oral 4 yr Calcium and Vitamin D supplements</td>
<td>N=4432</td>
<td>Postmenopausal women without baseline VFx and FN BMD ≤0.68g/cm² (T score ≤-1.6)</td>
<td>4 yr (mean 4.2 yr)</td>
<td>Fracture 1º: CFx (nonspine, hip, wrist, CVFx, other) 2º: RVFx Predictors: FN BMD Harms: any, death, GI Rare harms: NR</td>
</tr>
<tr>
<td>Donaldson, 2012 Pooled FIT-I and FIT-II RCTs</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral FIT-I: mean 2.9 yr FIT-II: mean 4.3 yr Calcium and Vitamin D: NR</td>
<td>Placebo QD Oral FIT-I: mean 2.9 yr FIT-II: mean 4.3 yr Calcium and Vitamin D: NR</td>
<td>N=6459 (FIT-I 2027, FIT-II 4432)</td>
<td>Postmenopausal 2+ yr., with or without baseline VFx. FN-BMD ≤ 0.68g/cm² (T score ≤ -1.6)</td>
<td>Median 1457 days (4 yr*) [31.4%* median 3 yr.; 68.6%* median 4 yr.]</td>
<td>Fracture: NonVFx; assessed CFx, MOF and RadVFx but results NR (interactions NS) Predictors: FRAX score (with/without FN-BMD). Baseline FRAX excluded baseline RadVFx); FN-BMD Harms: NR Rare harms: NR</td>
</tr>
<tr>
<td>Jamal, 2007 Pooled FIT-I and FIT-II RCTs</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>Placebo QD Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>N=6458</td>
<td>Postmenopausal women with or without baseline VFx and FN-BMD ≤ 0.68g/cm² (T score ≤ -1.6)</td>
<td>Average of 4 yrs (3 yr FIT-I, 4 yr FIT-II; pooled median NR)</td>
<td>Fracture: clinical, radiographic vertebral. Predictors: renal function (eGFR&lt;45 ml/min vs ≥45 ml/min) Harms: by renal function Rare harms: NR</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Sample</td>
<td>Inclusion criteria</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Quandt, 2005</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>Placebo QD Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>N=2037</td>
<td>Postmenopausal 2+ yr., with or without baseline VFx., FN-BMD ≤ 0.68g/cm² (T-score ≤ -1.6)</td>
<td>4 years</td>
<td>Fracture: NonVFx Predictors: FRAX score (with/without FN-BMD) Harms: NR Rare harms: NR</td>
</tr>
<tr>
<td>Ryder, 2008</td>
<td>Alendronate 5 mg QD x 2 yr then 10 mg QD thereafter Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>Placebo QD Oral FIT-I: 3 yr FIT-II: 4 yr Calcium and Vitamin D supplements</td>
<td>N=2785</td>
<td>Postmenopausal 2+ yr., with or without baseline VFx., FN-BMD ≤ 0.68g/cm² (T-score ≤ -1.6)</td>
<td>4 years</td>
<td>Fracture: NonVFx Predictors: FRAX score (with/without FN-BMD) Harms: NR Rare harms: NR</td>
</tr>
<tr>
<td>Schwartz, 2010</td>
<td>Alendronate 5 mg QD Oral 5 yr (5-10 yr total) Alendronate 10 mg Daily Oral 5 yr (5-10 yr total) (groups pooled for analyses)</td>
<td>Placebo (discontinued alendronate after 5 yr) QD Oral 5 yr</td>
<td>N=1099 (39% of 2852 ALN users from FIT) with mean of 5 yr prior ALN use Age 73 Female 100% White 98% Comorbidities NR Prior Fx ≥45 yr 60% Prevalent VFx 34%</td>
<td>Prior enrollment in ALN arm of FIT (I or II), took ≥3 yr ALN TH BMD &gt;-3.5 at FLEX baseline</td>
<td>5 yr (5-10 yr FIT extension)</td>
<td>Fractures: First NonVFx, CVFx, RVFx. (Pathologic, skull, or excess trauma fx not excluded) Predictors: baseline VFx, FN BMD change in BMD prior to FLEX Harms: NR Rare harms: NR</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Sample</td>
<td>Inclusion criteria</td>
<td>Followup</td>
<td>Outcomes Reported</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------</td>
<td>------------</td>
<td>--------</td>
<td>-------------------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>Tonino, 2000</strong></td>
<td>Alendronate</td>
<td>placebo/discontinue</td>
<td>N=350</td>
<td>Postmenopausal women with osteoporosis per LS BMD $\leq 0.80$g/cm$^2$ ($\leq -2.5$ SD)</td>
<td>6-7 yr</td>
<td>Fracture: symptomatic VFx, CFx (traumatic fx included) Other outcomes: change in LS BMD, change in FN, trochanteric, TH, total body, &amp; forearm BMD; NTX, BSALP Predictors: NR Harms: UGI, serious, any Rare harms: NR</td>
</tr>
<tr>
<td>Prospective CCT (2 pooled partial RCT extension with discontinuation)</td>
<td>Continue ALN 5 mg QD oral</td>
<td>Oral 500mg calcium, no vitamin D</td>
<td>Age NR Female 100% Comorbidities NR Fx history or risk NR</td>
<td>BMD NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multicenter Industry</td>
<td>Continue ALN 10 mg QD oral</td>
<td>Placebo/discontinue QD Oral</td>
<td>500mg calcium, no vitamin D</td>
<td>N=1233</td>
<td>Osteoporotic women BMD g/cm$^2$ FN BMD 0.57 TH BMD 0.69 LS BMD 0.82 FN T-score -2.57</td>
<td></td>
</tr>
<tr>
<td>Industry Medium</td>
<td>Placebo</td>
<td>with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>Age 76 Female 100% Race NR Comorbidities NR Fx history or risk NR</td>
<td>6 yr (3 yr extension of a 3 yr trial)</td>
<td>Fracture: CFx and RVFx Predictors: BMD (percentage change) Rare harms: ONJ, AFF, Afib</td>
<td></td>
</tr>
<tr>
<td><strong>Black, 2012</strong></td>
<td>Zoledronic acid 5mg once per year IV 6 yr with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>Placebo with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>N=190</td>
<td>Osteoporotic women BMD g/cm$^2$ FN BMD 0.58 TH BMD 0.70 FN T-score -2.44</td>
<td>9 yr (second 3 yr extension of a 3 yr trial)</td>
<td>Fracture: CFx and RVFx Predictors: TH BMD (percentage change) Rare harms: ONJ, AFF, Afib</td>
</tr>
<tr>
<td>HORIZON RCT Multinational Industry Low</td>
<td></td>
<td></td>
<td>Age 78 Female 100% Race NR Comorbidities NR Fx history or risk NR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Black, 2015</strong></td>
<td>Zoledronic acid 5mg once per year IV 9 yr with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>Placebo with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>N=190</td>
<td>Osteoporotic women BMD g/cm$^2$ FN BMD 0.58 TH BMD 0.70 FN T-score -2.44</td>
<td>9 yr (second 3 yr extension of a 3 yr trial)</td>
<td>Fracture: CFx and RVFx Predictors: TH BMD (percentage change) Rare harms: ONJ, AFF, Afib</td>
</tr>
<tr>
<td>HORIZON RCT Multinational Industry Low</td>
<td></td>
<td></td>
<td>Age 78 Female 100% Race NR Comorbidities NR Fx history or risk NR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Black, 2015</strong></td>
<td>Zoledronic acid 5mg once per year IV 9 yr with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>Placebo with calcium 1000-1500 mg QD and vitamin D 400-1200 IU QD</td>
<td>N=190</td>
<td>Osteoporotic women BMD g/cm$^2$ FN BMD 0.58 TH BMD 0.70 FN T-score -2.44</td>
<td>9 yr (second 3 yr extension of a 3 yr trial)</td>
<td>Fracture: CFx and RVFx Predictors: TH BMD (percentage change) Rare harms: ONJ, AFF, Afib</td>
</tr>
<tr>
<td>HORIZON RCT Multinational Industry Low</td>
<td></td>
<td></td>
<td>Age 78 Female 100% Race NR Comorbidities NR Fx history or risk NR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Intervention Dosage Delivery Route Duration</td>
<td>Comparison Dosage Delivery Route Duration</td>
<td>Sample # Randomized Mean age % Female % White Race Comorbidities (% with DM, CKD, CVD) Fracture History Fracture Risk</td>
<td>Inclusion criteria Baseline BMD</td>
<td>Followup Duration</td>
<td>Outcomes Reported</td>
</tr>
<tr>
<td>-------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>-----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>McClung, 2009 RCT</td>
<td>Zoledronic acid 5mg/yr for 1yr then placebo for 1yr or 5mg/yr for 2yr IV 2yr with calcium 500-1200 mg QD and vitamin D 400-800 IU QD</td>
<td>Placebo with calcium 500-1200 mg QD and vitamin D 400-800 IU QD</td>
<td>N=581 Postmenopausal women &gt;45 yr with T-score &lt;-1.0 and &gt;-2.5 at LS and T-score &gt;=-2.5 at FN BMD g/cm² FN BMD 0.69 TH BMD 0.83 LS BMD 0.86</td>
<td>Fracture: CFx Predictors: none Harms: Serious adverse events Rare harms: Afib, ONJ</td>
<td>2 yr</td>
<td></td>
</tr>
<tr>
<td>Reid, 2018 RCT</td>
<td>Zoledronic acid 5 mg at 18-month intervals IV 6 years With vitamin D 1.25 mg per month + advised calcium 1 g per day</td>
<td>Placebo (normal saline)</td>
<td>N=2000 Postmenopausal women ≥65 yr with T-score of -1.0 to -2.5 at either TH or FN on either side BMD g/cm² LS BMD 1.08 TH BMD 0.85 FN BMD 0.81 Total body DMB 1.06 T-score LS T-score -0.89 TH T-score -1.26 FN T-score -1.64 Total body T-score -0.81</td>
<td>Fracture: fragility Fx (primary outcome, defined as any NVF or X-ray confirmed VFx); CFx, VFx (secondary outcomes); nonvertebral fragility Fx, CVFx, Hip Fx, forearm or wrist Fx (exploratory outcomes) Predictors: baseline VFx status, baseline Fx risk, osteoporosis status Harms: death, sudden death, MI, stroke, TIA, composite vascular event measure Rare harms: ONJ, Afib</td>
<td>6 yr</td>
<td></td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Location</td>
<td>Funder</td>
<td>Risk of Bias</td>
<td>Intervention</td>
<td>Dosage</td>
<td>Delivery Route</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
<td>--------------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>Miller, 2008</td>
<td>US</td>
<td>Industry</td>
<td>Medium</td>
<td>Denosumab</td>
<td>various doses (6 mg or 14 mg every 3 months; or 14 mg, 60 mg, or 100 mg every 6 months) for 2 yr; then 60 mg every 6 months subcutaneously for 2 yr with calcium 1000 mg QD and vitamin D 400 IU QD</td>
<td>Placebo QD 4 yr with calcium 1000 mg QD and vitamin D 400 IU QD</td>
</tr>
<tr>
<td>Barrett-Connor, 2002</td>
<td>US Multinational</td>
<td>Industry (Eli Lilly)</td>
<td>Low (all reports from MORE trial have low RoB except Johnell 2004, which has medium)</td>
<td>Raloxifene 60 mg QD (1 placebo and one 60 mg tablet) 120 mg QD (two 60 mg tablets) Oral with calcium 500 mg QD and cholecalciferol 400 to 600 mg QD</td>
<td>Placebo 2 tablets QD Oral 4 yr with calcium 500 mg QD and cholecalciferol 400 to 600 mg QD</td>
<td>N=7705</td>
</tr>
<tr>
<td>Study, Year Design</td>
<td>Location</td>
<td>Funder</td>
<td>Risk of Bias</td>
<td>Intervention Dosage Delivery Route Duration</td>
<td>Comparison Dosage Delivery Route Duration</td>
<td>Sample # Randomized Mean age % Female % White Race Comorbidities (% with DM, CKD, CVD) Fracture History Fracture Risk</td>
</tr>
<tr>
<td>--------------------</td>
<td>---------</td>
<td>--------</td>
<td>-------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Grady 2010 MORE/CORE RCT</td>
<td>Multinational Industry (Eli Lilly)</td>
<td>Medium</td>
<td></td>
<td>Raloxifene 60 mg QD Oral 8 yr</td>
<td>Placebo QD Oral 8 yr</td>
<td>From beginning of MORE trial, raloxifene 60 mg and placebo groups: N=5133 Age 67 Female 100% White 96% Hypertension 30% CVD 2% Stroke 6% DM 3% Yr post menopause 19 Fx history or risk NR</td>
</tr>
</tbody>
</table>

Delmas 2002 Barrett-Connor 2002
<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Intervention Dosage Delivery Route Duration</th>
<th>Comparison Dosage Delivery Route Duration</th>
<th>Sample # Randomized Mean age % Female % White Race Comorbidities (% with DM, CKD, CVD) Fracture History Fracture Risk</th>
</tr>
</thead>
</table>
| Ensrud, 2006       | Raloxifene 60 mg QD Oral 8 yr               | Placebo QD Oral 8 yr                       | From beginning of MORE trial, placebo and raloxifene 60 mg groups:  
N=4011  
Age 66  
Female 100%  
White 96%  
Hypertension 28%  
CVD NR  
Stroke 0.3%  
DM 3%  
BMD NR  
History of VFx NR  
Yr post menopause 18  
Fx history or risk NR |
| MORE/CORE RCT      |                                             |                                            | Postmenopausal for ≥2 yr Osteoporosis (prior VFx or T-score <-2.5)  
Was enrolled in MORE |
| Funder             | Multinational Industry (Eli Lilly)          |                                            |                                                                 |
| Risk of Bias       | Medium                                      |                                            |                                                                 |
| Martino, 2004      |                                             |                                            |                                                                 |
| Martino, 2005      |                                             |                                            |                                                                 |
| Siris, 2005        |                                             |                                            |                                                                 |
| USA, 40 centers    |                                             |                                            |                                                                 |
| Government         | Low                                         |                                            |                                                                 |
| Cauley, 2003       | Conjugated equine estrogen 0.625 mg QD + medroxyprogesterone acetate 2.5 mg QD Oral 6 yr |
| WHI RCT            |                                             |                                            | N=16608  
N for BMD subset=1024  
Age 63  
Female 100%  
White 84%  
Comorbidities NR  
Prior Fx 39%  
High fracture risk 23%  
Postmenopausal women with intact uterus, 50-79 years at baseline  
No BMD requirement for inclusion  
TH BMD 0.83 (T score -0.93)  
LS BMD 0.94 (T score -1.28) |
| Followup Duration  | 8 yr                                        |                                            |                                                                 |
| Outcomes Reported  | Fracture: CFx (nonvertebral), fracture at 6 sites  
Predictors: age, BMD, prevalent vertebral Fx  
Harms: Stroke, PE, DVT, hot flashes, mortality, serious adverse events  
Rare harms: none |
<p>| Inclusion criteria  | Baseline BMD                                |                                            |                                                                 |
| Study, Year Design |                                             |                                            |                                                                 |
| Martino, 2004      |                                             |                                            |                                                                 |
| Martino, 2005      |                                             |                                            |                                                                 |
| WHI RCT            |                                             |                                            |                                                                 |
| Government         |                                             |                                            |                                                                 |
| Risk of Bias       | Low                                         |                                            |                                                                 |
| Cauley, 2003       |                                             |                                            |                                                                 |</p>
<table>
<thead>
<tr>
<th>Study, Year Design</th>
<th>Intervention Dosage Delivery Route Duration</th>
<th>Comparison Dosage Delivery Route Duration</th>
<th>Sample # Randomized Mean age % Female % White Race</th>
<th>Comorbidities (% with DM, CKD, CVD)</th>
<th>Fracture History Duration</th>
<th>Inclusion criteria Baseline BMD</th>
<th>Followup Duration</th>
<th>Outcomes Reported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jackson, 2006 WHI RCT USA, 40 centers Government Low</td>
<td>Conjugated equine estrogen 0.625 mg QD Oral 7 yr</td>
<td>Placebo QD Oral 7 yr</td>
<td>N=10739 N for BMD subset=938 Age 64 Female 100% White 75% Comorbidities NR Prior Fx 39% High fracture risk 28%</td>
<td>Postmenopausal women with prior hysterectomy, 50-79 years at baseline No BMD requirement for inclusion TH BMD T score -0.81 LS BMD T score -1.16</td>
<td>Mean 7.1 yr</td>
<td>Fractures: Cfx, hip Predictors: History of fracture, baseline BMD; fracture risk Harms: global index Rare harms: none</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wimalawansa 1998 RCT Unclear, possibly the UK</td>
<td>HRT, prempack C (premarin 0.625 mg daily and norgestrel 150mg 12 days each month) Oral with 1 g elemental calcium (Sandocal 1,000 effervescent) and 400 units of vitamin D (ergocalciferol 10 mg) QD Oral 4 yr</td>
<td>Control, 1 g elemental calcium (Sandocal 1,000 effervescent) and 400 units of vitamin D (ergocalciferol 10 mg) QD Oral 4 yr</td>
<td>N=36 Age 65 Female 100% White 100% Comorbidities NR Prior RVFx 100%</td>
<td>Evidence of osteoporosis as determined by at least 1 (but not more than 4) radiographically demonstrable atraumatic thoracic vertebral crush fractures, and spine BMD 2.0 standard deviations below the reference range using DXA. Mean LS BMD 0.82 g/cm² Mean FN BMD 0.67 g/cm².</td>
<td>4 yr</td>
<td>Fractures: NonVFx, RVFx Predictors: NR Harms: withdrawals due to AEs or other medical problems Rare harms: NR Other outcomes: Change in TH BMD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: Afib=atrial fibrillation; AFF=atypical femoral fracture; ALN=alendronate; BMD=bone mineral density; BSALP=bone specific alkaline phosphatase; CFx=clinical fracture; CHD=coronary heart disease; CKD=acute kidney disease; CORE=Continuing Outcomes Relevant to Evista; CVD=cardiovascular disease; DM=diabetes mellitus; DVT=deep vein thrombosis; FIT=Fracture Intervention Trial; FLEX=Fracture Intervention Trial Long Term Extension; FN=femoral neck; Fx=fraction; g/cm²=grams per centimeter squared; GI=gastrointestinal; HORIZON=Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly trial; IU=international units; IV=Intravenous infusion; LS=lumbar spine; MORE=Multiple Outcomes of Raloxifene Evaluation; mg=milligram; N=number; NSAID=nonsteroidal anti-inflammatory drug; NR=not reported; ONJ=osteonecrosis of the jaw; PE=pulmonary embolism; RCT=randomized controlled trial; RVFx=Radiographic vertebral fracture; SD=standard deviation; TH=total hip; UGI=upper gastrointestinal; VFx=vertebral fracture; yr=year

* Includes pooled RCTs and RCT subgroup analyses
### Table D3. Key Questions 1 and 2 evidence overview

<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Final Outcomes Incident Clinical Fracture</th>
<th>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs placebo</td>
<td>Bauer, 2006 (Pooled FIT-I &amp; FIT-II RCTs 3 yr FIT-I; 4 yr FIT-II) Medium</td>
<td>NR</td>
<td>NR</td>
<td>Women with osteoporosis or prevalent spine fracture (subgroup n=3495) NonVFx: ALN: n (%) NR Placebo: n (%) NR HR 0.69 (0.58, 0.83), p=NR RVFx: ALN: n (%) NR Placebo: n (%) NR HR 0.50 (0.39, 0.65), p=NR Pretreatment P1NP (tertile), osteoporotic women only (subgroup*subgroup): NonVFx: ALN efficacy (vs. placebo) greater among women with high pretreatment P1NP: Lowest P1NP tertile: HR 0.88 (0.65, 1.21) Highest P1NP tertile: HR 0.54 (0.39, 0.74) p for trend=0.03; interaction (P1NP continuous) = 0.02 Pretreatment BSALP and sCTx on nonvertebral fracture in osteoporotic women: similar direction but not statistically significant. Women without osteoporosis, no prevalent spine fracture (subgroup n=2689) Non VFx: ALN: n (%) NR Placebo: n (%) NR HR 1.11 (0.88, 1.44), p=NR (NS) RVFx: ALN: n (%) NR Placebo: n (%) NR HR 0.65 (0.37, 1.16), p=NR (NS) Pretreatment P1NP (tertile),</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>

**Drug Comparison**

**Study (Trial) Followup Risk of Bias**

**Final Outcomes Incident Clinical Fracture**

**Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change**

**Predictors (Patient, Bone, Drug)**

**Cummings 1998 (FIT-II)**

**4 yr**

**Up to 4 yr Multiple fxs in respective categories:**

Any CFx, baseline N/group:
- ALN: 272/2214 (12.3%)
- Placebo: 312/2218 (14.1%)
- HR=0.86 (95% CI 0.73, 1.01) p=0.07

Any nonVFx:
- ALN: 261/2214 (11.8%)
- Placebo: 294/2218 (13.3%)
- RH=0.88 (95% CI 0.74, 1.04) p=0.13

Hip:
- ALN: 19/2214 (0.9%)
- Placebo: 24/2218 (1.1%)

**New RVFx:**

≥1:
- ALN: 43/2057 (2.1%)
- Placebo: 78/2077 (3.8%)
- HR=0.56 (95% CI 0.39, 0.80) p=0.002

≥2:
- ALN: 4/2057 (0.2%)
- Placebo: 10/2077 (0.5%)
- RH=0.40 (95% CI 0.13, 1.24) p=0.11

**Change in BMD over 4 yr (denominators per measure NR):**

FN BMD:

-2.5 to -2.0:
- ALN: 92/NR (12.7%)
- Placebo: 87/NR (12.3%)
- HR=1.03 (95% CI 0.77, 1.39)

-2.0 to -1.6:
- ALN: 73/NR (10.9%)
- Placebo: 66/NR (9.5%)

**CFx by baseline FN BMD T-score. Interaction p value (BMD continuous)=0.01;**

<~2.5:
- ALN: 107/NR (13.1%)
- Placebo: 159/NR (19.6%)
- HR=0.64 (95% CI 0.50, 0.82)

-2.5 to -2.0:
- ALN: 92/NR (12.7%)
- Placebo: 87/NR (12.3%)
- HR=1.03 (95% CI 0.77, 1.39)

**NonVFx:**

ALN vs. placebo
- Low P1NP tertile: HR 1.46 (0.98, 2.17)
- High P1NP tertile: HR 0.77 (0.51, 1.17)
- p for trend=0.03; interaction (P1NP continuous) = 0.08

No evidence of a relationship between treatment BSALP and sCTx on nonvertebral fracture in nonosteoporotic women.

**RVFx:**

Higher pretreatment BSALP was associated with greater alendronate efficacy on vertebral fracture
- Low BSALP tertile: OR 1.22 (0.45, 3.33)
- High BSALP tertile: OR 0.25 (0.07, 0.88)
- p for trend=0.05; interaction p value (BSALP continuous) = 0.08

Results did not change when authors adjusted for baseline BMD (data NR)
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup</th>
<th>Risk of Bias</th>
<th>Final Outcomes Incident Clinical Fracture</th>
<th>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR=0.79 (95% CI 0.43, 1.44) p=0.44</td>
<td>RH=1.14 (95% CI 0.82, 1.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Wrist:</td>
<td>Post hoc incident hip fracture by dichotomized baseline FN-BMD T-score:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ALN: 83/2214 (3.7%) Placebo: 70/2218 (3.2%)</td>
<td>&lt;-2.5:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR=1.19 (95% CI 0.87, 1.64) p=0.28</td>
<td>ALN: 8/NR (1.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Other than hip, wrist or spine:</td>
<td>Placebo: 18/NR (2.2%); HR (95% CI): 0.44 (0.18, 0.97), p=NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ALN: 182/2214 (8.2%) Placebo: 227/2218 (10.2%)</td>
<td>&gt;2.5:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HR=0.79 (95% CI 0.65, 0.96) p=0.02</td>
<td>ALN: 11/NR (0.8%)</td>
</tr>
<tr>
<td>Donaldson 2012</td>
<td>(Pooled FIT I &amp; II RCTs)</td>
<td>3 yr FIT-I; 4 yr FIT-II</td>
<td>Overall: First NonVFx: ALN: NR/3236 (% NR) Placebo: NR/3223 (% NR) IRR=0.86 (95% CI 0.75, 0.99) p=NR</td>
<td>ALN: +3.8% Placebo: -0.8% Difference: 4.6% (p&lt;0.001) TH BMD: ALN: +3.4% Placebo: -1.6% Difference: 5.0% (p&lt;0.001) LS BMD: ALN: +8.3% Placebo: +1.5% Difference: 6.6% (p&lt;0.001)</td>
<td>First NonVFx by FRAX (with FN BMD); by FRAX tertiles, per 100 person years: FRAX 4.75-22.06: ALN rate: 2.21 Placebo rate: 2.55 IRR=0.87 (95% CI 0.65, 1.15) RD=-0.34 (95% CI -1.0, 0.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medium</td>
<td></td>
<td>R VFx: Data not reported/shown. Interaction between treatment (ALN/placebo) and (FRAX with BMD): p=0.88</td>
<td>FRAX 22.07-34.19: ALN rate: 3.32 Placebo rate: 3.73 IRR=0.89 (95% CI 0.70, 1.13)</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td>Jamal 2007</td>
<td>(Pooled FIT-I &amp; FIT-II RCTs 3 yr FIT-I; 4 yr FIT-II)</td>
<td>Medium</td>
<td>RD=-0.41 (95% CI -1.3, 0.43) FRAX 34.2-85.36: ALN rate: 4.75 Placebo rate: 5.66 IRR=0.84 (95% CI 0.68, 1.03) RD=-0.91 (95% CI -2.0, 0.17) First NonVFx by baseline FN BMD: Interaction between treatment and FN BMD: p=0.014. Within FN T-score ≤-2.5 stratum: ALN: NR/stratum NR (% NR) Placebo: NR/stratum NR (% NR) IRR=0.76 (95% CI 0.62, 0.93) p=NR &quot;No evidence of benefit&quot; of ALN in women with FN T-score &gt;-2.5; data not shown: IRR=0.96 (95% CI 0.80, 1.16) Also assessed CFx, MOF and RVFx but results NR (interactions NS)</td>
<td>CFx: &quot;ALN (vs. placebo) reduced the risk of clinical fractures regardless of renal function&quot;: OR (95% CI): 0.8 (0.70, 0.9); n, %, and p value NR RVFx: &quot;Overall, alendronate (vs. placebo) reduced the risk of spine fractures&quot; OR (95% CI): 0.54 (0.37, 0.78) n, %, and p value NR All women, by renal function (eGFR&lt;45 ml/min vs ≥ 45 ml/min). n=6458 CFx: P value for interaction=0.89: With reduced eGFR: Alendronate: n (%)=NR Placebo: n (%)=NR OR (95% CI): 0.78 (0.51, 1.21) Without reduced eGFR: Alendronate: n (%)=NR Placebo: n (%)=NR OR (95% CI): 0.80 (0.70, 0.93) RVFx: P value for interaction=0.44: With reduced eGFR: Alendronate: n (%)=NR Placebo: n (%)=NR</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Placebo: n (%)=NR OR (95% CI): 0.72 (0.31, 1.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: Alendronate: n (%)=NR OR (95% CI): 0.50 (0.32, 0.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Placebo: n (%)=NR+ Subgroup with osteoporosis (FN BMD T score &lt; -2.5, n=3214)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CFx: P value for interaction=0.72: With reduced eGFR: Alendronate: n (%)=NR OR (95% CI): 0.84 (0.45, 1.54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Placebo: n (%)=NR Without reduced eGFR: Alendronate: n (%)=NR OR (95% CI): 0.74 (0.61, 0.91)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RVFx: P value for interaction=0.49: With reduced eGFR: Alendronate: n (%)=NR OR (95% CI): 1.01 (0.29, 3.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Placebo: n (%)=NR Without reduced eGFR: Alendronate: n (%)=NR OR (95% CI): 0.62 (0.36, 1.10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ryder 2008 (FIT-II RCT subgroup) 4.2 yr Medium</td>
<td>NonVFx, in FIT-II women with low BMD: ALN 158/1389 (11.4%) Placebo: 143/1396 (10.2%) HR=1.12 (95% CI 0.89, 1.40)</td>
<td>NR</td>
<td>NonVFx, in FIT-II women with low BMD, stratified by prior fx after age 45: interaction p value=0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No prior fx since age 45: ALN: 88/961 (9.3%)* Placebo: 86/954 (9.0%)* HR=1.02 (95% CI 0.76, 1.38)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Yes, prior fx since age 45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td>ALN: 70/438 (16.0%)* Placebo: 57/442 (12.9%)* HR=1.26 (95% CI 0.89, 1.79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quandt 2005 (FIT-I and FIT-II subgroup) 3-4.5 yr Medium</td>
<td>FIT-II subgroup with &gt;3 yr followup: CVFx (women without baseline VFx): ALN: 6/1394 (0.43%)* Placebo: 13/1403 (0.93%)* (denominators/group from Table 1) RR=0.46 (95% CI 0.16, 1.17)</td>
<td>FIT-II subgroup (&gt;3 yr followup): RVFx (&gt;4 yr), women without baseline VFx: ALN: 21/UTD Placebo: 33/UTD RR=0.64 (95% CI 0.38, 1.10) Denominators NR for FIT-II subgroup, this outcome. Not all women had x-rays (total N for RVFx = 3532; total N for this study = 3737)</td>
<td>NR for outcomes &gt;3 yr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zoledronic acid vs placebo Black 2012 (HORIZON) 6 yr Low</td>
<td>Incident CFx, all 6 yr HR=1.04 (95% CI 0.71, 1.54) Incident CVFx HR=1.81 (95% CI 0.53, 6.2) Incident nonCVFx Zoledronic acid: 45/616 (7.3%)* Placebo: 47/617 (7.6%)* HR=0.99 (95% CI 0.7, 1.5) Incident CFx, hip Zoledronic acid: 7/616 (1.1%) Placebo: 8/617 (1.3%) HR=0.90 (95% CI 0.33, 2.49)</td>
<td>Incident RVFx 6 yr Zoledronic acid: 14/469 (3.0%) Placebo: 30/486 (6.2%) OR=0.51 (95% CI 0.26, 0.95), p=0.035 adjusted for number of baseline VFx BMD, mean % change from randomization (yr 3) to 6 yr FN BMD Zoledronic acid: 0.24 Placebo: -0.80 Mean % difference 1.04 (95% CI 0.43, 1.65), p&lt;0.001 TH BMD Zoledronic acid: -0.36 Placebo: -1.58 Mean % difference 1.22 (95% CI 0.75, 1.70), p&lt;0.001 LS BMD</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Zoledronic acid: 3.2 Placebo: 1.18 Mean % difference 2.03 (95%CI 0.76, 3.29), p=-0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black 2015 (HORIZON) 9 yr Low</td>
<td>Incident CFx, all 9 yr Zoledronic acid: 10/95 (12.2%)* Placebo: 9/95 (9.5%)* HR=1.11 (95%CI 0.45, 2.73)</td>
<td>Incident RVFx 9 yr Zoledronic acid: 3/68 (4.4%) Placebo: 5/69 (7.2%) OR=0.61 (95% CI 0.14, 2.77) BMD, mean % change from randomization (yr 6) to 9 yr FN BMD Zoledronic acid: -1.11 Placebo: -1.17 Mean % difference=0.06 (95% CI -1.41, 1.53) p=0.935 TH BMD Zoledronic acid: -0.54 Placebo: -1.31 Mean % difference=0.78 (95% CI -0.37, 1.93) p=0.183</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reid, 2018 6 yr Low</td>
<td>Fragility Fx Zoledronic acid: 122/1000 (12%) Placebo: 190/1000 (19%) HR=0.63 (95% CI 0.50, 0.79) P&lt;0.001</td>
<td>Percent change in LS BMD (estimated from Figure S1) Baseline to Year 3 Zoledronic acid: 5.5% Placebo: -1% Baseline to Year 6: Zoledronic acid: 7% Placebo: -1% Percent change in TH BMD (estimated from Figure S1) Baseline to Year 3 Zoledronic acid: 3.5%</td>
<td>Predictor: Baseline VFx status Outcome: Fragility Fx p for interaction =0.27 HR among participants without baseline VFx=0.65 (95% CI 0.50, 0.83) Predictor: Osteoporosis status Outcome: Fragility Fx HR excluding osteoporotic participants=0.63 (95% CI 0.49, 0.80) Predictor: Baseline Fx risk Outcome: Fragility Fx</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------------------------</td>
<td>------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td></td>
</tr>
<tr>
<td>Denosumab vs placebo</td>
<td>Miller 2008 4 yr Medium</td>
<td>CFx Denosumab 33/314 (11%) Placebo 5/46 (11%) RR=0.97 (95% CI 0.40, 2.35)* p&gt;0.99</td>
<td>BMD, % change from baseline LS Continued denosumab: (n=153) 9.4 to 11.8 Placebo: (n=30) -2.4 p&lt;0.001</td>
<td>TH Continued denosumab: 4.0 to 6.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Osteoporotic fx Denosumab: 22/314 (7%) Placebo: 4/46 (9%) RR=0.81 (95% CI 0.29, 2.23)*</td>
<td>Discontinued denosumab (n=31) at 36 months: -6.6 Discontinued/re-treatment denosumab (n=30): 9.0 p&lt;0.05 for both groups vs placebo</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total VFx Zoledronic acid: 23/1000 (2.3%) Placebo: 49/1000 (4.9%) OR=0.45 (95% CI 0.27, 0.73)</td>
<td>Placebo: -2% Baseline to Year 6: Zoledronic acid: 3.5% Placebo: -3.8%</td>
<td>HR excluding participants with hip Fx risk &gt;3% or osteoporotic Fx risk &gt;20%=0.60 (95% CI 0.44, 0.81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CVFx Zoledronic acid: 14/1000 (1.4%) Placebo: 34/1000 (3.4%) HR=0.41 (95% CI 0.22, 0.75)</td>
<td>Percent change in Total Body BMD (estimated from Figure S1) Baseline to Year 3 Zoledronic acid: 1.3% Placebo: -1.6%</td>
<td>HR excluding participants with hip Fx risk&gt;3% or osteoporotic Fx risk&gt;20% or hx of NVFx after age 45=0.60 (95% CI 0.39, 0.91)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NV Fragility Fx Zoledronic acid: 101/1000 (10%) Placebo: 148/1000 (15%) HR=0.66 (95% CI 0.27, 1.16)</td>
<td>Baseline to Year 6: Zoledronic acid: 1.8% Placebo: -2%</td>
<td>Predictor: Baseline Fx risk Outcome: NVFx</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forearm or wrist Fx Zoledronic acid: 36/1000 (3.6%) Placebo: 63/1000 (6.3%) HR=0.56 (95% CI 0.37, 0.85)</td>
<td>All counts represent number of participants with event, not total number of events</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Denosumab vs placebo

Miller 2008 4 yr Medium

CFx Denosumab 33/314 (11%) Placebo 5/46 (11%) RR=0.97 (95% CI 0.40, 2.35)* p>0.99

Osteoporotic fx Denosumab: 22/314 (7%) Placebo: 4/46 (9%) RR=0.81 (95% CI 0.29, 2.23)*

BMD, % change from baseline LS Continued denosumab: (n=153) 9.4 to 11.8 Placebo: (n=30) -2.4 p<0.001

Discontinued denosumab (n=31) at 36 months: -6.6 Discontinued/re-treatment denosumab (n=30): 9.0 p<0.05 for both groups vs placebo

TH Continued denosumab: 4.0 to 6.1
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Final Outcomes Incident Clinical Fracture</th>
<th>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raloxifene vs placebo</td>
<td>Sontag 2010 Johnell 2004 (Raloxifene 60 mg and placebo groups only) (MORE) 3.3 yr (mean) Low (Sontag) Medium (Johnell)</td>
<td>Incident CVFx 4 yr Raloxifene 60 mg: 62/2549 (2%) Placebo: 107/2565 (4%)* p=0.003 (Data from Sontag 2010; also reported in Johnell 2004)</td>
<td>Incident RVFx Raloxifene: 60 mg 181/2259 (8%) Placebo: 288/2292 (13%)* (Data from Sontag 2010; also reported in Johnell 2004)</td>
<td>Predictor: Baseline VFx status Outcome: CVFx No baseline VFx Raloxifene 60 mg: 8/1574 (0.5%) Placebo: 26/1629 (2%) With baseline VFx Raloxifene 60 mg: 54/975 (6%) Placebo: 81/936 (9%) p for interaction=0.127 Outcome: RVFx No baseline VFx Raloxifene 60 mg: 41/1574 (3%) Placebo: 84/1629 (5%) With baseline VFx Raloxifene 60 mg: 140/975 (14%) Placebo: 204/936 (22%) p for interaction=0.402 (Data from Sontag 2010; also reported in Johnell 2004)</td>
</tr>
<tr>
<td>Raloxifene vs placebo</td>
<td>Delmas 2002 (subjects with a baseline and ≥1 post-baseline vertebral Xray)</td>
<td>Incident Hip fx Raloxifene 60 mg + 120 mg: 56/4536 (1.1%) Placebo: 29/2292 (1.1%) RR=0.97 (95% CI 0.62 to 1.52) (Data from Delmas 2002; also reported in)</td>
<td>Incident RVFx, Kaplan-Meier log-rank test Raloxifene 60 mg: n/N=NR Raloxifene 120 mg: n/N=NR Placebo: n/N=NR Raloxifene 60 mg vs placebo</td>
<td>Predictor: Baseline VFx Outcome: RVFx No baseline VFx Raloxifene 60 mg vs placebo RR=0.51 (95% CI 0.35, 0.73) Raloxifene 120 mg vs placebo</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Raloxifene vs placebo</td>
<td>n=6828) Barrett-Connor 2004 (MORE) 3.3 yr (mean) Low</td>
<td>Barrett-Conner 2004)</td>
<td>RR=0.64 (95% CI 0.53, 0.76) Raloxifene 120 mg vs placebo RR=0.57 (95% CI 0.48, 0.69) p&lt;0.001 for difference between raloxifene groups and placebo Change in LS BMD vs placebo, ANOVA Raloxifene 60 mg: 2.6% Raloxifene 120 mg: 2.5% p&lt;0.001 Change in FN BMD vs placebo, ANOVA Raloxifene 60 mg: 2.1% Raloxifene 120 mg: 2.3% p&lt;0.001</td>
<td>RR=0.62 (95% CI 0.44, 0.87) With baseline VFx Raloxifene 60 mg vs placebo RR=0.66 (95% CI 0.55, 0.81) Raloxifene 120 mg vs placebo RR=0.54 (95% CI 0.44, 0.66) (Data from Delmas 2002)</td>
</tr>
<tr>
<td>Siris 2005 (MORE/CORE) 8 yr Medium</td>
<td>Incident clinical nonvertebral fracture Raloxifene 60 mg n/N= 22.8% Placebo n/N= 22.9% Raloxifene 60 mg vs placebo HR=1.00 (0.82,1.21) Incident clinical nonvertebral-6 fracture (clavicle, humerus, wrist, pelvis, hip, lower leg) Raloxifene 60 mg n/N= 17.5% Placebo n/N= 17.5% Raloxifene 60 mg vs placebo HR=1.01 (0.81, 1.26) BMD Reported in Siris 2005 (high ROB)</td>
<td>Baseline age Treatment-by-subgroup interaction NS Baseline BMD Treatment-by-subgroup interaction NS Prevalent vertebral fracture Treatment-by-subgroup interaction p&lt;0.10 (significant) No significant reduction in nonvertebral or nonvertebral-6 fracture in either subgroup (with or without vertebral fracture)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hormone therapy vs control</td>
<td>Wimalawansa 4 yr Medium</td>
<td>Incident nonCVFx 4 yrHormone therapy: 1/15 (6.7%) Control: 1/14 (7.1%)P=1.0 (Fisher’s exact test) *</td>
<td>Incident RVFx 4 yrHormone therapy: 2/15 (13.3%) Control: 5/14 (35.7%)</td>
<td>NR</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>
| Hormone therapy (estrogen + progestin) vs placebo | Cauley 2003 6 yr Low | Incident CFx, all 5.6 yr  
Hormone therapy: 733/8506 (8.6%, annualized percentage 1.5%)  
Placebo: 896/8102 (11.1%, annualized percentage 2.0%)  
HR 0.76 (95% CI 0.69, 0.83)  
Incident Hip Fx 5.6 yr  
Hormone therapy: 52/8506 (0.61%, annualized percentage 0.11%)  
Placebo: 73/8102 (0.90%, annualized percentage 0.16%)  
HR 0.67 (95% CI 0.47, 0.96)  
Incident CVF 5.6 yr  
Hormone therapy: 41/8506 (0.48%, 11 per 10,000 person years)  
Placebo: 60/8102 (0.74%, 17 per 10,000 person years) | RR 0.37 (95%CI 0.09, 1.62)*  
P=0.19*  
BMD, mean % change from randomization to 4 yr  
LS BMD Hormone therapy: 7 (SE 1.0)  
Control: -2.5 (SE 0.8)  
Mean % difference 9.50 (95%CI 6.99, 12.01)*  
p<0.001  
TH BMD Hormone therapy: 4.8 (SE 0.9)  
Control: -4.4 (SE 0.8)  
Mean % difference 9.20 (95%CI 6.83, 11.57)*  
p<0.001 | Predictor: History of fracture  
Outcome: Incident CFx  
With history of fracture:  
Hormone therapy: 330/2950 (11.2%, annualized percentage 2.1%)  
Placebo: 417/2947 (14.2%, annualized percentage 2.6%)  
HR 0.78 (95% CI 0.68, 0.91)  
Without history of fracture:  
Hormone therapy: 298/5556 (5.4%, annualized percentage 1.2%)  
Placebo: 391/5155 (7.6%, annualized percentage 1.6%)  
HR 0.74 (95% CI 0.63, 0.86)  
p for interaction=0.59  
Predictor: Baseline BMD T score  
Outcome: Incident CFx  
T score >2.5:  
Hormone therapy: 11/NR (annualized percentage 1.4%) |
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Final Outcomes Incident Clinical Fracture</th>
<th>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone therapy (estrogen) vs placebo</td>
<td>Jackson 2006 7 yr Low</td>
<td>Incident CFx, all 7.1 yr</td>
<td>Hormone therapy: 540/5310 (10.2%, annualized percentage 1.4%) Placebo: 761/5429 (14.0%, annualized percentage 2.0%) HR 0.71 (95% CI 0.64, 0.80)</td>
<td>Predictor: History of fracture Outcome: Incident CFx With history of fracture: Hormone therapy: 267/1916 (13.9%, annualized percentage 2.0%) Placebo: 354/1900 (18.6%, annualized percentage 2.7%) HR 0.73 (95% CI 0.62, 0.86) Without history of fracture: Hormone therapy: 220/3394 (6.5%, annualized percentage 1.1%) Placebo: 321/3529 (9.1%, annualized percentage 1.6%) HR 0.68 (95% CI 0.57, 0.81)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident Hip Fx 7.1 yr</td>
<td>Hormone therapy: 46/5310 (0.87%, annualized percentage 0.12%) Placebo: 73/5429 (1.3%, annualized percentage 0.19%)</td>
<td>NR for subgroups of interest</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR 0.65 (95% CI 0.45, 0.94)</td>
<td></td>
<td>p for interaction=0.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Incident CVF</td>
<td></td>
<td>Predictor: Baseline BMD T score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.1 yr</td>
<td></td>
<td>Outcome: Incident CFx</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hormone therapy: 43/5310 (0.81%, 11 per 10,000 person years)</td>
<td>T score ≤2.5 (osteoporotic): HR 0.83 (95% CI 0.17, 3.91)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Placebo: 69/5429 (1.3%, 18 per 10,000 person years)</td>
<td>T score &gt; -1.0 to -2.5 (osteopenic): HR 0.83 (95% CI 0.49, 1.40)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HR 0.64 (95% CI 0.44, 0.93)</td>
<td>T score &gt; -1.0 (normal bone mass): HR 0.99 (95% CI 0.53, 1.84)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Note these outcomes are for the total study population, which includes women who are neither osteoporotic nor osteopenic.</td>
<td>p for interaction=0.17</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** ALN=alendronate; BMD=bone mineral density; BMI=body mass index; CFx=clinical fracture; CORE=Continuing Outcomes Relevant to Evista; DXA=Dual-energy x-ray absorptiometry; FIT=Fracture Intervention Trial; FLEX=Fracture Intervention Trial Long Term Extension; FN=femoral neck; FRAX=World Health Organization (WHO) Fracture Risk Assessment Tool; Fx=fracture; HORIZON=Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly trial; HR=hazard ratio; IRR=incidence rate ratio; LS=lumbar spine; mg=milligrams; MOF=major osteoporotic fracture; MORE=Multiple Outcomes of Raloxifene Evaluation; N=number; NR=not reported; NS=not significant; RD=risk difference; RH=relative hazard; RVFx=radio graphic vertebral fracture; RR=risk ratio; TH=total hip; VFx=vertebral fracture; vs=versus; yr=years

*Calculated by EPC
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs placebo</td>
<td>Bauer 2000 (Pooled FIT I &amp; II) Mean 3.8 yr Medium</td>
<td>Any upper GI AE ALN: 1536/3236 (47.5%) Placebo: 1490/3223 (46.2%) RR 1.02 (95% CI 0.95, 1.10)</td>
<td>NR</td>
<td>Age (Fig 3): “perforation/ulcer/bleed increased with age but no evidence of disproportionate # of events with ALN” (pg 522)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Any gastric or duodenal AE ALN: 130/3236 (4.0%) Placebo: 129/3223 (4.0%) RR: NR</td>
<td></td>
<td>History of UGI disease Gastroduodenal perforation/ulcer/bleed, esophageal events more common with history of UGI disease; no evidence of disproportionate events on ALN.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gastroduodenal perforation/ulcer/bleed ALN: 53/3236 (1.6%) Placebo: 61/3223 (1.9%) RR: NR</td>
<td></td>
<td>NSAIDs Events more common on NSAIDs (both groups)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Any esophageal AE ALN: 322/3236 (10.0%) Placebo: 309/3223 (9.4%) RR: NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black 2010 (pooled Fit-I and Fit-II) Mean 3.8 yr Low</td>
<td>NR</td>
<td>Subtrochanteric or diaphyseal femur fracture ALN: 1/3236 (0.031%) Placebo: 1/3223 (0.031%) RH= 1.03 (95% CI 0.06, 16.46) p=0.98</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Cummings 1998 (FIT-II) 4 yr Low</td>
<td>Any upper GI event ALN: 1052/2214 (47.5%) Placebo: 1047/2218 (47.2%) RH = 1.00 (95% CI 0.92, 1.09) Serious AE category: NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Any AE resulting in hospitalization: ALN: 644/2214 (29.1%) Placebo: 596/2218 (26.9%) HR (95% CI): 1.09 (0.98, 1.22), p = NR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Death ALN: 37/2214 (1.7%) Placebo: 40/2218 (1.8%) HR = 0.92 (95% CI 0.59, 1.45)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ALN: 1052/2214 (47.5%) Placebo: 1047/2218 (47.2%) RH = 1.00 (95% CI 0.92, 1.09) Serious AE category: NR

Any AE resulting in hospitalization: ALN: 644/2214 (29.1%) Placebo: 596/2218 (26.9%) HR (95% CI): 1.09 (0.98, 1.22), p = NR

Death ALN: 37/2214 (1.7%) Placebo: 40/2218 (1.8%) HR = 0.92 (95% CI 0.59, 1.45)
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jamal 2007</td>
<td>Pooled (FIT-I + FIT-II) Mean follow-up 3 yr FIT-I, 4 yr FIT-II Medium</td>
<td>NR</td>
<td>NR</td>
<td>Reduced renal function (eGFR&lt;45 ml/min, n=581) vs without reduced renal function (eGFR≥45 ml/min, n=5877).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GI events</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 4.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 5.2% p=0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrhythmias</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 2.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 2.1% p=0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cardiovascular events</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 2.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 3.2% p=0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Renal AEs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 12/581 (2.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 137/5877 (2.3%), p=0.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cancer (any)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 4.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 5.0% p=0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Death</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>With reduced eGFR: 1.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Without reduced eGFR: 1.9% p=0.5</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Alendronate vs active comparator</td>
<td>Chiu 2014 4 yr Medium</td>
<td>NR</td>
<td>ONJ: defined by 3 criteria [Principal hospital diagnosis of inflammatory dental condition (526.3X, 526.4X, 730.1X, 733.4X; Claim for ≥1 dentistry visit at hospital; Review x-ray, chart, surgical and pathology reports for ONJ confirmation] Crude incidence ONJ: 40/7332 (0.05%) Incidence per 100,000 persons per year: Alendronate: 318/100,000 Raloxifene: 35/100,000 Attributable risk of ONJ associated with alendronate: 283/100,000 person years aHR 7.42 (95% CI 1.02, 54.09) p=0.48</td>
<td>Risk of ONJ From 1 logistic regression model of age, sex, drug duration, and DM associated with ONJ on alendronate. Sex, Female vs male aOR=4.77 (95% CI 0.65, 35.07) p=0.125 Age (yr) 80+ vs. &lt;65: aOR=5.65 (95% CI 1.57, 20.38) p=0.008 65-80 vs. &lt;65: aOR=4.14 (95% CI 1.24, 13.89) p=0.021 Drug duration, 3+ vs. &lt;3 yr aOR=5.73 (95% CI 2.97, 11.05) p&lt;0.001 DM (yes vs. no) aOR=2.00 (95% CI 1.04, 3.87) p=0.039</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Alendronate vs raloxifene/calcitonin</td>
<td>Lin 2014 6 yr Medium</td>
<td>NR</td>
<td>ONJ: diagnostic codes (ICD-9: 73008, 73000, 73340, 73349, 73018, 73010, 73020, 73345, 73399, 52689, 7339, 5264, 5289, 5259, and 5269). Confirmed the diagnosis persisted for 8 weeks or longer while allowing a 30-day gap between records. Incidence ALN: 24/16003 (0.15%) Raloxifene or calcitonin: 13/16003 (0.08%) Incidence per 10,000 person-yr ALN: 8.2 Raloxifene or calcitonin: 6.9 aHR=0.86 (95% CI 0.44, 1.69) p=0.659</td>
<td>NR</td>
</tr>
<tr>
<td>Alendronate vs no treatment</td>
<td>Abrahamsen 2009 0-9 yr</td>
<td>NR</td>
<td>AFF defined by ICD-10 codes for subtrochanteric (S72.2) and femoral diaphysis (S72.30) fractures Longterm alendronate use (&gt;6 yr) + highly refill compliant (MPR&gt;80%) ALN and no treatment combined: 5/534 (0.94%) aHR= 1.37 (95% CI, 0.22, 8.62), p=0.74 Full cohort ALN: 35/5187 (0.67%) No treatment: 41/10374 (0.40%) HR= 1.64 (95% CI 1.05, 2.58) aHR= 1.46 (95% CI 0.91, 2.35), p=0.12</td>
<td>NR</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>
| Zoledronic acid vs placebo | Black 2012 (HORIZON extension) 6 yr Low | Mortality  
Zoledronic acid: 26/613 (4.2%)  
Placebo: 18/616 (2.9%)  
p=0.22  
Myocardial infarction, any  
Zoledronic acid: 6/613 (1.0%)  
Placebo: 4/616 (0.6%)  
Myocardial infarction, severe  
Zoledronic acid: 5/13 (0.8%)  
Placebo: 4/616 (0.6%) | Afib, any  
Zoledronic acid: 21/613 (3.4%)  
Placebo: 13/616 (2.1%)  
p=0.17  
Afib, severe  
Zoledronic acid: 12/613 (2.0%)  
Placebo: 7/616 (1.1%)  
p=0.26  
AFF  
Zoledronic acid: 0/613 (0%)  
Placebo: 0/616 (0%)  
ONJ  
Zoledronic acid: 1/613 (<1%)  
Placebo: 0/616 (0%) |  |
| | Black 2015 (HORIZON extension) 9 yr Low | Mortality  
Zoledronic acid: 1/92 (1.1%)  
Placebo: 5/95 (5.3%)  
p=0.21 | Afib, any  
Zoledronic acid: 5/92 (5.4%)  
Placebo: 1/95 (1.1%)  
p=0.11  
Afib, severe  
Zoledronic acid: 1/92 (1.1%)  
Placebo: 1/95 (1.1%)  
p=1.0  
AFF  
Zoledronic acid: 0/92 (0%)  
Placebo: 0/95 (0%)  
ONJ  
Zoledronic acid: 0/92 (0%)  
Placebo: 0/95 (0%) | NR |
<table>
<thead>
<tr>
<th>Event</th>
<th>Zoledronic acid</th>
<th>Placebo</th>
<th>OR (95% CI)</th>
<th>NR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious adverse events</td>
<td>Zoledronic acid: 400/1000 (40%)</td>
<td>Placebo: 443/1000 (44%)</td>
<td>OR=0.84 (95% CI, 0.70, 1.00)</td>
<td>NR</td>
</tr>
<tr>
<td>Death</td>
<td>Zoledronic acid: 27/1000 (2.7%)</td>
<td>Placebo: 41/1000 (4.1%)</td>
<td>OR=0.65 (95% CI, 0.40, 1.05)</td>
<td>NR</td>
</tr>
<tr>
<td>Sudden death</td>
<td>Zoledronic acid: 3/1000 (0.3%)</td>
<td>Placebo: 1/1000 (0.1%)</td>
<td>OR=3.01 (95% CI, 0.3, 28.9)</td>
<td>NR</td>
</tr>
<tr>
<td>MI</td>
<td>Zoledronic acid: 24/1000 (2.4%)</td>
<td>Placebo: 39/1000 (3.9%)</td>
<td>OR=0.61 (95% CI, 0.36, 1.02)</td>
<td>NR</td>
</tr>
<tr>
<td>Stroke</td>
<td>Zoledronic acid: 17/1000 (1.7%)</td>
<td>Placebo: 20/1000 (2.0%)</td>
<td>OR=0.85 (95% CI, 0.44, 1.63)</td>
<td>NR</td>
</tr>
<tr>
<td>TIA</td>
<td>Zoledronic acid: 23/1000 (2.3%)</td>
<td>Placebo: 14/1000 (1.4%)</td>
<td>OR=1.66 (0.85, 3.24)</td>
<td>NR</td>
</tr>
<tr>
<td>Composite of vascular events</td>
<td>Zoledronic acid: 53/1000 (5.3%)</td>
<td>Placebo: 69/1000 (6.9%)</td>
<td>OR=0.76 (95% CI 0.52, 1.09)</td>
<td>NR</td>
</tr>
<tr>
<td>Cancer</td>
<td>Zoledronic acid: 84/1000 (8.4%)</td>
<td>Placebo: 121/1000 (12%)</td>
<td>OR=0.67 (0.50, 0.89)</td>
<td>NR</td>
</tr>
</tbody>
</table>

All counts represent number of participants with event, not total number of events.
reviewed by team blinded to exposure status

Any bisphosphonate

Time since first bisphosphonate prescription, >3 yr
Case: 6/44 (14%)
Control: 8/220 (4%)

Duration of use, >3 yr
Case: 5/44 (11%)
Control: 3/220 (1%)

Time since first bisphosphonate prescription
Model 1
aOR=4.71 (95% CI 1.52, 14.6)
Model 2
aOR=9.46 (95% CI 2.17, 41.3)

Duration
Model 1
aOR=9.18 (95% CI 2.12, 38.9)
Model 2
aOR=31.9 (95% CI 4.05, 251)

Model 1: Adjusted for matching variables
Model 2: Adjusted for matching variables, smoking, alcoholism, BMI, previous fracture, kidney disease, malabsorption, stroke, dementia, rheumatoid arthritis, diabetes, epilepsy, Parkinson disease, thyroid disease, PPI, anxiolytics, sedatives, antidepressants, antihypertensives, corticosteroids, raloxifene, hormone therapy and thiazolidinediones.
p=0.0007 for trend over duration of use
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koh 2017</td>
<td>5.7 yr Medium</td>
<td>NR</td>
<td>AFF defined by subtrochanteric or diaphyseal femoral fracture (ICD-10 S722, S723, S724, or S728), plus x-ray confirmation per 2014 ASBMR AFF criteria</td>
<td>Risk of AFF Bisphosphonate without drug holiday aHR=5.17 (95% CI 2.0, 13.36)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Glucocorticoid ≥ 1 yr aHR=3.04 (95% CI 1.40, 6.57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BMI aHR=1.24 (95% CI 1.1, 1.39)</td>
</tr>
</tbody>
</table>

Bisphosphonate users from which cases & controls were selected: 43/35104 (0.12%)
Mean bisphosphonate use: Cases: 7.3 yr Controls: 5.2 yr
Cox Model 1: factors associated with AFF
Bisphosphonate without drug holiday aHR=5.17 (95% CI 2.0, 13.36)
Glucocorticoid ≥ 1 yr aHR=3.04 (95% CI 1.40, 6.57)
BMI aHR=1.24 (95% CI 1.1, 1.39)

| Lim 2018 | 5.2 yr (mean) Medium | NR | AFF defined by subtrochanteric or diaphyseal femoral fracture (ICD-10 S72.0-S72.9) plus x-ray confirmation per 2013 ASBMR AFF criteria | |
| | | | | |
| | | | | |
| | | | | |

Bisphosphonate use
In AFF cases: 140/196
In TFF cases: 10/94
aOR=25.65 (95% CI 10.74, 61.28)
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Park-Wyllie 2011 4.0 yr Medium</td>
<td>NR</td>
<td>AFF, subtrochanteric or femoral shaft fracture in bisphosphonate users Overall incidence, bisphosphonate users &gt;7 yr: 716/205466 (0.35%) Absolute risk, bisphosphonate use &gt;5 yr: 6th yr: 71/52595 (0.13%) 7th yr: 117/52595 (0.22%) No x-ray confirmation. Only first fractures counted. Reference=bisphosphonate use &lt;100 days. In women on bisphosphonates &gt;5yr, 64% of subtrochanteric/femoral shaft fractures were attributable to bisphosphonate use.</td>
<td>Risk of subtrochanteric or femoral shaft fracture, by drug duration. 3-5 yr bisphosphonate use Cases: 204/716 (28.5%) Controls: 1070/3580 (29.9%) aOR=1.59 (95% CI 0.80, 3.15) 5+ yr bisphosphonate use Cases: 121/716 (16.9%) Controls: 460/3580 (12.9%) aOR=2.74 (95% CI 1.25, 6.02) Adjusted for socioeconomic status, comedications, drug count, comorbidities, recent medical visits, prior fall, prior osteoporotic fracture, BMD test past 5 yr.</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Schilcher 2014</td>
<td>4-5 yr Medium</td>
<td>NR</td>
<td>Cohort study age-adjusted RR by duration of use vs. no use ≥4 yr RR=126 (95% CI 55, 288) Absolute risk=11 (95% CI 7, 14) fractures per 10,000 patient-yr</td>
<td>Risk of AFF Age ≤80 yr RR=100 (95% CI 40, 253) Absolute risk=10 (95% CI 5, 15) fractures per 10,000 patient-yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case/control study 172 cases with AFF vs. 952 controls with ordinary femoral shaft Fx AFF defined by ICPC-2 codes for subtrochanteric or diaphyseal fracture and manually reviewed by team blinded to exposure status</td>
<td>Age &gt;80 yr RR=163 (95% CI 39, 687) Absolute risk=11 (95% CI 7, 16) fractures per 10,000 patient-yr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Any bisphosphonate 4-5 yr Case: 13/172 (8%) Control (ordinary shaft fx): 3/952 (0.3%) aOR=116 (95%CI 58, 234)</td>
<td>*Adjusted by age, sex, cortisone use, and Charlson’s comorbidity Index</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5+ yr Case: 11/172 (6%) Control (ordinary shaft fx): 5/952 (0.5%) aOR=93 (95% CI 66, 132)</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Biphosphonates vs raloxifene/calcitonin</td>
<td>Kim 2011 &lt;9.5 yr (mean 2 yr) Medium</td>
<td>NR</td>
<td>Subtrochanteric or diaphyseal femur fracture defined by ICD-9 codes 820.22 and 821.0x: 3-5 yr treatment Biphosphonates: 15/2591 (0.6%) Raloxifene/calcitonin: 11/2309 (0.5%) HR=1.20 (95% CI 0.55, 2.61) ≥5 yr treatment Biphosphonates: 6/2371 (0.3%) Raloxifene/calcitonin: 2/1726 (0.1%) HR=2.02 (95% CI 0.41, 10.00) Cox proportional hazards Propensity score matched on demographics, health care utilization, comorbidities, other medications</td>
<td>NR</td>
</tr>
<tr>
<td>Denosumab vs placebo</td>
<td>Miller 2008 4 yr Medium</td>
<td>Serious AEs</td>
<td>Denosumab: 56/314 (18%) Placebo: 5/46 (11%) p=0.30 Arthralgia Denosumab: 74/314 (24%) Placebo: 14/46 (30%) p=0.36 Infections Denosumab: 10/314 (3%) Placebo: 0/46 (0%) p=0.62 Malignant neoplasms Denosumab: 15/314 (5%) Placebo: 2/46 (4%) p&gt;0.99 (Fisher’s exact) (calculated by extractor)</td>
<td>NR</td>
</tr>
<tr>
<td>Event</td>
<td>No baseline vertebral fracture</td>
<td>With baseline vertebral fracture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------</td>
<td>----------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke</td>
<td>Raloxifene 60 mg: 23/1574 (1.5%)</td>
<td>Placebo: 27/1629 (1.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE</td>
<td>Raloxifene 60 mg: 20/975 (2.05%)</td>
<td>Placebo: 29/936 (3.1%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p for interaction=0.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>Raloxifene 60 mg: 6/1574 (0.2%)</td>
<td>Placebo: 3/1629 (0.18%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>With baseline vertebral fracture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raloxifene 60 mg: 5/975 (0.51%)</td>
<td>Placebo: 1/936 (0.11%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p for interaction=0.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVT</td>
<td>Raloxifene 60 mg: 12/1574 (0.76%)</td>
<td>Placebo: 6/1629 (0.37%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>With baseline vertebral fracture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raloxifene 60 mg: 8/975 (0.82%)</td>
<td>Placebo: 2/936 (0.21%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p for interaction=0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Raloxifene 60 mg: 158/1574 (10.0%)</td>
<td>Placebo: 103/1629 (6.3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>With baseline vertebral fracture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Raloxifene 60 mg: 79/975 (8.1%)</td>
<td>Placebo: 47/936 (5.02%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p for interaction=0.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mortality</td>
<td>No baseline vertebral fracture</td>
<td>Raloxifene 60 mg: 10/1574 (0.64%)</td>
<td>Placebo: 13/1629 (0.80%)</td>
<td></td>
</tr>
</tbody>
</table>

**Comparison:**

- **Raloxifene vs placebo**

  (Raloxifene 60 mg and placebo groups only)

  Barrett-Connor 2004

  (combined raloxifene 60 mg and raloxifene 120 mg groups)

  (MORE)

  3.3 yr (mean)

  **Low**
| Drug Comparison | Study (Trial) Followup Risk of Bias | Harms Drug class-specific harms | Rare Harms AFF, ONJ, Afib | Predictors (Patient, Bone, Drug) |
|----------------|------------------------------------|--------------------------------|=--------------------------|---------------------------------|
|                |                                    |                                |                          | With baseline vertebral fracture |
|                |                                    |                                |                          | Raloxifene 60 mg: 13/975 (1.33%) |
|                |                                    |                                |                          | Placebo: 23/936 (2.46%)         |
|                |                                    |                                |                          | p for interaction=0.47          |
|                |                                    |                                |                          | Logistic regression             |
|                |                                    |                                |                          | (Data from Sontag 2010)         |
| Delmas 2002    | DVT                                | Raloxifene 60 mg: 20/2557 (0.8%)| NR                       | NR                              |
| Cauley 2001    |                                    | Raloxifene: 120 mg 24/2572 (0.9%)|                          |                                 |
| (MORE)         | Low                                | Placebo: 8/2576 (0.3%)         |                          |                                 |
| 3.3 yr (mean)  |                                    | p=0.006 (pooled raloxifene groups vs placebo) |              |                                 |
| Low            |                                    | Rowling’s chi-square           |                          |                                 |
|                |                                    | (Data from Delmas 2002 and Cauley 2001) |              |                                 |
| Duvernoy 2005  | DVT or PE                          | Raloxifene 60 mg vs placebo    | NR                       | DVT or PE Increased cardiovascular risk |
| (Raloxifene 60 mg and placebo groups only) (MORE) |                                    | HR=1.86 (95% CI 0.97, 3.56)   |                          | Raloxifene 60 mg vs placebo      |
| 3.3 yr (mean)  | Low                                | Not significant                |                          | HR=1.32 (95% CI 0.37, 4.69)     |
| Low            |                                    | Cox proportional hazards       |                          | Not significant                  |
|                |                                    |                                |                          | Cox proportional hazards        |

Table Notes:
- **DVT** indicates deep vein thrombosis.
- **PE** indicates pulmonary embolism.
- **HR** indicates hazard ratio.
- **CI** indicates confidence interval.
- **NR** indicates not reported.
- **(Data from Delmas 2002 and Cauley 2001)** indicates the study details.
- **(Data from Sontag 2010)** indicates the study details.
- **Logistic regression** indicates the statistical method used.
- **Cox proportional hazards** indicates the statistical method used.

D-43
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grady 2004 (MORE) 3.3 yr (mean) Low</td>
<td>DVT</td>
<td>Raloxifene 60 mg + 120 mg: 43/5129 (0.8%) Placebo: 7/2576 (0.3%) RR=3.1 (95% CI 1.4, 6.9) p&lt;0.01</td>
<td>NR</td>
<td>DVT or PE Drug dose Raloxifene 60 mg vs placebo: RR=1.9 (95% CI 1.0, 3.6) Raloxifene 120 mg vs placebo: RR=2.4 (95% CI 1.3, 4.4) P for difference between RRs=0.4 Treatment duration DVT p for time x treatment interaction=0.01 PE p for time x treatment interaction=0.8</td>
</tr>
<tr>
<td></td>
<td>PE</td>
<td>Raloxifene 60 mg + 120 mg: 18/5129 (0.4%) Placebo: 2/2576 (0.08%) RR=4.5 (95% CI 1.1, 19.5) p=0.05 Logistic regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stroke</td>
<td>Raloxifene 60 mg: 22/2557 (1.2%) Raloxifene 120 mg: 26/2572 (0.9%) Placebo: 32/2576 (1.0%)</td>
<td>NR</td>
<td>Stroke Increased cardiovascular risk Raloxifene 60 mg: 6/359 (1.7%) Raloxifene 120 mg: 6/359 (1.7%) Placebo: 14/317 (4.4%) Raloxifene 60 mg vs placebo: RR=0.38 (0.15, 0.94) Raloxifene 120 mg vs placebo: RR=0.38 (0.15, 0.94) Prevalent CHD Raloxifene 60 mg: 1/56 (1.7%) Raloxifene 120 mg: 2/80 (2.5%) Placebo: 6/66 (9%) Raloxifene 60 mg vs placebo: RR=0.20 (95% CI 0.03, 1.25) Raloxifene 120 mg vs placebo: RR=0.28 (95% CI 0.06, 1.18)</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------</td>
<td>--------------------------------</td>
</tr>
</tbody>
</table>
| Grady 2010 (MORE/CORE) 8 yr Medium | Mortality  
Raloxifene: 45/2557 (1.8%)  
Placebo: 65/2576 (2.5%)  
HR=0.68 (95% CI 0.46, 0.99)  
p=0.04 | NR | NR |
| Martino 2004 Martino 2005 Ensrud 2006 Grady 2010 (MORE/CORE) 8 yr Medium | Stroke  
Raloxifene: 49/2725 (1.8%)  
Placebo: 19/1286 (1.5%)  
HR=1.22 (95% CI 0.72, 2.08)  
p=0.45  
PE  
Raloxifene: 17/2725 (0.6%)  
Placebo: 2/1286 (0.2%)  
p=0.048  
DVT  
Raloxifene: 31/2725 (1.1%)  
Placebo: 10/1286 (0.8%)  
p=0.32  
Hot flashes  
Raloxifene: 342/2725 (12.6%)  
Placebo: 89/1286 (6.9%)  
p<0.001  
Mortality  
Raloxifene: 47/2725 (1.7%)  
Placebo: 29/1286 (2.3%)  
p=0.27 | NR | NR |
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hormone therapy vs control</td>
<td>Wimalawansa 4 yr Medium</td>
<td>Withdrawals due to estrogen-related AEs  HRT: 3/18 (17%) Control: 0/18  Withdrawals due to inability to tolerate medications  HRT: 0/18 Control: 1/18 (6%)  Withdrawals due to other medical problems  HRT: 0/18 Control: 2/18 (11%)</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Hormone therapy (estrogen + progestin) vs placebo</td>
<td>Cauley 2003 6 yr Low</td>
<td>Global index 15% increase in hormone therapy group vs placebo group indicating more harm than benefit  Note this outcome is for the total study population, which includes women who are neither osteoporotic nor osteopenic</td>
<td>NR</td>
<td>Predictor: Fracture risk  Outcome: Global index  Lowest tertile fracture risk: HR 1.20 (95% CI 0.93, 1.58) Medium tertile fracture risk: HR 1.23 (95% CI 1.04, 1.46) Highest tertile fracture risk: HR 1.03 (95% CI 0.88, 1.24) p for interaction= 0.54  Note this outcome is for the total study population, which includes women who are neither osteoporotic nor osteopenic</td>
</tr>
<tr>
<td>Hormone therapy (estrogen) vs placebo</td>
<td>Jackson 2006 7 yr Low</td>
<td>NR</td>
<td>NR</td>
<td>Predictor: Fracture risk  Outcome: Global index  Lowest tertile fracture risk: HR 0.81 (95% CI 0.62, 1.05) Medium tertile fracture risk: HR 1.09 (95% CI 0.92, 1.30) Highest tertile fracture risk: HR 1.04 (95% CI 0.88, 1.23) p for interaction= 0.42  Note this outcome is for the total study population, which includes women who are neither osteoporotic nor osteopenic</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Osteoporosis drugs vs active comparator</td>
<td>Vestergaard 2012 3.8 yr Medium</td>
<td>NR</td>
<td>Inflammatory jaw event (proxy for ONJ) Any osteoporosis drug: 33/103,562 (0.03%) No osteoporosis drug: 37/310,683 (0.01%) RR=2.68 (95% CI 1.70, 4.20) p&lt;0.01 Alendronate: 15/55,090 (0.03%) No osteoporosis drug: 13/310,683 (0.004%) RR=3.46 (95% CI 1.72, 6.95) p&lt;0.01 HR=3.15 (95% CI 1.44, 6.87) Cox proportional hazards Adjusted for DMs, Sjögren’s syndrome, chemotherapy, irradiation, use of systemic corticosteroids, alcoholism, and jaw events before start of drug Raloxifene: 0/4,831 (0%) No osteoporosis drug: 2/310,683 (0.0006%) RR=0 (unadjusted) p=0.41</td>
<td>NR</td>
</tr>
<tr>
<td>Vestergaard 2011</td>
<td>NR</td>
<td>Defined daily dose per day (proportion of usual dose)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----</td>
<td>-----------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.8 yr Medium</td>
<td></td>
<td>Subtrochanteric fracture (proxy for AFF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Alendronate**

- 96/55,090
- No osteoporosis drug
- HR = 2.41 (1.78, 3.27)
- 96/165,270

- 0.66 ≤ HR = 2.93 (2.03, 4.22)
- 0.66 ≤ 0.99 HR = 3.44 (2.37, 5.00)*
- ≥ 1 HR = 0.49 (0.20, 1.22)

- **p for trend** = 0.01

**Raloxifene**

- 5/4,831
- No osteoporosis drug
- RR = 1.06 (0.34, 3.32)
- 9/14,493

- ≤ 0.35 NA
- 0.35 ≤ 0.99 HR = 1.99 (0.58, 6.85)
- ≥ 1 HR = 0.82 (0.10, 6.74)

- **p for trend** = 0.48

**Femoral shaft fracture (proxy for AFF)**

**Alendronate**

- 65/55,090
- No osteoporosis drug
- HR = 2.90 (1.97, 4.26)
- 52/165,270

- ≤ 0.66 HR = 3.69 (2.43, 5.83)
- 0.66 ≤ 0.99 HR = 3.60 (2.21, 5.88)
- ≥ 1 HR = 1.03 (0.44, 2.41)

- **p for trend** = 0.01

**Raloxifene**

- 3/4,831
- No osteoporosis drug
- HR = 0.82 (0.21, 3.20)
- 8/14,493

- ≤ 0.35 HR = 0.90 (0.11, 7.25)
- 0.35 ≤ 0.99 HR = 0.58 (0.07, 4.82)
- ≥ 1 HR = 1.21 (0.15, 9.94)

- **p for trend** = 0.84
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class-specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Risedronate</td>
<td>0/1,452 No osteoporosis drug</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0/4,356 HR undefined</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Harms Drug class-specific harms</td>
<td>Rare Harms AFF, ONJ, Afib</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
<td>--------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Vestergaard 2010 3.8 yr Medium</td>
<td>NR</td>
<td>AFib or atrial flutter Alendronate: 729/55,090 (1.3%) No osteoporosis drug: 1,635/310,683 (0.53%) HR=1.05 (95% CI 0.96, 1.15) Raloxifene: 55/4,831 (1.1%) No osteoporosis drug: 168/310,683 (0.05%) HR=0.82 (95% CI 0.60, 1.13) Risedronate: 0/1,452 (0%) No osteoporosis drug: 0/310,683 (0%) HR undefined Zolendronate: 0/22 (0%) No osteoporosis drug: 0/310,683 (0%) HR undefined Ibandronate: 0/612 (0%) No osteoporosis drug: 0/310,683 (0%) HR undefined Teriparatide: 0/303 (0%) No osteoporosis drug: 0/310,683 (0%) HR undefined Cox proportional hazards HRs adjusted for AFib before starting drug, heart valve disease, CHF, hyperthyroidism, diuretics, cardiovascuclar drugs, COPD, drugs treating COPD, alcoholism</td>
<td>NR</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** AE=Adverse effect; Afib=atrial fibrillation; AFF=atypical femoral fracture; ALN=alendronate; aHR=adjusted hazard ratio; aOR=adjusted odds ratio; BMI=mody mass index; CHF=congestive heart failure; CI=confidence interval; COPD=chronic obstructive pulmonary disease; CORE=Continuing Outcomes Relevant to Evista trial;
Table D5. Key Questions 5 and 6 evidence overview

<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Final Outcomes Incident Clinical Fracture</th>
<th>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs Discontinue Alendronate (Placebo)</td>
<td>Black 2006 (FLEX: RCT extension with rerandomization and discontinuation) 5-10 yr Low</td>
<td>Incidence n (%), RR (95% CI) adjusted for clinic, and high vs low fx risk stratum: Pooled ALN (n=662), Placebo (n=437): Any CFx Pooled ALN: 132/662 (19.9%) Discontinue/Placebo: 93/437 (21.3%) ARR=0.93 (95% CI 0.71, 1.21) p=NR but NS Clinical nonVFx Pooled ALN: 125/662 (18.9%) Discontinue/Placebo: 83/437 (19.0%) ARR=1.00 (95% CI 0.76, 1.32) p=NR but NS Clinical VFx Pooled ALN: 16/662 (2.4%) Discontinue/Placebo: 23/437 (5.3%) ARR=0.45 (95% CI 0.24, 0.85) p= significant but NR Hip Fx Pooled ALN: 20/662 (3.0%) Discontinue/Placebo: 13/437 (3.0%) ARR=1.02 (95% CI 0.51, 2.10) p=NR Forearm Fx Pooled ALN: 31/662 (4.7%) Discontinue/Placebo: 19/437 (4.3%)</td>
<td>RVFx (lateral spine x-rays at 36 &amp; 60 mo.): Pooled ALN: 60/662 (9.8%) Discontinue/Placebo: 46/437 (11.3%) ARR=0.86 (95% CI 0.60, 1.22), p=NR but NS Change in BMD during FLEX, Mean % change (SE) N=1071 of 1099 with ≥1 hip BMD; pooled ALN groups: TH BMD Pooled ALN (n=643): -1.02% (0.18) Discontinue/Placebo (n=428): -3.38% (0.22) Mean difference=2.36% (95% CI 1.81%, 2.90%), p=0.001 FN BMD Pooled ALN (n=643): 0.46% (0.24) Discontinue/Placebo (n=428): -1.48% (0.30) Mean difference=1.94 (95% CI 1.20, 2.68), p&lt;0.001 LS BMD N=1023 (per group NR) Pooled ALN: 5.26% (0.24) Discontinue/Placebo: 1.52% (0.29) Mean difference=3.74 (95% CI 3.03,4.45), p&lt;0.001</td>
<td>Clinical VFx, unadjusted RR. Baseline FN BMD T score p value for interaction=0.72 ≤-2.5: denominator for stratum=322; within stratum group denominators NR: Pooled ALN: 9/NR (4.7%) Discontinue/Placebo: 11/NR (8.3%) RR=0.57 (95% CI 0.23, 1.40) &gt;-2.5 to ≤-2.0: denominator for stratum=311; within stratum group denominators NR: Pooled ALN: 3/NR (1.6%) Discontinue/Placebo: 9/NR (7.1%) RR=0.22 (95% CI 0.05, 0.74) &gt;-2.0: denominator for stratum=461; within stratum group denominators NR: Pooled ALN: 4/NR (1.4%) Discontinue/Placebo: 3/NR (1.7%) RR=0.84 (95% CI 0.18, 4.2) Baseline VFx p value for interaction=0.86: Yes baseline VFx: denominator for stratum = 376; within stratum group denominators NR: Pooled ALN: 9/NR (4.0%) Discontinue/Placebo: 12/NR (8.0%) RR=0.47 (95% CI 0.19, 1.1) No baseline VFx denominator for stratum=723; within stratum group denominators NR: Pooled ALN: 7/NR (1.6%) Discontinue/Placebo: 11/NR (3.8%) RR=0.42 (95% CI 0.16, 1.1)</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Schwartz 2010 Post hoc FLEX analysis</td>
<td>NR</td>
<td>NR</td>
<td>Risk of first nonVFx in women without baseline VFx (n=720 of 1094*), by FLEX baseline T score. P value for interaction=</td>
<td></td>
</tr>
<tr>
<td><strong>ARR</strong>=1.09 (95% CI 0.62, 1.96) p=NR</td>
<td></td>
<td></td>
<td></td>
<td>NonVFxs, unadjusted RR. <em>Baseline FN BMD T score</em> p value for interaction = 0.40 ≤-2.5: denominator for stratum=322; within stratum group denominators NR: Pooled ALN: 43/NR (22.6%) Discontinue/Placebo: 39/NR (29.5%) RR=0.77 (95% CI 0.50, 1.2) &gt;-2.5 to ≤-2.0: denominator for stratum=311; within stratum group denominators NR: Pooled ALN: 38/NR (20.5%) Discontinue/Placebo: 26/NR (20.6%) RR=0.77 (95% CI 0.50, 1.2) &gt;-2.0: denominator for stratum=461; within stratum group denominators NR: Pooled ALN: 42/NR (14.9%) Discontinue/Placebo: 18/NR (10.1%) RR=1.5 (95% CI 0.86, 2.6) <strong>Prevalent VFx</strong> p value for interaction=0.23: Yes baseline VFx: denominator for stratum=376; within stratum group denominators NR: Pooled ALN: 62/NR (27.7%) Discontinue/Placebo: 35/NR (23.3%) RR=1.20 (95% CI 0.80, 1.8) No baseline VFx denominator for stratum=723; within stratum group denominators NR: Pooled ALN: 61/NR (14.1%) Discontinue/Placebo: 35/NR (23.3%) RR=0.86 (95% CI 0.59, 1.3)</td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------</td>
<td>-------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------</td>
<td>---------------------------------</td>
</tr>
</tbody>
</table>
| 5-10 yr Medium  |                                     |                                           | 0.019:  
≤-2.5: denominator for stratum=184; within stratum group denominators NR:  
Pooled ALN: 16/NR (14.7%)  
Discontinue/Placebo: 21/NR (28.0%)  
RR=0.50 (95% CI 0.26, 0.96)  
> -2.5 to ≤-2.0: denominator for stratum=203; within stratum group denominators NR:  
Pooled ALN: 15/NR (12.4%)  
Discontinue/Placebo: 13/NR (15.9%)  
RR=0.79 (95% CI 0.37, 1.66)  
>-2.0: denominator for stratum=333; within stratum group denominators NR:  
Pooled ALN: 30/NR (14.8%)  
Discontinue/Placebo: 14/NR (10.8%)  
RR=1.41 (95% CI 0.75, 2.66)  
All other subgroup*subgroup interactions NS |
| Bone 2004 Prospective CCT 8-10 yr with discontinuation Medium | CFx: collected, NR  
First NonVFx 8-10 yr. N=247  
ALN 10 mg: 8.1%  
ALN 5 mg: 11.5%  
ALN 20/ALN 5/placebo: 12.0%  
p=NR but NS | RVFx 6-10 yr (n=228)  
ALN 10 mg: 5.0%  
ALN 5 mg: 13.9%  
ALN 20/ALN 5/placebo: 6.6%  
p=NR but NS  
Mean % change in BMD 8-10 yr. Reports denominator range for 3 groups together, each comparison:  
LS BMD  
n=71-81/group  
ALN 10 mg: 2.3% (95% CI 1.4, 3.1)  
p<0.001  
ALN 5 mg: 1.2% (95% CI 0.2, 2.1) | NR |
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Final Outcomes</th>
<th>Intermediate Outcomes</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Incident Clinical Fracture</td>
<td>Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p&lt;0.05 Placebo /discontinue: 0.2% (9% CI -0.7, 1.1) NS</td>
<td>FN BMD n=71-76/group ALN 10 mg: 1.0% (95% CI -0.3, 2.4) NS ALN 5 mg: 0.3% (95% CI -1.2, 1.7) NS Placebo /discontinue: -1.7% (95% CI -3.0, -0.3) p&lt;0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ALN 10 mg: 1.0% (95% CI -0.3, 2.4) NS ALN 5 mg: 0.3% (95% CI -1.2, 1.7) NS Placebo /discontinue: -1.7% (95% CI -3.0, -0.3) p&lt;0.05</td>
<td>TH BMD n=46-50/group ALN 10 mg: 0.1% (95% CI -1.1, 1.3) NS ALN 5 mg: -0.2% (-1.4, 1.0) NS Placebo/discontinue: -1.6 (95% CI -2.8, -0.4) p&lt;0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ALN 10 mg: 8/122 (6.6%) ALN 5 mg: 8/113 (7.1%) ALN 20/ALN 5/placebo: 9/115 (7.8%)</td>
<td>Denominators: ALN 10 mg (n=122), ALN 5 mg (n=113), ALN 20/ALN 5/placebo (n=115)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NonVFx Symptomatic VFx 6-7 yr (p= not calculated by authors) ALN 10 mg: 8/122 (6.6%) ALN 5 mg: 7/113 (6.2%) ALN 20/ALN 5/placebo: 8/115 (7.0%)</td>
<td>LS BMD 6-7 yr: p=NR (all 3) ALN 10 mg: +1.6% ALN 5 mg: +1.5% ALN 20/ALN 5/placebo: +0.2%</td>
<td></td>
</tr>
<tr>
<td>Tonino 2000</td>
<td>6-7 yr with discontinuation Medium</td>
<td></td>
<td>FN BMD 6-7 yr: ALN 10 mg: +0.5%, p=NS ALN 5 mg: +0.3%, p=NS ALN 20/ALN 5/placebo: -0.5%</td>
<td></td>
</tr>
<tr>
<td>Drug Comparison</td>
<td>Study (Trial) Followup Risk of Bias</td>
<td>Final Outcomes Incident Clinical Fracture</td>
<td>Intermediate Outcomes Incident Radiographic Vertebral Fracture DXA BMD Change</td>
<td>Predictors (Patient, Bone, Drug)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Zoledronic Acid vs Zoledronic Acid then placebo</td>
<td>McClung, 2009 2 yr Low</td>
<td>CFx: ZOL 5mg/yr x1 year (discontinuation): 4/181 (2%) ZOL 5mg/yr x2 year: 6/198 (3%) RR 0.73 (0.21, 2.54) *</td>
<td>% change LS BMD 2 yr: ZOL 5mg/yr x1 year (discontinuation): +4.4% ZOL 5mg/yr x2 year: +5.2% MD=-0.76 (-0.82, -0.70) *</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>*Calculated</td>
<td>% change TH BMD 2 yr: ZOL 5mg/yr x1 year (discontinuation): +2.3% ZOL 5mg/yr x2 year: +2.9% MD=-0.63 (-0.67, -0.59) *</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>% change FN BMD 2 yr: ZOL 5mg/yr x1 year (discontinuation): +1.6% ZOL 5mg/yr x2 year: +2.2% MD=-0.56 (-0.62, -0.50) *</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*Calculated</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** ALN=alendronate; ARR=absolute risk reduction; BMD=bone mineral density; BP=bisphosphonates; CI=confidence interval; DXA=Dual-energy x-ray absorptiometry; FIT=Fracture Intervention Trial; FLEX=Fracture Intervention Trial Long Term Extension; FN=femoral neck; Fx=fracture; LS=lumbar spine; mg=milligrams; N=number; NR=not reported; NS=not significant; RCT=randomized controlled trial; RR=risk ratio; SE=standard error; TH=total hip; VFx=vertebral fracture; yr=year
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial)</th>
<th>Followup</th>
<th>Risk of Bias</th>
<th>Harms</th>
<th>Drug class specific harms</th>
<th>Rare Harms</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate vs Discontinue Alendronate (Placebo)</td>
<td>Black 2006 (FLEX)</td>
<td>5 yr: 5-10 year FIT extension Low</td>
<td>Low</td>
<td>Serious AEs, discontinuations due to AEs, rates of death, UGI or serious UGI AEs</td>
<td>Data not shown. Pooled ALN vs. discontinue/placebo p=not significant</td>
<td>None</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Black 2010 (FLEX)</td>
<td>5 yr: 5-10 year FIT extension Low</td>
<td>NR</td>
<td></td>
<td></td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Bone 2004 8-10 yr CCT extension with discontinuation</td>
<td>Medium (subset of Tonino 2000)</td>
<td></td>
<td>Medium</td>
<td>Any serious clinical event: ALN 10 mg: 18/86 (20.9%) ALN 5 mg: 25/78 (32.1%) ALN 20/ALN 5/placebo: 18/83 (21.7%)</td>
<td></td>
<td>Reported &quot;no insufficiency fractures&quot;</td>
<td>8 -10 yr: 30-36% of women in each group used aspirin and 41-53% used nonsteroidal or glucocorticoid drugs. No apparent adverse interaction between these drugs and alendronate.</td>
</tr>
<tr>
<td></td>
<td>Ensrud 2004 (FLEX 3 year interim analysis)</td>
<td></td>
<td>Low</td>
<td>Any UGI event(s)</td>
<td>Pooled ALN: 197/662 (29.8%) Discontinue/placebo: 156/437(35.7%)</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Black 2006 3 yr</td>
<td></td>
<td></td>
<td>Death Pooled ALN: 16/662 (2.4%) Discontinue/placebo: 8/437 (1.8%)</td>
<td></td>
<td></td>
<td>NR</td>
</tr>
</tbody>
</table>

Table D6. Key Questions 7 and 8 evidence overview
<table>
<thead>
<tr>
<th>Drug Comparison</th>
<th>Study (Trial) Followup Risk of Bias</th>
<th>Harms Drug class specific harms</th>
<th>Rare Harms AFF, ONJ, Afib</th>
<th>Predictors (Patient, Bone, Drug)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tonino 2000 6-7 yr CCT extension with discontinuation</td>
<td>Medium</td>
<td>Any UGI AE, (≥1, 6-7 year) ALN 10 mg: 21/122 (17.2%) ALN 5 mg: 18/113 (15.9%) ALN 20/ALN 5/placebo: 21/115 (18.3%) p=NR, but “similar”</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Bisphosphonate drug holiday (≥1 yr without use) vs Bisphosphonate use ≥3 yr</td>
<td>Adams 2018 Mean 4 yr</td>
<td>Any AE (≥1), 6-7 yr ALN 10 mg: 111/122 (91.0%) ALN 5 mg: 96/113 (85.0%) ALN 20/ALN 5/placebo: 104/115 (90.4%) Any serious AE (≥1): ALN 10 mg: 15/122 (12.3%) ALN 5 mg: 13/113 (11.5%) ALN 20/ALN 5/placebo: 13/115 (11.3%) p=NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Zoledronic Acid vs Zoledronic Acid then placebo</td>
<td>McClung, 2009 2 yr Low</td>
<td>Serious AEs ZOL 5mg/yr x1 year (discontinuation): 17/181 (9%) ZOL 5mg/yr x2 year: 21/198 (11%) RR 0.89 (0.48, 1.62) * Death ZOL 5mg/yr x1 year (discontinuation): 0/181 ZOL 5mg/yr x2 year: 1/198 (&lt;1%) *Calculated</td>
<td>Afib 0 ONJ 0</td>
<td>NR</td>
</tr>
</tbody>
</table>

**Abbreviations:** Afib=atrial fibrillation; AFF=atypical femoral fracture; AE=adverse event; ALN=alendronate; mg=milligram; BP=bisphosphonates; NR=not reported; ONJ=osteonecrosis of the jaw; RCT=randomized controlled trial; UGI=upper gastrointestinal
<table>
<thead>
<tr>
<th>Intervention</th>
<th>Study, Year Design (Trial) Location Funding</th>
<th>Study Demographics Comorbidities</th>
<th>Duration of Treatment Drug Holiday?</th>
<th>Outcomes Reported</th>
<th>Patient, Bone, Drug Predictors Reported?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alendronate</td>
<td>Abrahamsen, 2010 Retrospective cohort, administrative data Denmark Industry</td>
<td>Male and female incident alendronate users and untreated matched controls</td>
<td>NR 2 months 25% 1.1 yr 25% 3.7 yr 25% 8.7 yr 25% 3.4 yr (mean followup time) No</td>
<td>Incident hip fractures Adverse events: incident fractures of the subtrochanteric femur or femoral shaft</td>
<td>Sex, treatment duration</td>
</tr>
<tr>
<td></td>
<td>Abrahamsen, 2016 Case control, administrative data Denmark Industry</td>
<td>Male and female incident alendronate users and untreated matched controls aged 50-94</td>
<td>NR &lt; 5 yr 76% 5-10 yr 21% ≥10 yr 3% 6.9 yr (median ascertainment) No</td>
<td>Incident hip fractures Adverse events: incident subtrochanteric femur or femoral shaft fractures</td>
<td>Treatment duration</td>
</tr>
<tr>
<td></td>
<td>Chiu, 2018 Retrospective cohort, administrative data Taiwan Government</td>
<td>Women aged ≥50 or men aged ≥60 using alendronate ≥30 days and unmatched controls taking raloxifene ≥30 days</td>
<td>1-8 yr No</td>
<td>Adverse events: ONJ</td>
<td>Treatment duration</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study, Year Design (Trial) Location Funding</td>
<td>Study Demographics Comorbidities</td>
<td>Duration of Treatment Drug Holiday?</td>
<td>Outcomes Reported</td>
<td>Patient, Bone, Drug Predictors Reported?</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Chapurlat, 2005 RCT (FIT)</td>
<td>US Industry</td>
<td>Postmenopausal women aged 55 to 80 with femoral neck T-score ≤ −1.6 who adhered to treatment but lost BMD during first yr of treatment DM NR CVD NR CKD NR BMD, mean, g/cm² Hip 0.700 LS 0.826 Fracture since age 45: 41% Fracture risk NR</td>
<td>3.2 yr (mean) No</td>
<td>Radiographic vertebral fractures</td>
<td>BMD change during 1 or 2 yr of treatment</td>
</tr>
<tr>
<td>Nevitt, 1999 RCT (FIT)</td>
<td>US Industry</td>
<td>Postmenopausal (at least 2 yr) women aged 55-80 with FN BMD ≤0.68 g/cm² (T-score ≤ −1.6) who had lateral spine radiographs and baseline and one followup visit DM NR CVD NR CKD NR BMD, g/cm² LS 0.83 FN 0.59 TH 0.69 Vertebral fracture NR Fracture risk NR</td>
<td>3.8 No</td>
<td>Incident radiographic vertebral fractures Bone density as a predictor of fractures</td>
<td>Fracture history BMD (LS, FN, TH)</td>
</tr>
<tr>
<td>Wang, 2016 Retrospective cohort</td>
<td>Taiwan University</td>
<td>Women aged ≥50 with osteoporosis Comorbidities NR BMD NR Fracture during 1 yr prior to study 20%</td>
<td>5 yr No</td>
<td>Adverse events: AFF</td>
<td>Treatment duration</td>
</tr>
<tr>
<td>Zoledronic Acid Cosman, 2014</td>
<td>RCT (HORIZON extension)</td>
<td>Postmenopausal women aged 65 to 89 yr with FN T-score ≤ −2.5, or FN T-score ≤ −1.5, with ≥ 2 mild radiologic</td>
<td>3 yr Yes</td>
<td>Clinical nonvertebral fractures, radiographic vertebral fractures</td>
<td>Incident radiographic vertebral</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study, Year Design (Trial) Location Funding</td>
<td>Study Demographics</td>
<td>Duration of Treatment Drug Holiday?</td>
<td>Outcomes Reported</td>
<td>Patient, Bone, Drug Predictors Reported?</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------</td>
<td>--------------------</td>
<td>--------------------------------------</td>
<td>-------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>US Industry</td>
<td></td>
<td>vertebral fractures or one moderate vertebral fracture; who had received zoledronate for 3 yr in the core study Comorbidities at start of extension study DM NR CVD NR CKD NR BMD FN T-score ≤ -2.5: 55% TH T-score ≤ -2.5: 26% Spine fracture 61% Hip fracture NR Fracture risk NR</td>
<td>5 yr No</td>
<td>Vertebral fractures, nonvertebral fractures, hip fractures, BMD changes</td>
<td>None</td>
</tr>
<tr>
<td>Risedronate</td>
<td>Sorensen, 2002 RCT (VERT) Multinational Industry</td>
<td>Postmenopausal women aged ≤85 yr with ≥2 vertebral fractures; compliant in core study, no new medications known to affect bone metabolism DM NR CVD NR CKD NR BMD, g/cm² LS 0.80 FN 0.64 Number of vertebral fractures 4.4 Fracture risk NR</td>
<td>5 yr No</td>
<td>Vertebral fractures, nonvertebral fractures, hip fractures, BMD changes Adverse events: serious adverse events, GI intolerance</td>
<td>None</td>
</tr>
<tr>
<td>Ste-Marie, 2004 RCT (VERT) Multinational Industry</td>
<td>Postmenopausal women with 2 prevalent vertebral fractures or 1 prevalent vertebral fracture and LS T-score ≤ -2; completed core study and underwent 3 yr iliac crest biopsy DM NR CVD NR CKD NR</td>
<td>5 yr No</td>
<td>Adverse events: vertebral fractures, nonvertebral fractures, BMD change</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>Study, Year Design (Trial) Location Funding</td>
<td>Study Demographics Comorbidities</td>
<td>Duration of Treatment Drug Holiday?</td>
<td>Outcomes Reported</td>
<td>Patient, Bone, Drug Predictors Reported?</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------</td>
<td>------------------------------------</td>
<td>------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Ibandronate</td>
<td>Miller, 2011 RCT (MOBILE long-term extension) Multinational Industry</td>
<td>Postmenopausal women aged 55 to 80 yr with LS T-score &lt; -2.5 and ≥ -5.0, ≥75% compliant in core study</td>
<td>5 yr No</td>
<td>Adverse events: upper GI intolerance, musculoskeletal pain, serious adverse events</td>
<td>Dose</td>
</tr>
<tr>
<td>Bisphosphonates (any)</td>
<td>Meier, 2012 Case control, administrative data US Funding NR</td>
<td>Women and men aged ≥50 yr admitted to a trauma center with a subtrochanteric or femoral shaft fracture</td>
<td>NR None 87% &lt; 2 yr 3% 2-5 yr 4% 5-9 yr 4% ≥ 9 yr 2% 12 yr ascertainment No</td>
<td>Adverse events: AFF</td>
<td>Treatment duration</td>
</tr>
<tr>
<td>Bisphosphonates (any)</td>
<td>Lee, 2018 Retrospective cohort, administrative data Korea Foundation</td>
<td>Women and men ≥45 yr without history of femoral surgery who took bisphosphonate at study hospital</td>
<td>4.5 yr (mean in AFF patients) 10 yr ascertainment No</td>
<td>Adverse events: AFF</td>
<td>Treatment duration</td>
</tr>
<tr>
<td>Intervention</td>
<td>Study, Year Design (Trial) Location Funding</td>
<td>Study Demographics Comorbidities</td>
<td>Duration of Treatment Drug Holiday?</td>
<td>Outcomes Reported</td>
<td>Patient, Bone, Drug Predictors Reported?</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------------</td>
<td>-------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Denosumab Teriparatide</td>
<td>Leder, 2015 RCT (DATA-Switch) US Industry</td>
<td>Postmenopausal women aged ≥45 with T-score ≤ -2.5 at the spine, hip, or FN; T-score ≤ -2.0 with ≥1 BMD-independent risk factor; or T-score ≤ -1.0 with history of fragility fracture DM NR CVD NR Previous MI NR CKD NR BMD, g/cm² Posterior-anterior spine 0.84 FN 0.64 TH 0.75 Clinical fracture age &gt;45 42%</td>
<td>4 yr No</td>
<td>Adverse events: hypercalcemia, serious adverse events</td>
<td>None</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; BMD=bone mineral density; CKD=chronic kidney disease; CVD=cardiovascular disease; DM=diabetes mellitus; FIT=Fracture Intervention Trial; FN=femoral neck; HORIZON=Health Outcomes and Reduced Incidence with Zoledronic Acid Once Yearly; LS=lumbar spine; MOF=major osteoporotic fracture; NR=not reported; ONJ=osteonecrosis of the jaw; RCT=randomized controlled trial; TH=total hip; yr=years
Strength of Evidence Tables

- **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.

Alendronate

Table D8. Strength of evidence assessments: Long-term treatment with alendronate versus placebo from RCTs and CCTs

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Population</th>
<th>Summary statistics [95% CI]</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Components</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture 0-4 yrs</td>
<td>1 (584/4432) PM women with FN BMD T-score ≤-1.6 and no baseline RVF</td>
<td>Cummings 1998 NS difference in risk with ALN vs. PBO: ALN 12.3%, PBO 14.1%; HR = 0.86 [0.73, 1.01]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>1 (266/1631*) PM women with FN BMD T-score &lt; -2.5 and no baseline RVF</td>
<td>Cummings 1998 Lower risk with ALN vs. PBO: ALN 13.1%, PBO 19.6%; HR = 0.64 [0.50, 0.82]</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis. Interaction p value 0.01.</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>1 (179/1436) PM women with FN BMD T-score -2.5 to</td>
<td>Cummings 1998</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis.</td>
<td>Low</td>
</tr>
<tr>
<td>Outcome</td>
<td>Time frame</td>
<td># Studies (n/N events/participants evaluated) Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>Optional Components</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>------------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 and no baseline RVF</td>
<td>NS difference in risk with ALN vs. PBO: ALN 12.7%, PBO 12.3%; HR = 1.03 [0.77, 1.39]</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis. Interaction p value 0.01.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 (139/1315) PM women with FN BMD T-score -2 to -1.6 and no baseline RVF</td>
<td>Cummings 1998 NS difference in risk with ALN vs. PBO: ALN 10.9%, PBO 9.5%; HR = 1.14 [0.82, 1.60]</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
</tr>
<tr>
<td>Incident nonvertebral fracture 0-4 yrs</td>
<td></td>
<td>1 (555/4432) PM women with FN BMD T-score ≤-1.6 and no baseline RVF</td>
<td>Cummings 1998 NS difference in risk with ALN vs PBO: ALN 11.8%, PBO 13.3%; HR = 0.88 [0.74, 1.04]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
</tr>
<tr>
<td>Incident hip fracture 0-4 yrs</td>
<td></td>
<td>1 (43/4432) PM women with FN BMD T-score ≤-1.6 and no baseline RVF</td>
<td>Cummings 1998 NS difference in risk with ALN vs. PBO: ALN 0.9%, PBO 1.1%; HR = 0.79 [0.43, 1.44]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td></td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Outcome</td>
<td>Time frame</td>
<td># Studies (n/N events/participants evaluated)</td>
<td>Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>---------------------------------------------</td>
<td>------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture† 0-4 yrs</td>
<td>1 (121/4134) PM women with FN BMD T-score ≤-1.6 and no baseline RVF</td>
<td>Cummings 1998 Lower risk with ALN vs. PBO: ALN 2.1%, PBO 3.8%; HR = 0.56 [0.39, 0.80]</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td>1 (66/1631) PM women with FN BMD T-score ≤-2.5 and no baseline RVF</td>
<td>Cummings 1998 Lower risk with ALN vs. PBO: ALN 2.9%, PBO 5.8%; HR = 0.50 [0.31, 0.82]</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis. Interaction p value NR.</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>1 (37/1436) PM women with FN BMD T-score -2.5 to -2 and no baseline RVF</td>
<td>Cummings 1998 Lower risk with ALN vs. PBO: ALN 1.9%, PBO 3.6%; HR = 0.54 [0.28, 1.04]</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis. Interaction p value NR.</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>1 (18/1315) PM women with FN BMD T-score -2 to -1.6 and no baseline RVF</td>
<td>Cummings 1998 Lower risk with ALN vs. PBO: ALN 1.3%, PBO 1.5%;</td>
<td>Medium (subgroup)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>A priori subgroup analysis. Interaction p value NR.</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome Time frame</td>
<td># Studies (n/N events/participants evaluated) Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>Optional Components</td>
<td>SOE</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------------</td>
<td>--------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>-----</td>
</tr>
<tr>
<td>Serious adverse events 0-4 yrs</td>
<td>1 (4432) PM women with FN-BMD T-score ≤-1.6 and no baseline RVF</td>
<td>HR = 0.82 [0.33, 2.07]</td>
<td>Cummings 1998 AE causing hospitalization NS difference with ALN vs. PBO: ALN 29.1%, PBO 26.9%; HR = 1.09 [0.98, 1.22]</td>
<td>Low</td>
<td>Indirect</td>
<td>Precise</td>
<td>Unknown but multisite trial</td>
<td>Possible (serious AE outcome NR)</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>1 (4432) PM women with FN-BMD T-score ≤-1.6 and no baseline RVF</td>
<td>HR = 0.92 [0.59, 1.45]</td>
<td>Cummings 1998 Mortality NS difference with ALN vs. PBO: ALN 1.7%, PBO 1.8%; HR = 0.92 [0.59, 1.45]</td>
<td>Low</td>
<td>Indirect</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
</tr>
<tr>
<td>ST/FS fracture (rare x-ray review for confirmation of AFF features) 3-4.5 yrs</td>
<td>1 (6459) PM women with FN-BMD T-score ≤-1.6</td>
<td>HR= 1.03 [0.06, 16.46]</td>
<td>Black 2010 NS difference with ALN vs. PBO: ALN 0.031% (n=1), PBO 0.031% (n=1) HR= 1.03 [0.06, 16.46]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>NA</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

**Abbreviations:** AE=adverse events; ALN=alendronate; BMD=bone mineral density; CI=confidence interval; FIT=Fracture Intervention Trial; FN=femoral neck; HR=hazard ratio; LS=lumbar spine; NA=not applicable; NR=not reported; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; SOE=strength of evidence; ST/FS=subtrochanteric/femoral shaft; TH=total hip; yr=years
*Calculated by EPC
†95% of survivors to the year 4 visit completed followup vertebral x-rays at that visit, but the percentage of completers is not reported by BMD subgroup, so the number in each BMD population subgroup is some unknown number likely slightly smaller than that listed.

Table D9. Strength of evidence assessments: Alendronate continuation versus discontinuation

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated) Population</th>
<th>Summary statistics [95% CI]</th>
<th>Study limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1 (225/1099) PM women previously received 5 yrs ALN for FN BMD T-score ≤ -1.6</td>
<td>Black 2006 NS difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs ALN + 5 yrs PBO (ALN5/PBO5): ALN10 19.9%, ALN5/PBO5 21.3%; RR = 0.93 [0.71, 1.21]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>1 (208/1099) PM women previously received 5 yrs ALN for FN BMD T-score ≤ -1.6</td>
<td>Black 2006 NS difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs ALN + 5 yrs PBO (ALN5/PBO5): ALN10 18.9%, ALN5/PBO5 19.0%; RR = 1.00 [0.76, 1.32]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>1 (25/350) PM women previously received 5 yrs ALN for LS BMD T-score ≤ -2.5</td>
<td>Tonino 2000 No apparent difference in risk with 7 yrs ALN (ALN7) vs. 5 yrs ALN + 2 yrs PBO (ALN5/PBO2): ALN7 5 mg/d 7.1%, ALN7 10 mg/d 6.6%, ALN5/PBO2 7.8% ALN7 vs. ALN5/PBO RR = 0.87 [0.40, 1.91]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>1 (26/247) PM women previously received 5 yrs ALN for LS BMD T-score ≤ -2.5</td>
<td>Bone 2004 No apparent difference in risk with 10 yrs ALN (ALN10) vs. 7 yrs ALN + 3 yrs PBO (ALN7/PBO3): ALN10 5 mg/d 11.5%, ALN10 10 mg/d 8.1%, ALN7/PBO3 12.0%</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated) Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>-------</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>1 (33/1099) PM women previously received 5 yrs ALN for FN BMD T-score ≤ 1.6 and no baseline RVF</td>
<td>ALN10 vs. ALN7/PBO RR = 0.81 [0.38, 1.71]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>1 (39/1099) PM women previously received 5 yrs ALN for FN BMD T-score ≤ 1.6</td>
<td>Black 2006 NS difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs ALN + 5 yrs PBO (ALN5/PBO5): ALN10 3.0%, ALN5/PBO5 3.0%; RR = 1.02 [0.51, 2.10]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>1 (23/350) PM women previously received 5 yrs ALN for LS BMD T-score ≤ 2.5</td>
<td>Tonino 2000 No apparent difference in risk with 7 yrs ALN (ALN7) vs. 5 yrs ALN + 2 yrs PBO (ALN5/PBO2): ALN7 5 mg/d 6.2%, ALN7 10 mg/d 6.6%, ALN5/PBO2 7.0% ALN7 vs ALN5/PBO RR = 0.92 [0.40, 2.10]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>1 (106/1099) PM women previously received</td>
<td>Black 2006 NS difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>fracture</td>
<td>5 yrs ALN for FN BMD T-score ≤-1.6</td>
<td>ALN + 5 yrs PBO (ALN5/PBO5): ALN10 9.8%, ALN5/PBO5 11.3%; RR = 0.86 [0.60, 1.22]</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>1 (NR/228) PM women previously received 5 yrs ALN for LS BMD T-score ≤-2.5</td>
<td>Bone 2004 No apparent difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs ALN + 5 yrs PBO (ALN5/PBO2): ALN7 5 mg/d 5%, ALN7 10 mg/d 6.6%, ALN5/PBO2 13.9%; For ALN10 vs. ALN5/P5: RR 1.40 [0.52, 3.74]</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>1 (NR/1099) PM women previously received 5 yrs ALN for FN-BMD T-score ≤-1.6</td>
<td>Black 2006 NS difference in risk of serious AEs, mortality, UGI AEs, or serious UGI AEs between ALN10 and ALN5/PBO5 (no numerical data provided)</td>
<td>Low</td>
<td>Direct</td>
<td>Unknown (no data)</td>
<td>Unknown but multisite trial</td>
<td>Detected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>1 (41/350) PM women previously received 5 yrs ALN for LS BMD T-score ≤-2.5</td>
<td>Tonino 2000 NS difference in risk with 7 yrs ALN (ALN7) vs. 5 yrs ALN + 2 yrs PBO (ALN5/PBO2): ALN7 5 mg/d 11.5%, ALN7 10 mg/d 12.3%, ALN5/PBO2 11.3% ALN7 vs ALN5/PBO RR = 1.05 [0.57, 1.96]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Detected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>1 (61/247) PM women previously received 5 yrs ALN for LS BMD T-score ≤-2.5</td>
<td>Bone 2004 NS difference in risk with 10 yrs ALN (ALN10) vs. 7 yrs ALN + 3 yrs PBO (ALN7/PBO3): ALN10 5 mg/d 32.1%, ALN10 5 mg/d 20.9%, ALN7/PBO3</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Detected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated) Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------------------</td>
<td>----------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>ST/FS fracture (rare x-ray review for confirmation of AFF features)</td>
<td>1 (3/1099) PM women previously received 5 yrs ALN for FN-BMD ≤-1.6</td>
<td>Black 2006 NS difference in risk with 10 yrs ALN (ALN10) vs. 5 yrs ALN + 5 yrs PBO (ALN5/PBO5): ALN10 0.030%, ALN5/PBO5 0.023% HR= 1.33 [0.12, 14.67]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>ST/FS fracture without x-ray review for confirmation of AFF features</td>
<td>1 (47/39502) Women aged ≥45 yr with ≥3 yr prior ≥50% adherent bisphosphonate use (99% alendronate)</td>
<td>Adams 2018 Increased risk with bisphosphonate (alendronate) continuation vs. discontinuation Pooled persistent and nonpersistent continuation groups 0.15% (44/28005) vs. discontinuation 0.03% (3/11497) OR=6.03 [1.87, 19.42]</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (0/1099) PM women previously received 5 yrs ALN for FN BMD T-score ≤-1.6</td>
<td>Black 2006 &quot;No cases of ONJ were observed.&quot;</td>
<td>Low</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations**: AE=adverse events; ALN=alendronate; BMD=bone mineral density; CF=clinical fracture; CI=confidence interval; FIT=Fracture Intervention Trial; FN=femoral neck; HR=hazard ratio; LS=lumbar spine; NA=not applicable; NR=not reported; ONJ=osteonecrosis of the jaw; PBO=placebo; PM=postmenopausal; RCT=randomized controlled trial; RR=risk ratio; SOE=strength of evidence; TH=total hip; UGI=upper gastrointestinal; yr=years
<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated) Study description</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Components</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated)</td>
<td>Study description</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>Optional Components</td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>ST/FS fracture without x-ray confirmation of AFF features</td>
<td>1 (309/220,360) Retrospective cohort. AFF defined by diagnosis codes only. ALN treatment duration range 0-11 yrs (mean ~3.8)</td>
<td>Vestergaard 2012 Higher risk with ALN vs. no osteoporosis drug treatment Subtrochanteric fracture: ALN 0.017%, No treatment 0.006%; HR=2.41 [1.78, 3.27] Femoral shaft fracture: ALN 0.012%, No treatment 0.003%; HR=2.90 [1.97, 4.26]</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>NA</td>
</tr>
<tr>
<td>1 (5/534) Retrospective cohort, AFF defined by diagnosis codes only. Subgroup with ALN treatment duration &gt;6 years</td>
<td>Abrahamsen 2009 No difference in risk with ALN vs. no bisphosphonate treatment ALN % NR, No bisphosphonate % NR; HR=1.37 [0.22, 8.62]</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
</tbody>
</table>
### Outcome

**Osteonecrosis of the jaw**

<table>
<thead>
<tr>
<th># Studies</th>
<th>Study description</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Components</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (28/220,360) retrospective cohort. ONJ defined by diagnosis codes only. ALN treatment duration range 0-11 yrs (mean ~3.8)</td>
<td>Vestergaard 2012 Higher risk with ALN vs no treatment Any jaw event: ALN 0.03% vs No treatment 0.01%; HR=3.15 [1.44, 6.87]</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>Low</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; AE=adverse events; ALN=alendronate; CI=confidence interval; HR=hazard ratio; ICF=incident clinical fracture; LS=lumbar spine; MOF=major osteoporotic fracture; NA=not applicable; NR=not reported; ONJ=osteonecrosis of the jaw; SOE=strength of evidence; ST/FS=subtrochanteric/femoral shaft; yr=years

### Table D11. Strength of evidence assessments: Long-term alendronate treatment versus raloxifene or calcitonin or versus raloxifene alone

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Components</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated) Study description</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>Optional Components</td>
<td>SOE</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0 NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0 NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0 NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0 NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (353/8,354) ONJ defined by diagnosis codes, radiographic and pathological confirmation. ALN duration range 0-11 yrs (mean ~4 yr in ONJ group).</td>
<td>Chiu 2014 Higher risk with ALN vs. raloxifene: HR=7.42 [1.02, 54.09]</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>1 (37/43,645) ONJ defined only by diagnosis codes. ALN duration up to 6 yrs.</td>
<td>Lin 2014 No difference in risk with ALN vs. raloxifene or calcitonin: ALN 0.15% vs. raloxifene/calcitonin 0.08%; HR=0.86 [0.44, 1.69]</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>
Abbreviations: AE=adverse events; ALN=alendronate; CI=confidence interval; HR=hazard ratio; MOF=major osteoporotic fracture; NA=not applicable; NR=not reported; ONJ=osteonecrosis of the jaw; SOE=strength of evidence; yr=years
## Zoledronic Acid

### Table D12. Strength of evidence assessments: Long-term zoledronic acid treatment (6 yr) versus placebo

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1(377/2000)</td>
<td>Reid 2018 Lower risk with zoledronic acid vs. placebo: ZOL 16% vs. placebo 21% HR=0.73 [0.60, 0.90]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>1 (249/2000)</td>
<td>Reid 2018 Lower risk with zoledronic acid vs. placebo ZOL 10% vs placebo 15% HR=0.66 [0.51, 0.85]</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>High</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>1 (20/2000)</td>
<td>Reid 2018 ND in risk with zoledronic acid vs. placebo ZOL 0.8% vs placebo 1.2% HR=0.66 [0.27, 1.16]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>1 (48/2000)</td>
<td>Reid 2018 Lower risk with zoledronic acid vs. placebo: ZOL 1.4% vs placebo 3.4% HR=0.41 [0.22, 0.75]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>NR</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated)</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>1 (843/2000)</td>
<td>Reid 2018 ND in risk with zoledronic acid vs. placebo: ZOL 40% vs placebo 44%; OR=0.84 [0.70, 1.00]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (0/2000)</td>
<td>Reid 2018 No cases occured</td>
<td>Low</td>
<td>Direct</td>
<td>NA</td>
<td>Unknown,</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations:** CI=confidence interval; HR=hazard ratio; NA=not applicable; ND=no difference; SOE=strength of evidence; yr=years; ZOL=zoledronic acid

**Table D13. Strength of evidence assessments: zoledronic acid continuation (2 years zoledronic acid) versus discontinuation (1 year zoledronic acid followed by 1 year placebo)**

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1(10/379)</td>
<td>McClung 2009 NS difference in risk with zoledronic acid 2 yrs (Z2) vs. zoledronic acid 1 yr followed by placebo 1 yr (Z1/P1); Z2 3.0% vs. Z1/P1 2.2%; RR=1.37 [0.39, 4.78]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>
### Table D14. Strength of evidence assessments: zoledronic acid continuation (6 years zoledronic acid) versus discontinuation (3 years zoledronic acid followed by 3 years placebo)

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0 (NR)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>1 (38/379)</td>
<td>McClung 2009 NS difference in risk with Z2 vs. Z1/P1: Z2 9.4% vs. Z1/P1 10.6%; RR=0.91 [0.50, 1.67]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0 (NR)</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (0/379)</td>
<td>McClung 2009 No cases occurred</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations:** AE=adverse events; CI=confidence interval; NA=not applicable; NR=not reported; P=placebo; RR=risk ratio; SOE=strength of evidence; yr=years; Z=zoledronic acid
<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>1 (NR/1233)</td>
<td>Black 2012 NS difference in risk with Z6 vs. Z3/P3: HR=1.81 [0.53, 6.2]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Number of events not reported</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>1 (44/1233)</td>
<td>Black 2012 Lower risk with Z6 vs. Z3/P3: Z6 3.0% vs. Z3/P3 6.2%; OR=0.51 [0.26, 0.95]</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>1 (359/1233)</td>
<td>Black 2012 NS difference in risk with Z6 vs. Z3/P3: Z6 31% vs. Z3/P3 27%; RR=1.14 [0.96, 1.36]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>&quot;Atypical femoral fracture&quot; (not defined)</td>
<td>1 (0/1233)</td>
<td>Black 2012 No cases occurred</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (1/1233)</td>
<td>Black 2012 One case of ONJ in Z6 group, none in Z3/P3 group</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

Abbreviations: AE=adverse events; CI=confidence interval; HR=hazard ratio; NA=not applicable; NR=not reported; ONJ=osteonecrosis of the jaw; P=placebo; RR=risk ratio; SOE=strength of evidence; yr=years; Z=zoledronic acid
Table D15. Strength of evidence assessments: zoledronic acid continuation (9 years zoledronic acid) versus discontinuation (6 years zoledronic acid followed by 3 years placebo)

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Evidence Summary</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1 (19/190)</td>
<td>Black 2015</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS difference in risk with zoledronic acid 9 yrs (Z9) vs. zoledronic acid 6 yrs followed by placebo 3 yrs (Z6/P3): Z9 11% vs Z6/P3 9%; HR=1.11 [0.45, 2.73]</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>1 (8/190)</td>
<td>Black 2015</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS difference in risk with Z9 vs. Z6/P3: Z9 3.2% vs Z6/P3 5.3%; OR=0.58 [0.13, 2.55]</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>1 (52/190)</td>
<td>Black 2015</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/participants evaluated)</td>
<td>Evidence Summary</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------------------------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Atypical femoral fracture with confirmed radiologic features</td>
<td>1 (0/190)</td>
<td>Black 2015 No cases occurred</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (0/190)</td>
<td>Black 2015 No cases occurred</td>
<td>Low</td>
<td>Direct</td>
<td>Highly imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** CI=confidence interval; HR=hazard ratio; NA=not applicable; NR=not reported; P=placebo; RR=risk ratio; SOE=strength of evidence; yr=years; Z=zoledronic acid
### Any Bisphosphonate

Table D16. Strength of evidence assessments: any bisphosphonate use >3 years versus no bisphosphonate

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Components</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N events/ participants evaluated)</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>Optional Components</td>
<td>SOE</td>
</tr>
<tr>
<td>-------------------------------------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>-----------------</td>
<td>----------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>2 (216 /~2.8 million for cohort; 244 BP users/1124 participants for case-control)</td>
<td>Schilcher 2015 Higher risk of AFF Retrospective cohort analysis: &gt;4 yr BP use: RR=126 [55, 288] Case-control analysis: 3-4 yr BP use: OR=40 [17, 91] 4-5 yr BP use: OR=116 [58, 234] &gt;5 yr BP use: OR=93 [66, 132]</td>
<td>Medium</td>
<td>Direct</td>
<td>Precise</td>
<td>Consistent</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST/FS fracture without radiologic confirmation of AFF features</td>
<td>1 (158 BP users/290 participants)</td>
<td>Lim 2018 Case control Higher risk of AFF with BP use (mean 5.2 yr) OR=25.65 [10.74, 61.28]</td>
<td>Medium</td>
<td>Direct</td>
<td>Precise</td>
<td>Consistent</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>0</td>
<td></td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; BP=bisphosphonate; NA=not applicable; NR=not reported; OR=adjusted odds ratio; RR=risk ratio; SOE=strength of evidence; yr=years
*calculated by EPC
Table D17. Strength of evidence assessments: Any bisphosphonate use >3 years versus past bisphosphonate use

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/ participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Factors</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Atypical femoral fracture with confirmed AFF radiologic features</td>
<td>1 (43/172)</td>
<td>Koh 2017 Higher risk AFF with &gt;5 yr BP use vs. past BP use lasted &gt;1 yr but stopped &gt;6 mo ago: HR=5.17 [2.0, 13.36] HR=2.37 [1.68, 11.41] HR=3.36 [1.77, 11.91]</td>
<td>Medium</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Large magnitude of effect</td>
<td>Low</td>
</tr>
</tbody>
</table>
### Table D18. Strength of evidence assessments: Any bisphosphonate use >3 years versus different osteoporosis drug treatment

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N events/participants evaluated)</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>Optional Factors</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; BP=bisphosphonate; HR=hazard ratio; NR=not reported; NA=not applicable; OR=adjusted odds ratio; RR=risk ratio; SOE=strength of evidence; ST/FS=subtrochanteric/femoral shaft; yr=years

*calculated by EPC
| Outcome                                                                 | # Studies (n/N events/participants evaluated) | Summary statistics [95% CI] | Study Limitations | Directness | Precision | Consistency | Reporting Bias | Optional Factors | SOE            |
|------------------------------------------------------------------------|-----------------------------------------------|-------------------------------|-------------------|------------|-----------|-------------|----------------|------------------|----------------|----------------|
| Incident radiographic vertebral fracture                              | 0                                             | NR                            | NA                | NA         | NA        | NA          | NA             | NA               | Insufficient    |
| Serious adverse events                                                 | 0                                             | NR                            | NA                | NA         | NA        | NA          | NA             | NA               | Insufficient    |
| Subtrochanteric or femoral shaft fracture without confirmed radiologic AFF features | 1 3-5 yr use (26/4900) >5 yr use (8/4097) | Kim 2011 NS difference in risk of subtrochanteric or femoral shaft fracture with longterm use of any BP vs. longterm raloxifene or calcitonin: 3-5 yr BP use: HR=1.20 [0.55, 2.61] >5 yr BP use: HR=2.02 [0.41,10.0] | Medium          | Direct      | Highly Imprecise | Unknown     | Undetected | None          | Insufficient    |
| Osteonecrosis of the jaw                                               | 0                                             | NR                            | NA                | NA         | NA        | NA          | NA             | NA               | Insufficient    |

**Abbreviations:** AFF=atypical femoral fracture; BP=bisphosphonate; HR=hazard ratio; NA=not applicable; NR=not reported; NS=not significant; OR=adjusted odds ration; RR=risk ratio; SOE=strength of evidence; yr=years
*calculated by EPC
### Denosumab

Table D19. Strength of evidence assessments: long-term treatment with denosumab versus placebo

<table>
<thead>
<tr>
<th>Outcome Time frame</th>
<th># Studies (n/N=events/ participants) Population</th>
<th>Summary Statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture 0-4 yrs</td>
<td>1 (38/360) PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>Miller 2008 NS difference in risk with denosumab vs. placebo: 10.5% vs 10.9%; RR=0.97 [0.40, 2.35]</td>
<td>Medium</td>
<td>Indirect</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events 0-4 yrs</td>
<td>1 (61/360) PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>Miller 2008 NS difference in risk with denosumab vs. placebo: 17.8% vs 10.9%; RR=1.64 [0.69, 3.88]</td>
<td>Medium</td>
<td>Indirect</td>
<td>Highly imprecise</td>
<td>Unknown but multisite trial</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

**Abbreviations:** AE=adverse events; BMD=bone mineral density; LS=lumbar spine; NA=not applicable; NR=not reported; PM=postmenopausal; RR=risk ratio; SOE=strength of evidence; yr=years
Table D20. Strength of evidence assessments: denosumab continuation (4 years denosumab) versus discontinuation (2 years denosumab followed by 2 years placebo or by 1 year placebo and 1 year of denosumab)

<table>
<thead>
<tr>
<th>Outcome Time frame</th>
<th>Summary Statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture 0-4 yrs</td>
<td>Among women assigned denosumab, &quot;no increase in fracture incidence among small number of patients who discontinued denosumab,&quot; but no numerical data provided.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events 0-4 yrs</td>
<td>Miller 2008 Authors only reported results for denosumab continuation and discontinuation groups pooled together.</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Atypical femoral fracture 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw 0-4 yrs</td>
<td>Miller 2008 PM women with LS-BMD T-score -1.8 to -4 or hip T-score -1.8 to -3.5</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>
Abbreviations: AE=adverse events; BMD=bone mineral density; LS=lumbar spine; NA=not applicable; NR=not reported; PM=postmenopausal; RR=risk ratio; SOE=strength of evidence; yr=years
### Raloxifene

**Table D21. Strength of evidence assessments: long-term treatment with raloxifene versus placebo**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Time frame</th>
<th># Studies (n/N=events/participants evaluated) Population</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0-4 yrs</td>
<td>1 (844/6828) PM women with T-score ≤-2.5 and/or baseline RVF</td>
<td>Barrett-Connor 2004 NS difference in risk with raloxifene (pooled 60 mg/d + 120 mg/d) vs. placebo: 11% vs. 12%; RR=0.93 [0.81, 1.06]</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>High</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>5-8 yrs</td>
<td>1 (915*/4011) PM women previously received raloxifene vs. placebo for T-score ≤-2.5 and/or baseline RVF</td>
<td>Sirs 2005 NS difference in risk with raloxifene 60 mg/d vs. placebo: 23% vs. 23%; HR=1.00 [0.82, 1.21]</td>
<td>Medium</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>0-4 yrs</td>
<td>1 (85/6828) PM women with T-score ≤-2.5 and/or baseline RVF</td>
<td>Delmas 2002 NS difference in risk with raloxifene (pooled 60 mg/d + 120 mg/d) vs. placebo: 1.1% vs. 1.1%; RR=0.97 [0.62, 1.52]</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>5-8 yrs</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral nonhip fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident clinical</td>
<td>1 (169/5114) PM women with T-score</td>
<td>Sontag 2010</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Outcome Time frame</td>
<td># Studies (n=events/participants evaluated) Population</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------------</td>
<td>-------------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>vertebral fracture 0-4 yrs</td>
<td>&lt;2.5 and/or baseline RVF</td>
<td>Lower risk with raloxifene 60 mg/d vs placebo: 2% vs. 4%; RR=0.58 [0.43, 0.79]</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident clinical vertebral fracture 5-8 yrs</td>
<td>0</td>
<td>Delmas 2002 Lower risk with raloxifene vs. placebo: 60 mg/d vs. placebo 8% vs. 12%; RR = 0.64 [0.53, 0.76] 120 mg/d vs placebo: 7% vs. 12%; RR=0.57 [0.48 to 0.69]</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture 0-4 yrs</td>
<td>1 (615*/6828) PM women with T-score &lt;2.5 and/or baseline RVF</td>
<td>Martino 2004 NS difference in risk with raloxifene 60 mg/d vs. placebo: 42% vs. 46%; RR=0.93 [0.86,1.00]*</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Serious adverse events 0-4 yrs</td>
<td>0</td>
<td>Delmas 2002 Lower risk with raloxifene vs. placebo: 60 mg/d vs. placebo 8% vs. 12%; RR = 0.64 [0.53, 0.76] 120 mg/d vs placebo: 7% vs. 12%; RR=0.57 [0.48 to 0.69]</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Serious adverse events 5-8 yrs</td>
<td>1 (1739/4011) PM women previously received raloxifene vs. placebo for T-score &lt;2.5 and/or baseline RVF</td>
<td>Martino 2004 NS difference in risk with raloxifene 60 mg/d vs. placebo: 42% vs. 46%; RR=0.93 [0.86,1.00]*</td>
<td>Medium</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but multisite trial</td>
<td>Undetected</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Atypical femoral fractures</td>
<td>0</td>
<td>Delmas 2002 Lower risk with raloxifene 60 mg/d vs placebo: 2% vs. 4%; RR=0.58 [0.43, 0.79]</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
<td></td>
</tr>
</tbody>
</table>
### Table D22. Strength of evidence assessments: long-term treatment with raloxifene versus no osteoporosis drug treatment

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N=events/participants evaluated) Study description</th>
<th>Summary statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N=events/participants evaluated)</td>
<td>Study description</td>
<td>Summary statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>ST/FS fracture without radiologic confirmation of AFF features</td>
<td>1 (25/19324) Retrospective cohort. AFF defined by diagnosis codes only. Duration 0-11 yr (mean ~3.8).</td>
<td>Vestergaard 2011 NS difference in risk with raloxifene vs. no osteoporosis drug treatment: Subtrochanteric fracture: HR=1.06 [0.34, 3.32] Femoral shaft fracture: HR=0.82 [0.21, 3.20]</td>
<td>Medium Direct Highly Imprecise Unknown Undetected Insufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>1 (2/19324) Retrospective cohort. ONJ defined by diagnosis codes only. Duration 0-11 yrs (mean ~3.8).</td>
<td>Vestergaard 2012 2 cases of inflammatory jaw events in no treatment group</td>
<td>Medium Direct Highly Imprecise Unknown Undetected Insufficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** AFF=atypical femoral fracture; CI=confidence interval; IRR=incident risk ratio; NA=not applicable; NR=not reported; NS=not significant ONJ=osteonecrosis of the jaw; SOE=strength of evidence; yr=years
## Hormone Therapy

### Table D23. Strength of evidence assessments: long-term treatment with hormone therapy (estrogen/progestin) versus control

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N=events/participants evaluated)</th>
<th>Summary Statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1 (WHI trial, PM women with an intact uterus at baseline) (1629/16608) Full study population</td>
<td>Cauley 2003 Lower risk with estrogen/progestin vs. placebo: 8.6% vs. 11.1%; HR 0.76 (95% CI 0.69, 0.83)</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but large multisite trial</td>
<td>Undetected</td>
<td>High</td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>1 (747/5897) History of fracture</td>
<td>Cauley 2003 Lower risk with estrogen/progestin vs. placebo: 11.2% vs. 14.1%; HR 0.78 (95% CI 0.68, 0.91)</td>
<td>Medium (low for the WHI trial but this subgroup analysis was not specified a priori)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>1 (33/NR) T score ≤-2.5</td>
<td>Cauley 2003 No difference in risk with estrogen/progestin vs. placebo: Annualized percentage 1.4% vs 3.2%; HR 0.53 (0.25, 1.10)</td>
<td>Medium (see above)</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown, and only included 3/40 study sites</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>1 (2/29)</td>
<td>Wimalawansa 1998 NS difference in risk with estrogen/progestin vs. control: 7% vs. 7%; RR=0.93 [0.06, 13.5]</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td></td>
<td>1 (125/16608) Full study population</td>
<td>Cauley 2003 Lower risk with estrogen/progestin vs. placebo: 0.61% vs. 0.90%; HR 0.67 (95% CI 0.47, 0.96)</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N=events/participants evaluated)</td>
<td>Summary Statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
<td>SOE</td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>------</td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>1 (72/5897) History of fracture</td>
<td>Cauley 2003 No difference in risk with estrogen/progestin vs. placebo: 1.1% vs 1.4%; HR 0.77 (95% CI 0.48, 1.22)</td>
<td>Medium (see above)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial</td>
<td>Undetected</td>
<td>Low</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>1 (101/16608) Full study population</td>
<td>Cauley 2003 Lower risk with estrogen/progestin vs. placebo: 0.48% vs. 0.74%; HR 0.65 (0.46, 0.92)</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial</td>
<td>Undetected</td>
<td>Moderate</td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>1 (7/29)</td>
<td>Wimalawansa 1998 NS difference in risk with estrogen/progestin vs. control: 13% vs. 36%; RR=0.37 [0.09, 1.62]</td>
<td>Medium</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown</td>
<td>Undetected</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
</tbody>
</table>

Abbreviations: NA=not applicable; NR=not reported; PM=postmenopausal; RR=risk ratio; SOE=strength of evidence; yr=years
## Table D24. Strength of evidence assessments: long-term treatment with hormone therapy (estrogen) versus placebo

<table>
<thead>
<tr>
<th>Outcome</th>
<th># Studies (n/N=events/participants evaluated) Population</th>
<th>Summary Statistics [95% CI]</th>
<th>Study Limitations</th>
<th>Directness</th>
<th>Precision</th>
<th>Consistency</th>
<th>Reporting Bias</th>
<th>SOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incident clinical fracture</td>
<td>1 (WHI trial, PM women with hysterectomy) (1301/10739) Full study population</td>
<td>Jackson 2006 Lower risk with estrogen vs. placebo: 10% vs. 14%; HR 0.71 (95% CI 0.64, 0.80)</td>
<td>Low</td>
<td>Direct</td>
<td>Precise</td>
<td>Unknown, but large multisite trial Undetected</td>
<td>High</td>
<td></td>
</tr>
<tr>
<td>Incident clinical fracture</td>
<td>1 (621/3816) History of fracture</td>
<td>Jackson 2006 Lower risk with estrogen vs. placebo: 14% vs. 19%; HR 0.73 (95% CI 0.62, 0.86)</td>
<td>Medium (low for the WHI trial but this subgroup analysis was not specified a priori)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial Undetected</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>1 (NR/363*) T score between -1.0 and -2.5</td>
<td>Jackson 2006 No difference in risk with estrogen vs. placebo: HR 0.83 (95% CI 0.49, 1.40)</td>
<td>Medium (see above)</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown, and only included 3/40 study sites Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>1 (NR/53*) T score ≤-2.5</td>
<td>Jackson 2006 No difference in risk with estrogen vs. placebo: HR 0.83 (95% CI 0.17, 3.91)</td>
<td>Medium (see above)</td>
<td>Direct</td>
<td>Highly Imprecise</td>
<td>Unknown, and only included 3/40 study sites Undetected</td>
<td>Insufficient</td>
<td></td>
</tr>
<tr>
<td>Incident nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Incident hip fracture</td>
<td>1 (119/10739) Full study population</td>
<td>Jackson 2006 Lower risk with estrogen vs. placebo: 0.87% vs 1.3%; HR 0.65 (95% CI 0.45, 0.94)</td>
<td>Low</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial Undetected</td>
<td>Moderate</td>
<td></td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>1 (57/3816) History of fracture</td>
<td>Jackson 2006 Lower risk with hormone therapy vs. placebo: 1.0% vs 1.9%; HR 0.55 (95% CI 0.32, 0.94)</td>
<td>Medium (see above)</td>
<td>Direct</td>
<td>Imprecise</td>
<td>Unknown, but large multisite trial Undetected</td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Incident nonhip nonvertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Insufficient</td>
</tr>
<tr>
<td>Outcome</td>
<td># Studies (n/N=events/participants evaluated)</td>
<td>Population</td>
<td>Summary Statistics [95% CI]</td>
<td>Study Limitations</td>
<td>Directness</td>
<td>Precision</td>
<td>Consistency</td>
<td>Reporting Bias</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Incident major osteoporotic fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Incident clinical vertebral fracture</td>
<td>1 (112/10739) Full study population</td>
<td>Jackson 2006 Lower risk with estrogen vs. placebo: 0.81% vs 1.3%; HR 0.64 (95% CI 0.44, 0.93)</td>
<td>Low Direct Imprecise Unknown, but large multisite trial Undetected</td>
<td>Moderate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incident radiographic vertebral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Atypical femoral fracture</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Osteonecrosis of the jaw</td>
<td>0</td>
<td>NR</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Abbreviations: NA=not applicable; NR=not reported; PM=postmenopausal; HR=hazard ratio; SOE=strength of evidence

*Calculated by EPC
Appendix E. Excluded References


47. Archer D, F. Pinkerton, J, V. Utian, W, H. Menegoci, J, C. de, Villiers T, J. Yuen,


97. Bjarnason N, H. Christiansen, C,. The influence of thinness and smoking on bone loss and response to hormone replacement therapy in early postmenopausal women. Journal of Clinical Endocrinology &


treated with monthly ibandronate or weekly
alendronate and risedronate. Annals of
PMID: 19318598. Ineligible treatment
duration.

109. Bock O, Borst, H, Beller, G, Armbrecht,
G, Degner, C, Martus, P, Roth, H, J.
Felsenberg, D, Impact of oral ibandronate
150 mg once monthly on bone structure and
density in post-menopausal osteoporosis or
osteopenia derived from in vivo mCT.
Bone. 2012 Jan;50(1):317-24. PMID:
22067902. Ineligible study design.

110. Body J, J. Gaich, G, A. Scheele, W, H.
Kulkarni, P, M. Miller, P, D, Peretz, A,
Dore, R, K. Correa, Rotter R,
Papaioannou, A, Cumming, D, C.
Hodsman, A, B. A randomized double-
blind trial to compare the efficacy of
teriparatide [recombinant human parathyroid
hormone (1-34)] with alendronate in
postmenopausal women with osteoporosis.
Journal of Clinical Endocrinology &
PMID: 12364430. Ineligible study design.

111. Body J, J. How to manage postmenopausal
osteoporosis? Acta Clinica Belgica. 2011
Ineligible study design.

112. Body J, J. Bergmann, P, Boonen, S,
Devogelaer, J, P. Gielen, E, Goemaere, S,
Kaufman, J, M. Rozenberg, S, Reginster,
J, Y. Extraskeletal benefits and risks of
calcium, vitamin D and anti-osteoporosis
medications. Osteoporosis International.
2012 February;23(SUPPL. 1):S1-S23.
PMID: 364305818. Ineligible study design.

113. Bogado C, E. Massari, F, E. Zanchetta, J,
R. Parathyroid hormone (1-84) and
teriparatide in the treatment of
postmenopausal osteoporosis. Women's
health. 2006 May;2(3):447-57. PMID:
19803916. Ineligible study design.

114. Bogado C, E. Boailchuk, J, A. Zanchetta,
M, B. Massari, F, E. Zanchetta, J, R.
Ineligible study design.

115. Bogado C, E. Zanchetta, M, B. Boailchuk,
J, A. Massari, F, E. Zanchetta, J, R.
Denosumab: What's new? Current
PMID: 21132405. Ineligible study design.

116. Bolanos R, Francia, J, Isoflavones versus
hormone therapy for reduction of vertebral
fracture risk: indirect comparison.
PMID: 20613672. Ineligible study design.

117. Bolognese M, Teglbjærg, Cs, Zanchetta,
Jr, Lippuner, McClung, Mr, Brandi, Mi,
Heiseth, Lakatos, Moffett, Ah, Lorenc,
Rs, Wang, Libanati, Denosumab
significantly increases DXA BMD at both
trabecular and cortical sites: results from the
FREEDOM study. Journal of clinical
densitometry. 2013;16(2):147-53. PMID:
CN-00969747. Ineligible study design.

Harris, S, T. Weinstein, R, S. Licata, A,
A. McClung, M, R. Kimmel, D, B.
Gertz, B, J. Hale, E, Polvino, W, J. Dose-
response relationships for alendronate
treatment in osteoporotic elderly women.
Alendronate Elderly Osteoporosis Study
Centers. Journal of Clinical Endocrinology &
PMID: 8989272. Ineligible treatment
duration.

119. Bone H, G. Greenspan, S, L. McKeever,
C, Bell, N, Davidson, M, Downs, R, W.
Emkey, R, Meunier, P, J. Miller, S, G.
Mullay, A, L. Recker, R, R. Weiss, S, R.
Heyden, N, Musliner, T, Suryawanshi, S,
Yates, A, J. Lombardi, A, Alendronate
and estrogen effects in postmenopausal
women with low bone mineral density.
Alendronate/Estrogen Study Group. Journal
of Clinical Endocrinology & Metabolism.
No outcomes of interest.

120. Bone H, G. Bolognese, M, A. Yuen, C, K.
Kendler, D, L. Wang, H, Liu, Y, San,
Martin J, Effects of denosumab on bone
mineral density and bone turnover in
postmenopausal women. Journal of Clinical
Endocrinology & Metabolism. 2008


152. Brown J, Roux, Töring, Ho, Pr, Beck, Jensen, Je Gilchrist, Recknor, Austin, Wang, Grauer, Wagman, Rb,. Discontinuation of denosumab and


Dec;86(12):5700-5. PMID: 11739424. *Ineligible treatment duration.*


224. Cole Z, Dennison, E. Cooper, C.. Update on the treatment of post-menopausal


Ineligible treatment duration.


Ineligible study design.


Ineligible study design.


380. Fraser L, A, Papaioannou, A, Adachi, J, D, Ma, J, Thabane, L, CaMos, Research Group,. Fractures are increased and bisphosphonate use decreased in individuals


Ineligible study design.


438. Greenspan S, L. Schneider, D, L. McClung, M, R. Miller, P, D. Schnitzer,


No outcomes of interest.


621. Kristensen S, Abrahamsen, Madsen, Js, Gram, Rejmark, Rud, Jespersen., Venous thrombosis is not increased in younger women on genuine oestrogen postmenopausal hormonal replacement therapy: results from the Danish Osteoporosis Prevention Study (DOPS). Thrombosis and haemostasis. 2006;95(5):915-6. PMID: CN-00565072. **Ineligible population.**


624. Kurland E, S. Cosman, F, McManus, D, J. Rosen, C, J. Lindsay, R, Bilezikian, J, P. Parathyroid hormone as a therapy for idiopathic osteoporosis in men: effects on bone mineral density and bone markers. Journal of Clinical Endocrinology & Metabolism. 2000 Sep;85(9):3069-76. PMID: 10999788. **No outcomes of interest.**


652. Lees B, Stevenson, J, C. The prevention of osteoporosis using sequential low-dose hormone replacement therapy with estradiol-


E-63


813. Murphy MM. Role of the Pharmacist in Medication Therapy Management Services in Patients With Osteoporosis. Clinical Therapeutics. 2015 Jul 01;37(7):1573-86. PMID: 25913920. Ineligible study design.


Ineligible study design.

No outcomes of interest.

No interventions of interest.


975. Riis B, J. Overgaard, K, Christiansen, C. Biochemical markers of bone turnover to monitor the bone response to postmenopausal hormone replacement therapy. Osteoporosis International.


980. Ringe J, D. Farahmand, P, Schacht, E, Rozehnal, A,. Superiority of a combined treatment of Alendronate and Alfacalcidol compared to the combination of Alendronate and plain vitamin D or Alfacalcidol alone in established postmenopausal or male osteoporosis (AAC-Trial). Rheumatology International. 2007 Mar;27(5):425-34. PMID: 17216477. No outcomes of interest.


992. Rizzoli R. Bisphosphonates for postmenopausal osteoporosis: are they all the same? Qjm. 2011 Apr;104(4):281-300. PMID: 21258058. Ineligible study design.


1012. Saarelainen J, Hassi, S, Honkanen, R, Koivumaa, Honkanen H, Sirola, J,


1023. Salmen, T., Heikkinen, A. M., Mahonen, A., Kroger, H., Komulainen, M., Pallonen, H,


1034. Sato M, Ye, W, Sugihara, T, Isaka, Y,. Fracture risk and healthcare resource utilization and costs among osteoporosis patients with type 2 diabetes mellitus and without diabetes mellitus in Japan: retrospective analysis of a hospital claims


1097. Stevenson J, C. Teter, P, Lees, B., 17beta-estradiol (1 mg/day) continuously combined with dydrogesterone (5, 10 or 20 mg/day) increases bone mineral density in postmenopausal women. Maturitas. 2001 Apr 20;38(2):197-203. PMID: 11306209. Ineligible treatment duration.


1109. Suresh E, Pazianas, M, Abrahamsen, B,. Safety issues with bisphosphonate therapy


Vestergaard P. Acute myocardial infarction and atherosclerosis of the coronary arteries in patients treated with drugs against osteoporosis: calcium in the vessels and not the bones? Calcified Tissue International.


1206. Wells G, Cranney, A, Peterson, J, Boucher, M, Shea, B, Robinson, V, Coyle, D, Tugwell, P,. Risedronate for the


