Raloxifene for Primary and Secondary Prevention of Osteoporotic Fractures in Postmenopausal Women: A Systematic Review of Efficacy and Safety Evidence
Canadian Coordinating Office for Health Technology Assessment

Raloxifene for Primary and Secondary Prevention of Osteoporotic Fractures in Postmenopausal Women: A Systematic Review of Efficacy and Safety Evidence

Howard M Schachter PhD
Tammy J Clifford PhD
Ann Cranney MBBCh
Nicholas J Barrowman PhD
David Moher MSc

January 2005

1 Chalmers Research Group, CHEO Research Institute, Ottawa
2 Clinical Epidemiology Program, Ottawa Health Research Institute, Ottawa
3 School of Mathematics and Statistics, Carleton University, Ottawa
4 Department of Pediatrics, University of Ottawa, Ottawa
5 Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa
Reviewers

These individuals kindly provided comments on this report.

External Reviewers

Allan Mills, PharmD CGP
Pharmacy Department
Baycrest Centre for Geriatric Care
3560 Bathurst Street
Toronto ON

CCOHTA Scientific Advisory Panel Reviewers

Jeffrey Barkun, MD CM FRCSC FACS MSc (Epidemiology)
Head, Hepatobiliary and Transplantation Unit
Director of Clinical Research in Transplantation
McGill University Health Centre
Associate Professor of Surgery
McGill University
Montreal QC

Jeff Mahon, MD FRCPC
Associate Professor of Medicine
London HSC, University Campus Site
University of Western Ontario
London ON

This report is a review of existing public literature, studies, materials and other information and documentation (collectively the “source documentation”) which are available to CCOHTA. The accuracy of the contents of the source documentation on which this report is based is not warranted, assured or represented in any way by CCOHTA and CCOHTA does not assume responsibility for the quality, propriety, inaccuracies or reasonableness of any statements, information or conclusions contained in the source documentation.

CCOHTA takes sole responsibility for the final form and content of this report. The statements and conclusions in this report are those of CCOHTA and not of its Panel members or reviewers.

Authorship

H.M. Schachter helped refine the research questions; developed and implemented methods used; interpreted the results; wrote all drafts of the report; and managed the review team.

T.J. Clifford, with H.M. Schachter, contributed to the conception and design of the review, including the eligibility criteria data abstraction forms; assisted in data abstraction; and critically reviewed drafts of the report.

A. Cranney provided content expertise, assisted with selection of relevant studies from search results and provided input on drafts of the report.
N.J. Barrowman contributed to the data synthesis plan including imputation strategy, statistical heterogeneity testing and assessment of publication bias; performed quantitative data synthesis; interpreted heterogeneity statistics; and participated in drafting and revising the report.

D. Moher contributed methodological expertise.

Acknowledgements

The authors thank Ms. Margaret Sampson for conducting a feasibility search and for reviewing the final search strategy, Mr. Jagadish Rangrej for helping to organize several of the analyses, Ms. Janet Joyce for performing electronic searches, Ms. Kaitryn Campbell for reference verification and Mr. Ken Bassett for providing his valuable input in writing the final draft.

Conflicts of Interest

H.M. Schachter, T.J. Clifford, N.J. Barrowman, J. Joyce and D. Moher disclosed no conflicts. A. Cranney has received consulting fees from Merck Frosst, Procter and Gamble and Eli Lilly.
**Technology Name**
Raloxifene (Evista®)

**Disease or Condition**
Osteoporosis is a progressive condition of decreased bone mass. This leads to fragile bones, which are prone to fracture. In women, osteoporotic fractures of the vertebrae can result in back pain, decreased quality of life and an increased risk of spinal fractures.

**Technology Description**
Raloxifene is a selective estrogen receptor modulator (SERM) drug. Licensed in Canada in 1999 for the prevention of osteoporosis in postmenopausal women, raloxifene does not heal existing vertebral deformity or fractures, but it is used to prevent subsequent fractures or to delay structural bone changes.

**The Issue**
With an aging Canadian population, vertebral fractures will become more common. There is a need to assess raloxifene’s efficacy and safety compared with other drugs approved for the prevention of osteoporosis.

**Assessment Objectives**
- To systematically assess, using a meta-analysis if appropriate, raloxifene’s safety and efficacy to prevent osteoporotic vertebral fractures in postmenopausal women.
- To compare raloxifene against placebo and other drug treatments, including estrogen alone in women with hysterectomies, estrogen-progestin combination therapies, bisphosphonates (alendronate, risedronate and etidronate) and salmon calcitonin.

**Methods**
An extensive literature search identified randomized controlled trials that assessed the safety and efficacy of raloxifene in postmenopausal women. The main outcome measure was vertebral fracture, with the secondary outcome measure being bone mineral density. A total of 17 studies met the inclusion criteria for the review.

**Conclusions**
- Raloxifene has no effect on the incidence of non-vertebral fractures such as hip fractures.
- Because of trial diversity, the planned meta-analysis to compare raloxifene’s efficacy with other drugs used to prevent osteoporosis could not be accomplished.
- Compared with placebo, raloxifene’s primary clinical benefit is a reduction in vertebral fracture in older postmenopausal women, particularly if some vertebral fractures are present at baseline.
- This benefit is offset by a similar increase in serious adverse events due to venous thromboembolic disease.
- Raloxifene causes significantly more mild to moderate adverse effects such as hot flashes and leg cramps compared with placebo, but significantly less vaginal bleeding than estrogen-progestin combination therapy.

This summary is based on a comprehensive health technology assessment available from CCOHTA’s web site (www.ccohta.ca): Schachter HM, Clifford TJ, Cranney A, Barrowman NJ, Moher D. Raloxifene for primary and secondary prevention of osteoporotic fractures in postmenopausal women: a systematic review of efficacy and safety evidence.
EXECUTIVE SUMMARY

The Issue

Osteoporosis is a progressive systemic disease characterized by low bone density and deterioration of bone tissue, with bones becoming fragile and prone to fracture. In women, symptomatic and asymptomatic osteoporotic vertebral fractures often occur 10 to 15 years after menopause, which may result in back pain, decreased quality of life and an increased risk of spinal fractures. With an aging Canadian population, vertebral fractures will become more common. In 1993, the cost of illness in Canada due to osteoporosis and its complications was estimated at $1.3 billion. In 1999, raloxifene was licensed in Canada for the prevention of osteoporosis in postmenopausal women. Raloxifene does not heal existing fractures or vertebral fracture, but is used to delay structural bone changes and prevent future fractures. Primary prevention is aimed at women who have no baseline radiologic evidence of vertebral fracture, while secondary prevention targets women with baseline evidence of vertebral fracture.

Objectives

The purpose of this systematic review is to evaluate, using a meta-analysis if appropriate, the safety and efficacy of raloxifene compared with placebo and other drugs approved in Canada for the primary and secondary prevention of osteoporotic vertebral fractures in postmenopausal women. These include estrogen alone in women with hysterectomies, estrogen-progestin therapies, bisphosphonates (alendronate, risedronate and etidronate) and salmon calcitonin. Vertebral fracture was identified \textit{a priori} as the primary outcome measure of efficacy, with bone mineral density (BMD) a secondary outcome measure. A meta-analysis was planned to evaluate these outcomes and adverse event data. Several subgroup and sensitivity analyses will evaluate the robustness and validity of the trials.

Methods

Published and unpublished randomized controlled trials (RCTs) were identified via electronic and manual searches. Trials were evaluated for methods-related quality and synthesized qualitatively and quantitatively.

Results

Seventeen RCTs met the eligibility criteria, including six that compared raloxifene to placebo, four versus estrogen-progestin therapies; two versus estrogen alone (women with hysterectomies), one versus alendronate, one versus tamoxifen and three versus estrogen alone (no hysterectomy). No RCT compared raloxifene to etidronate, risedronate or salmon calcitonin. Tamoxifen and estrogen alone in women without hysterectomy are not approved for osteoporosis prevention.

Postmenopausal women in secondary prevention studies were older and had spent more years post-menopause than those in primary prevention studies. Most trials used a dose of 60 mg/day, which is the recommended dosage in Canada. According to the Jadad scale, the methodological quality of the studies ranged from low to moderately high. The largest study was the Multiple
Outcomes of Raloxifene Evaluation (MORE) trial, which included primary and secondary prevention arms.

In all trials reporting the outcomes of interest, raloxifene was compared with placebo. Adverse event data were obtained from trials comparing raloxifene with placebo, estrogen or combined estrogen-progestin therapy.

In the MORE trial, raloxifene had no effect on the incidence of non-vertebral fractures such as hip fracture.

Two studies reported data on the primary outcome of vertebral fracture, although there were differences in its definition and assessment, in patient populations, in study size and in duration of treatment. As a result, the planned meta-analysis of vertebral fracture data was inappropriate because of issues related to the choice of statistical model and trial heterogeneity.

**Conclusion**

The best estimate of overall benefit and harm from raloxifene is derived from the results of the three-year, placebo-controlled MORE trial. In this trial, the absolute risk reduction for symptomatic vertebral fractures is 0.8% (1.4% in placebo and 0.6% in raloxifene groups, number needed to treat=125 over three years). This benefit needs to be balanced by an increase in the number of serious adverse events due to venous thrombo-embolic disease of 0.65% (0.31% in placebo and 0.96% in raloxifene, number needed to harm=154 over three years).

The largest observed incidence of vertebral fracture occurs among older postmenopausal women with vertebral fractures present at baseline. This is also the group in which raloxifene has the greatest impact. The group includes women with a range of BMD levels. The impact of raloxifene to prevent vertebral fractures in younger postmenopausal women or women with higher BMD levels but no pre-existing vertebral fractures, has not been established.

The impact of raloxifene on morbidity and mortality due to cardiovascular disease or cancer (breast or uterine) has not been established, although the evidence suggests a small decrease in the incidence of early breast cancer. A subgroup analysis of cardiovascular adverse events reported in the MORE trial indicates no significant increase in morbidity and mortality in raloxifene users as compared to placebo, after three years.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>iv</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Clinical Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Technology</td>
<td>2</td>
</tr>
<tr>
<td>2 OBJECTIVES</td>
<td>3</td>
</tr>
<tr>
<td>3 CLINICAL EFFECTIVENESS REVIEW</td>
<td>3</td>
</tr>
<tr>
<td>3.1 Methods</td>
<td>3</td>
</tr>
<tr>
<td>3.1.1 Trial identification</td>
<td>3</td>
</tr>
<tr>
<td>3.1.2 Eligibility criteria</td>
<td>4</td>
</tr>
<tr>
<td>3.1.3 Selection process</td>
<td>4</td>
</tr>
<tr>
<td>3.1.4 Data abstraction</td>
<td>4</td>
</tr>
<tr>
<td>3.1.5 Assessment of methodological quality of trials from reports</td>
<td>5</td>
</tr>
<tr>
<td>3.1.6 Data synthesis and analysis of efficacy and safety data</td>
<td>5</td>
</tr>
<tr>
<td>3.1.7 Publication bias</td>
<td>5</td>
</tr>
<tr>
<td>3.2 Results</td>
<td>6</td>
</tr>
<tr>
<td>3.2.1 Qualitative results</td>
<td>6</td>
</tr>
<tr>
<td>3.2.2 Quantitative results</td>
<td>13</td>
</tr>
<tr>
<td>4 DISCUSSION</td>
<td>25</td>
</tr>
<tr>
<td>5 CONCLUSIONS</td>
<td>29</td>
</tr>
<tr>
<td>6 REFERENCES</td>
<td>30</td>
</tr>
<tr>
<td>Appendix 1: Literature Search Strategy</td>
<td>39</td>
</tr>
<tr>
<td>Appendix 2: Additional Details Concerning Methods</td>
<td>41</td>
</tr>
<tr>
<td>Appendix 3: Tools to Assess Quality of Randomized Controlled Trials</td>
<td>43</td>
</tr>
<tr>
<td>Appendix 4: Raloxifene for Primary and Secondary Prevention of Osteoporosis in Postmenopausal Women</td>
<td>44</td>
</tr>
<tr>
<td>Appendix 5: Classification of Trials by Bone Mineral Density of Participants at Baseline</td>
<td>49</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

1.1 Clinical Background

Osteoporosis is defined as “a progressive systemic disease characterized by low bone density and micro-
architectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility
to fracture.”\(^1\) Its impact is a great personal and collective cost to Canadians.\(^2-4\) In 1996, 9,382 patients
were hospitalized in Ontario for hip fracture.\(^3\) In 1993, the cost of illness in Canada due to osteoporosis
and its complications was estimated at $1.3 billion.\(^4\) As of 2001, the annual burden of care related to hip
fractures in Canada was estimated at $650 million.\(^5\) Hip fracture rates increase exponentially with age.
The estimated annual incidence is 4.8 per 10,000 Canadian women aged 70, increasing to 12.5 per
10,000 of those aged 80.\(^6\)

The focus of our research is symptomatic (clinical) and asymptomatic vertebral fracture, which is the
most common morphometric change noted on x-rays of postmenopausal women. These changes are
often seen 10 to 15 years after natural or surgical menopause.\(^1,3\) The population with vertebral
deformities tends to be younger on average than women who suffer hip fractures, which usually occur
after age 75.

The clinical impact of vertebral fracture varies, with some people experiencing mild discomfort and
some who are asymptomatic.\(^7\) Asymptomatic vertebral fracture is often identified after radiographic
confirmation of a subsequent symptomatic vertebral fracture.\(^7,8\) Vertebral fractures can result in back
pain and decreased quality of life.\(^5,11\) Women who experience vertebral fractures are at high risk to
sustain another within the next 12 months.\(^12\) In addition, vertebral fractures are a marker or risk factor for
future hip fractures.\(^13\) With an aging Canadian population, vertebral fractures will become more common
in the future.\(^14\)

The crucial clinical manifestations of osteoporosis are symptomatic vertebral fracture and symptomatic
non-vertebral fractures, particularly hip fracture. Reflecting a combination of bone density and bone
quality (i.e., bone geometry; bone architecture and connectivity; remodelling status; accumulated micro-
damage), bone strength influences the propensity to fracture.\(^15\) Since it is impossible to directly assess
bone quality, bone mineral density (BMD) is typically used to assess the fracture risk associated with
bone aging,\(^15\) most commonly measured with dual energy x-ray absorptiometry (DXA).\(^16\)

BMD measurements are often interpreted according to guidelines produced by a study group on
densitometry hosted by the World Health Organization (WHO) in 1993.\(^17\) The guidelines define
osteoporosis as a BMD T-score (i.e., a standardized score estimated from sample parameters) of greater
than 2.5 standard deviations (SD) below the average value for peak young adult bone mass. In this
model, women with very low BMD are considered to be in need of treatment, presumably for their low
BMD level. Osteopenia is defined as a T-score between 2.5 and 1 SD below peak bone mass; and
“healthy” as a BMD value less than 1 SD below this reference point.\(^17,18\)
Labelling women as either having or not having osteoporosis, based on BMD measurements alone, is problematic, given the potential for misclassification. Other risk factors for osteoporotic fractures include low body weight, family history of osteoporotic fracture, history of fragility fracture, history of falls and increasing age, which is the dominant clinical risk factor. BMD is an important risk factor when it is combined with age over 65. All risk factors, including loss of bone mass, increase with age, but their association with fractures increases more rapidly starting at menopause. The absolute risk of the most clinically important fragility fracture – hip fracture – is low at menopause and occurs almost exclusively in women over age 70.

Raloxifene is prescribed to women identified as having osteoporosis through an incident non-vertebral fracture, a symptomatic vertebral fracture, an asymptomatic vertebral fracture found on a routine x-ray, a low BMD level or a combination of risk factors. Raloxifene is not used to treat the incident fracture or vertebral fracture. Rather, it is used to prevent the next incident vertebral fracture. Thus, the term “treatment” is not used in this report, because raloxifene and other anti-resorptive drugs for osteoporosis do not contribute to the healing of fractures. Raloxifene therapy is characterized as therapy for the “primary” or “secondary” prevention of fractures. Primary prevention is used to categorize patients if no vertebral fracture is apparent at baseline. Secondary prevention is used if vertebral fracture is present.

1.2 Technology

According to the federal government’s approved product monograph, raloxifene (Evista®: Eli Lilly) is a member of the benzothiophene class of compounds known as the selective estrogen receptor modulators (SERM). According to the product monograph, raloxifene has estrogen agonist effects on bone and lipid metabolism concomitant with an estrogen antagonist effect on uterine and breast tissue. SERMs can decrease the resorption of bone tissue by osteoclasts. This inhibits the process by which bone is lost.

According to the product monograph, raloxifene exerts various effects on the three-dimensional conformation of the ligand-bound receptor. This accounts for its agonist properties on some tissues and its antagonistic properties with respect to others. It acts antagonistically as a competitive inhibitor of 17-beta-estradiol for the estrogen receptor. It forms estrogen receptor-beta ligand complexes distinct from estrogen, with the estrogen alpha-receptor binding to a raloxifene response element; and codes for transforming growth factor (TGF)-beta, which is a regulator of bone remodelling.

Raloxifene is promoted as a long-term approach to primary and secondary prevention of osteoporosis in postmenopausal women. It has been approved in Canada for these two purposes (described as prevention and treatment of osteoporosis in the product monograph), with a recommended dose of 60 mg/day.
2 OBJECTIVES

The focus of this systematic review is to evaluate the safety and efficacy of raloxifene for the primary and secondary prevention of symptomatic and asymptomatic osteoporotic vertebral fractures. The required level of evidence is the randomized trial (RCT), given its status as the gold standard for research on the efficacy of health care interventions. No restrictions are placed on the type of control or comparator, thereby allowing the possible elucidation of raloxifene’s absolute (versus placebo) and relative (versus other standard treatments) utility.

A meta-analysis was planned to evaluate vertebral fracture (i.e., primary outcome), BMD (secondary outcome) and adverse event data. We also planned to use a meta-analysis of primary outcome data to investigate possible population (e.g., age; number of years post-menopause; race; geographic location as a surrogate index of vitamin D or sunlight exposure) and intervention sources of clinical heterogeneity (i.e., raloxifene dose; intervention length; use of supplements). We aimed to appraise the robustness and validity of raloxifene’s effect in light of trial quality, study design and publication bias.

3 CLINICAL EFFECTIVENESS REVIEW

3.1 Methods

3.1.1 Trial identification

Without restriction on the publication status or language of reports or the year of publication, several electronic databases were searched: MEDLINE® (1966 to February 2002 inclusive), EMBASE (1974 to July 2001 inclusive), Current Contents Search® (1990 to February 2002), Adis LMS Drug Alerts (1983 to April 2001), Pascal (1973 to July 2001), HealthSTAR (1975 to October 2000), Unlisted Drugs (1984 to July 1994), TOXLINE® (1965 to December 2000), Inside Conferences (1993 to April 1999), BioBusiness (1985 to August 1998), BIOSIS Previews® (1993 to April 1999) and International Pharmaceutical Abstracts (1970 to March 1999). These searches included a study design filter to capture RCTs. The topic was identified by focusing on terms in the titles, abstracts and subject fields of all records. These terms were contained in the MEDLINE search strategy described in Appendix 1. The searches also included a standard filter to capture systematic reviews and meta-analyses.

In a surrogate hand search, the Cochrane Library’s Controlled Trials Register was consulted (2002, Issue 1). The Cochrane Library was also searched for extant systematic reviews and meta-analyses.

Through contact with content experts and the manufacturer of raloxifene and through trial registries such as Current Controlled Trials (2000 to March 6, 2002), attempts were made to identify published and unpublished studies (e.g., grey literature reports such as conference
proceedings or ongoing studies that are “unpublished, have limited distribution, and are not included in bibliographic retrieval systems”\textsuperscript{30}). The Eli Lilly Canada representative provided trial reports, trial information and data, including guidance as to which reports referred to which unique studies. The reference lists from textbooks, reviews and reports of relevant primary studies were examined in an effort to identify additional trial material.


3.1.2 Eligibility criteria

A trial was eligible for inclusion if it met each of the following criteria:

- an RCT involving the use of raloxifene of any dose compared to other drugs or placebo for postmenopausal women with low BMD (i.e., a T-score of more than 2.5 SD below the average value for peak young adult bone mass) or women with higher BMD levels (above the 2.5 SD threshold)
- the inclusion of at least one of the following outcomes: incident radiographically confirmed vertebral fractures (primary outcome); BMD (secondary outcome), because of its key role in the clinical diagnosis of osteoporosis and despite its poor predictive value for future fractures;\textsuperscript{24} all reported adverse events (e.g., breast cancer, venous thromboembolic events, hot flashes).

All definitions of “postmenopausal” were included, e.g., natural or surgical (post-hysterectomy). We accepted trials that involved co-therapies such as calcium, vitamin D or exercise.

We excluded studies whose focus was the impact of raloxifene on biochemical markers of bone turnover, bone remodelling kinetics or bone histomorphometry; quality of life; cardiovascular risk (e.g., C-reactive protein, serum lipoprotein or plasma homocysteine levels); cognitive function; and neuromuscular function (e.g., balance or falls).

3.1.3 Selection process

The selection process by which evidence was organized and evaluated for relevance involved many steps,\textsuperscript{31} which are described in Appendix 2.

3.1.4 Data abstraction

After a calibration exercise involving five studies, two independent reviewers (HMS, TC) abstracted the content of each included trial using a form that focused on the report (language of publication, year of publication, published versus unpublished status and sources of funding); trial (design, number of centres); population (sample size, age, years post-menopause, country in which the study was conducted); intervention characteristics (intervention length; raloxifene
dosing schemes; types, magnitudes and frequencies of control or comparator interventions); and co-interventions (e.g., supplements). Primary (vertebral fracture incidence), secondary (BMD) and all reported adverse event outcome data (e.g., breast cancer, venous thromboembolic events, hot flashes)\textsuperscript{25-27} were also documented. Consensus was used to resolve disagreements. The original reports were not masked, as there was conflicting evidence regarding the benefit of this practice.\textsuperscript{32}

3.1.5 Assessment of methodological quality of trials from reports

The strategy by which the methodological quality of trials was assessed\textsuperscript{33,34} is described in Appendix 2. The instruments are listed in Appendix 3.

3.1.6 Data synthesis and analysis of efficacy and safety data

Report and trial characteristics were summarized qualitatively. Qualitative and quantitative evaluations were performed separately when different controls or comparators were involved.

Statistical analyses followed the intention-to-treat principle, focusing on data collected during the last follow-up visit at which participants were receiving the intervention. Conventions relating to data synthesis and analysis\textsuperscript{35-40} are presented in Appendix 2.

Sensitivity and subgroup analyses were planned to investigate possible sources of clinical heterogeneity in primary efficacy data. The following potential effect modifiers were to be tested: age; number of years post-menopause; race; geographic location of a study as a surrogate index of vitamin D (i.e., sunlight) exposure; dosage (i.e., low: $\leq 60$ mg/day; high: $\geq 120$ mg/day; combined: all doses); duration (e.g., with 12 months considered to be the minimum length needed to test efficacy and safety) and the use of vitamin D and calcium supplements.\textsuperscript{16,28,41} We also intended a priori to evaluate whether the features of the Multiple Outcomes of Raloxifene Evaluation (MORE) trial changed the meta-analytic picture of raloxifene’s efficacy: i.e., changing the therapeutic guidelines after year 3 (the primary trial endpoint). In year 4, participants were allowed to take additional bone-active agents other than oral estrogen.

Sensitivity analyses planned for primary outcome data included trial design and trial quality defined using the Jadad score. The Jadad score is a numerical quantity assigned to a trial based on the presence of defined characteristics (Appendix 3) (i.e., high versus low) and on the adequacy of concealment of allocation to trial arms. The only planned subgroup analysis involving safety data focused on the impact of dosage. Additional conventions are described in Appendix 2.

3.1.7 Publication bias

Publication bias is the tendency to preferentially publish statistically positive results.\textsuperscript{42} Methods\textsuperscript{43,44} to deal with it are described in Appendix 2.
3.2 Results

3.2.1 Qualitative results

Appendix 4 lists every report regardless of publication status; or whether they are duplications or describing non-overlapping information or data (e.g., efficacy versus safety). More than one report can refer to a given study (Figure 1). Qualitative summaries of reports and trials are included in Tables 1 to 3. All studies use a parallel group design.

a) Report characteristics

Of 486 citations included after screening, 262 were excluded and 224 were considered potentially relevant. The full reports for the latter were retrieved and assessed for relevance (Figure 1). Of these, 159 reports were excluded, leaving 65 reports describing 17 unique RCTs that were entered into a qualitative assessment and eligible for the meta-analysis. Reasons for exclusion at each stage of the systematic review are presented in Figure 1. One report was never found.

Fifteen reports were excluded because they integrated data from at least one trial already included in the systematic review; or from at least one trial whose identity and key population or intervention parameters were insufficiently defined to permit the extrapolation of its results. One included report involved two identical trials and described one study undertaken at two sites.

Most trials were described in at least one published report [70.6% (12/17)] (Table 3) and all reports were found in English-language journals between 1996 and March 2002 inclusive. Eli Lilly, Canada’s representative confirmed that the company had sponsored 12 of the 17 included trials.

Given the multiple reports for several trials, we refer to each trial by using the first author’s name (Appendix 4) (for example, Lufkin or Meunier). The exception is the Multiple Outcomes of Raloxifene Evaluation, which is known as the MORE trial.

b) Population and trial characteristics

The studies were categorized according to baseline BMD as defined a priori:

- category 1: trials with all participants with very low BMD that is more than 2.5 SD below peak young adult level (n=2) (MORE, Lufkin)
- category 2: trials with fewer than all participants with very low BMD (i.e., some had BMD less than 2.5 SD below the peak levels) (n=6)
- category 3: trials with all participants with higher BMD (i.e., less than 2.5 SD below peak levels) (n=11)
- category 4: trials with fewer than all participants identified as having higher BMD that is less than 2.5 SD below peak levels (i.e., some women had very low BMD) (n=15)
- category 5: mixed trials with participants having a range of BMD levels (n=4) (Appendix 4, Tables 1 to 3).
**Figure 1:** Progress through stages of systematic review

(More than one report of any publication status can refer to a given trial; RCT=randomized trial.)

- Potentially relevant citations identified and screened for possible retrieval (n=486)
- Citations excluded via broad screening, with reasons (n=262): review (n=114); inappropriate intervention or treatment (n=120); focus on mechanism of action (n=28)
- Reports retrieved for more detailed relevance assessment (n=224)
- Reports excluded via relevance assessment, with reasons (n=159): not a primary study (e.g., review) (n=65); not an RCT (n=11); inappropriate population (n=1); inappropriate outcome (n=40); extraneous focus (e.g., prediction study) (n=24); incorrect drug (n=2); report never found (n=1); data integration involving multiple RCTs, yet not all RCTs defined or identified sufficiently to preclude duplicate entry into systematic review (n=15)
- Reports (n=65) describing unique RCTs (n=17) entered into qualitative assessment and eligible for inclusion in meta-analysis
- RCTs excluded from meta-analysis, with reasons: no data could be pooled (n=2)
- RCTs with poolable data, by outcome (Tables 4, 6 to 8 and Figures 2 to 5)
**Table 1:** Population characteristic defining the included raloxifene studies by trial type based on prevalent vertebral fractures and BMD at baseline

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Sample Size, Mean (SD), Range</th>
<th>Age, Mean (SD), Range in Years</th>
<th>Number of Years Post-Menopause Mean (SD), Range</th>
<th>Primary Diagnostic Criterion Used to Identify Population (Mode)</th>
<th>Geographic Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary prevention trials, each with all pts osteoporotic* (n=2)</td>
<td>3,924 (5,347)</td>
<td>67.45 (1.34)</td>
<td>20.95 (2.33)</td>
<td>BMD or VFx (2)</td>
<td>North America and Europe</td>
</tr>
<tr>
<td></td>
<td>143 to 7,705</td>
<td>66.5 to 68.4</td>
<td>19.3 to 22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary prevention trials, some with fewer than all pts osteoporotic*</td>
<td>1,408.5 (3,086.1)</td>
<td>65.1 (6.6)</td>
<td>13.6 (7.9)</td>
<td>BMD or VFx (3), BMD (3)</td>
<td>North America and Europe</td>
</tr>
<tr>
<td>(n=6)</td>
<td>51 to 7,705</td>
<td>56 to 75</td>
<td>2 to 22.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=5/6)</td>
<td>(n=5/6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary prevention trials, each with all pts osteopenic** or “healthy”</td>
<td>316.4 (326.9)</td>
<td>56.0 (4.4)</td>
<td>5.9 (3.1)</td>
<td>“Healthy” (7)</td>
<td>North America, Europe and Asia</td>
</tr>
<tr>
<td>(n=11)</td>
<td>33 to 1,145</td>
<td>52.9 to 59.3</td>
<td>2 to 11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=10/11)</td>
<td>(n=7/11)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary prevention trials, some with fewer than all pts osteopenic** or “healthy” (n=15)</td>
<td>272.2 (292.2)</td>
<td>57.0 (4.8)</td>
<td>7.2 (4.8)</td>
<td>“Healthy” (7), BMD (7)</td>
<td>North America, Europe and Asia</td>
</tr>
<tr>
<td></td>
<td>33 to 1,145</td>
<td>52.9 to 75</td>
<td>2 to 18.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n=13/15)</td>
<td>(n=10/15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed primary and secondary prevention trials,</td>
<td>150.8 (123.7)</td>
<td>63.9 (8.2)</td>
<td>9.9 (7.0)</td>
<td>BMD (4)</td>
<td>Europe and Asia</td>
</tr>
<tr>
<td>each with osteoporotic and non-osteoporotic</td>
<td>51 to 330</td>
<td>56 to 75</td>
<td>2 to 18.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i.e. osteopenic or “healthy”) pts (n=4)</td>
<td>(n=3/4)</td>
<td>(n=3/4)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SD=standard deviation; pts=participants; BMD=bone mineral density values (e.g. WHO definitions); VFx=prevalent vertebral fracture.

*Osteoporotic defined as BMD<2.5 SD below young adult mean.

**Osteopenic defined as 2.5 SD<BMD<1 SD below mean.
Table 2: Intervention characteristics defining included raloxifene studies, by trial type

<table>
<thead>
<tr>
<th>Intervention Length:</th>
<th>Single RLX Dose mg/day (Mode)</th>
<th>RLX Dose Level Contrast mg/day (Mode)</th>
<th>Comparator (Mode)</th>
<th>Intervention: RLX (mg/day) versus Control Contrast (Mode)</th>
<th>Supplements Mandated Number (%); Type (Mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weeks Mean (SD), Range* Trials Lasting ≥12 months</td>
<td>RLX 60 (2), RLX 120 (2)</td>
<td>RLX 60 versus RLX 120 (2)</td>
<td>PB (2)</td>
<td>RLX 60 versus RLX 120 versus PB (2)</td>
<td>Calcium and vitamin D (2)</td>
</tr>
<tr>
<td>Secondary prevention trials, each with all pts osteoporotic* (n=2)</td>
<td>111.4 (83.9) 52 to 170.7* n=2</td>
<td></td>
<td></td>
<td>2 (100)</td>
<td>Calcium and vitamin D (2)</td>
</tr>
<tr>
<td>Secondary prevention trials, some with fewer than all pts osteoporotic (n=6)</td>
<td>69.5 (58.9) 12 to 170.7* n=4</td>
<td>RLX 60 (4)</td>
<td>RLX 60 versus RLX 120 (2)</td>
<td>PB (7)</td>
<td>Calcium and vitamin D (4)</td>
</tr>
<tr>
<td>Primary prevention trials, each with all pts osteopenic** or “healthy” (n=11)</td>
<td>62.6 (47.9) 8 to 156* n=7</td>
<td>RLX 60 (9)</td>
<td>RLX 60 versus RLX 150 (4)</td>
<td>PB (6) ERT (5)</td>
<td>Calcium and vitamin D (4)</td>
</tr>
<tr>
<td>Primary prevention trials, some with fewer than all pts osteopenic or “healthy” (n=15)</td>
<td>58.9 (45.1) 8 to 156* n=9</td>
<td>RLX 60 (12)</td>
<td>RLX 60 versus RLX 150 (5)</td>
<td>PB (9) ERT (6)</td>
<td>Calcium alone (6)</td>
</tr>
<tr>
<td>Mixed primary and secondary prevention trials, each with osteoporotic and non-osteoporotic (i.e. osteopenic or “healthy”) pts (n=4)</td>
<td>48.5 (40.5) 12 to 104* n=2</td>
<td>RLX 60 (3)</td>
<td>All unique</td>
<td>All unique</td>
<td>Calcium and vitamin D (2)</td>
</tr>
</tbody>
</table>

SD=standard deviation; pts=participants; RLX=raloxifene; PB=placebo; ERT=estrogen replacement therapy (CEE 0.625 mg/day); mode=measure of central tendency.
*Osteoporotic defined as BMD<2.5 SD below young adult mean.
**Osteopenic defined as 2.5 SD<BMD<1 SD below mean.
Table 3: Trial characteristics defining included raloxifene studies, by trial type

<table>
<thead>
<tr>
<th>Jadad Total Trial Quality Score, Mean (SD) Range</th>
<th>Trials with Jadad Total Trial Quality Score of 0 to 2 (Low) Number (%)</th>
<th>Trials with Maximum Jadad Total Trial Quality Score (=5), Number (%)</th>
<th>Trials with “Adequate” Allocation Concealment, Number (%)</th>
<th>Trials with “Unclear” Allocation Concealment, Number (%)</th>
<th>Trials Described by ≥1 Published Report,* Number (%)</th>
<th>Confirmed Eli Lilly Trials,** Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary prevention trials, each with all pts osteoporotic* (n=2)(^{49-63})</td>
<td>3.5 (0.71) 3 to 4</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td>2 (100)</td>
</tr>
<tr>
<td>Secondary prevention trials, some with fewer than all pts osteoporotic (n=6)(^{49-75})</td>
<td>2.7 (1.03) 1 to 4</td>
<td>2 (33.3)</td>
<td>0 (0)</td>
<td>1 (16.7)</td>
<td>5 (83.3)</td>
<td>4 (66.7)</td>
</tr>
<tr>
<td>Primary prevention trials, each with all pts osteopenic** or “healthy” (n=11)(^{47,48,75-110})</td>
<td>3.3 (1.10) 2 to 5</td>
<td>3 (27.3)</td>
<td>2 (18.2)</td>
<td>1 (9.1)</td>
<td>10 (90.9)</td>
<td>8 (72.7)</td>
</tr>
<tr>
<td>Primary prevention trials, some with fewer than all pts osteopenic or “healthy” (n=15)(^{47,48,64-75,75-110})</td>
<td>3.0 (1.13) 1 to 5</td>
<td>5 (33.3)</td>
<td>2 (13.3)</td>
<td>1 (6.7)</td>
<td>14 (93.3)</td>
<td>10 (66.7)</td>
</tr>
<tr>
<td>Mixed primary and secondary prevention trials, each with osteoporotic and non-osteoporotic (i.e. osteopenic or “healthy”) pts (n=4)(^{64-75})</td>
<td>2.3 (0.96) 1 to 3</td>
<td>2 (50)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>4 (100)</td>
<td>2 (50)</td>
</tr>
</tbody>
</table>

SD=standard deviation; pts=participants; *more than one report can refer to a given trial (Table 1); **confirmed by Eli Lilly’s Canadian representative.
*Osteoporotic defined as BMD<2.5 SD below young adult mean.
**Osteopenic defined as 2.5 SD<BMD<1 SD below mean.
These categories were not mutually exclusive. A trial could be included in more than one category, and categories 2 and 5 are subsets of categories 1 and 4 respectively.

For four trials, the patient populations were too heterogeneous to classify according to BMD criteria. One trial\(^{64-67}\) had 40% of participants at baseline with BMD levels greater than the threshold (2.5 SD below the peak young adult level). In two other studies, the BMD inclusion criterion was a level $>2$ SD below the mean value for young adults,\(^{68-71,74}\) As a result, both studies included women who were unclassifiable in this scheme. One study\(^{73-75}\) was not classifiable using the above criteria, because the BMD inclusion criterion was established between 3 SD below and 1 SD above the mean value.

Details of the similarities and differences in the trials for the five BMD categories are presented in Appendix 5. Because of the a priori plan to categorize trials by baseline BMD, the tables retain an organization based on five categories of baseline BMD. This system is not used in the main analysis, because most of the trials, with the exception of MORE and Luftkin, do not report vertebral fractures as an outcome. BMD, however, is retained as it is an important baseline patient characteristic.

Nine of the 17 trials that were identified\(^{47-83}\) included efficacy or safety data that could be pooled while the remaining eight\(^{75,84-110}\) reported efficacy data outside the focus of this review while providing relevant safety data. Each of the latter was a study of women with higher BMD levels. One study\(^{82,83}\) excluded explicit inclusion criteria and baseline population data.

Trials enrolling women with very low BMD differed from those that included women with higher BMD levels in terms of age and number of years post-menopause. Regarding age, there were trials with all women having very low BMD (mean=67.5 years) or fewer than all (65.1) participants with very low BMD versus trials with all women having higher BMD (56.0) or fewer than all with higher BMD (57.0). Regarding number of years post-menopause, there were trials with all women having very low BMD (mean=21 years) or fewer than all (13.6) participants with low BMD versus trials with all women having higher BMD (5.9) or fewer than all (7.2) participants with higher BMD (Table 1).

In addition to population-defined clinical heterogeneity, two\(^{68-70,101}\) and five studies\(^{68-70,81-83,101,108-110}\) respectively did not report data on the age of participants or the number of years post-menopause. Mixed BMD studies involved participants who were almost as old (63.9 years) as those in very low BMD studies; and who fell between purely low and purely higher BMD studies in the number of years post-menopause (9.9 years).

Both trials that reported vertebral fracture, the necessary outcome for inclusion in the meta-analysis, enrolled patients based on its prevalence and BMD levels. Three of the trials that included fewer than all participants with very low BMD (3/6) enrolled patients based on prevalent vertebral fracture, but they did not report it as an outcome. Three studies with mixed high and low BMD (3/6) used only BMD as an entry criterion (Table 1).
Most trials occurred in North America and Europe (Appendix 4, Table 1). There was insufficient information that could be used to summarize the racial composition of the samples.

c) Intervention characteristics

Length of follow-up

Data from the MORE trial described three and four years of follow-up. The primary predetermined endpoint was three years. Overall, 11 trials lasted at least 12 months.47-70,75-93,97-100,108-110 The trial type with the largest proportion of studies lasting more than 12 months was type I, which included those with all participants having very low BMD (2/2). Trials that included all (mean=111.4 weeks) or fewer than all participants (69.5 weeks) with very low BMD entailed longer therapy durations than both types of studies with higher BMD (62.6, 58.9 and 48.5 weeks for all, fewer than all and mixed BMD levels respectively) (Table 2).

Dose

There was consistency across all trials regarding the most popular single dose of raloxifene: 60 mg/day (Table 2), which is the approved dose in Canada and elsewhere. The most popular range of raloxifene doses was 60 versus 120 mg/day (mode for each=2) and 60 versus 150 mg/day for low BMD and higher BMD trials (modal values=4 and 5) respectively. Each trial with mixed BMD levels among participants at baseline used a unique range of raloxifene doses.

Comparator

The 17 identified RCTs include six versus placebo alone, four versus estrogen-progestin therapies; two versus estrogen alone (women with hysterectomies); one versus alendronate, one versus tamoxifen and three versus estrogen alone (no hysterectomy). No RCTs compared raloxifene to etidronate, risedronate or salmon calcitonin. Two active comparators tamoxifen and estrogen alone in women without hysterectomy are unapproved for osteoporosis prevention. Estrogen therapy was often used as a comparator in studies of women with higher BMD at baseline.

Supplements

Studies varied in terms of the types of mandated supplements. Calcium (range 500 to 1,000 mg/day) with vitamin D (range 300 to 800 IU/day) and calcium alone (range 400 to 600 mg/day) were given most often (Table 2). In one study, participants were asked not to take supplements.82,83 One trial involving patients with very low BMD73-75 and five trials involving women with higher BMD75,84-90,97-107 did not report whether supplements were used.

d) Trial quality

The mean Jadad total quality scores were comparable for trials with all participants diagnosed with very low BMD (mean=3.5) and those with low and higher BMD (mean=3.3) respectively. These indicated moderately high trial quality. Trials with mixed BMD levels had total scores that barely exceeded low quality (2.3). This trial type received a low Jadad total quality score (0 to 2) 50% (2/4) of the time. Trials involving women with higher BMD in either category were the only ones receiving a maximum or high (5) quality score. Two trials provided evidence that allocation to the intervention arms had been performed adequately.49-58,91-93 Most trials (15/17)
were rated as having unclear allocation concealment. Of the two trials reporting vertebral fracture as an outcome, the MORE trial was larger and exhibited higher methodological quality than the Lufkin trial (Appendix 4).

### 3.2.2 Quantitative results

The trial populations varied in terms of age and necessarily, the number of years post-menopause. Based on these two variables, we had planned subgroup analyses of efficacy (i.e., clinical fractures, vertebral fracture, BMD) and safety (mortality, serious adverse events, other adverse events). While 17 relevant trials were included in this systematic review, two provided data that were suitable for the planned meta-analysis, two provided data on the primary outcome of vertebral fracture, while others provided data on mild to moderate adverse effects.

#### a) Results of Main Analyses of Primary Outcome Data

**Symptomatic and asymptomatic vertebral fractures**

Three studies provided radiological and clinical data on vertebral fracture, all comparing raloxifene to placebo: the Meunier, MORE and Lufkin trials (Appendix 4, Table 4, Figure 2). The Meunier trial\(^{64-67}\) failed to report any vertebral fracture events, so it was excluded from the meta-analysis, leaving the MORE and Lufkin trials.

Using the meta-analyzed vertebral fracture results (MORE and Lufkin data), no benefit was seen in raloxifene recipients (for combined, low and high doses), even though a significant risk reduction was observed for all dose definitions when the MORE trial’s data at 40 or at 47.4 months were analyzed separately. Results obtained exclusively from the MORE trial showed a significant relative risk reduction (RRR) after three years \([0.59 (95\% \text{ CI} 0.51, 0.70)]\) and four years of intervention \([0.60 (0.52, 0.69)]\).

Given that the results of only two trials could be pooled, likely yielding an unreliable estimate of variation (and there was evidence of significant statistical heterogeneity), the random effects model was chosen as the correct statistical model. This model smoothed the weightings so that the larger trial’s contribution to the pooled result carried less weight.

While we report our findings in keeping with the pre-determined methodology, there is little justification for using a random effects model or for undertaking meta-analysis. The two trials that can be meta-analyzed differ in terms of number of randomized participants, patterns of incident vertebral fractures and trial quality when defined in two ways (Appendix 4). Thus, by virtue of its greater sample size and higher quality, the MORE study should “drive” the meta-analysis via a fixed effects model. Statistically, by using a fixed effects model, the larger trial (MORE: \(n=7,705\)) receives a larger weighting than a smaller trial (Lufkin: \(n=143\)). As a result, evidence from the MORE trial regarding raloxifene’s placebo-controlled efficacy in the prevention of vertebral fractures\(^{14}\) is revealed.

Certain factors may have played a role in producing the discordance across the vertebral fracture-related results between the Lufkin and MORE trials, with the most important factor being the definition of the vertebral fracture outcome.\(^{14}\) In the MORE trial, a vertebral fracture required confirmation via at least two of three types of assessments: two independent semi-quantitative readings (i.e., grade 0=None, 1=mild, 2=moderate, 3=severe) and one quantitative assessment.
(i.e., fracture=decrease of at least 20% and 4 mm via radiographic assessment). The Lufkin study defined an incident vertebral fracture as either a $\geq 15\%$ or a $\geq 30\%$ decrease in vertical height respectively. The results were not significant at the $\geq 15\%$ cutoff.

**Table 4:** Primary outcome and subgroup analyses of vertebral fracture data against placebo

<table>
<thead>
<tr>
<th>Dose</th>
<th>Number of Trials/ Number of Participants</th>
<th>Relative Risk of (95% CI)* Vertebral Fractures</th>
<th>Statistical Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>1/7,705$^1$</td>
<td>0.59 (0.51, 0.70)****</td>
<td>n/a</td>
</tr>
<tr>
<td>Low</td>
<td>1/5,133$^1$</td>
<td>0.65 (0.54, 0.79)****</td>
<td>n/a</td>
</tr>
<tr>
<td>High</td>
<td>1/5,148$^1$</td>
<td>0.54 (0.44, 0.65)****</td>
<td>n/a</td>
</tr>
<tr>
<td>Combined</td>
<td>2/7,848$^2$</td>
<td>0.80 (0.42, 1.53)******</td>
<td>p=0.00</td>
</tr>
<tr>
<td>Low</td>
<td>2/5,229$^2$</td>
<td>0.84 (0.48, 1.46)******</td>
<td>p=0.03</td>
</tr>
<tr>
<td>High</td>
<td>2/5,243$^2$</td>
<td>0.75 (0.36, 1.57)******</td>
<td>p=0.01</td>
</tr>
<tr>
<td>Combined</td>
<td>1/7,705$^3$</td>
<td>0.60 (0.52, 0.69)****</td>
<td>n/a</td>
</tr>
<tr>
<td>Low</td>
<td>1/5,133$^3$</td>
<td>0.64 (0.54, 0.75)****</td>
<td>n/a</td>
</tr>
<tr>
<td>High</td>
<td>1/5,148$^3$</td>
<td>0.57 (0.48, 0.68)****</td>
<td>n/a</td>
</tr>
<tr>
<td>Combined</td>
<td>2/7,848$^4$</td>
<td>0.81 (0.43, 1.51)******</td>
<td>p=0.01</td>
</tr>
<tr>
<td>Low</td>
<td>2/5,229$^4$</td>
<td>0.82 (0.46, 1.48)******</td>
<td>p=0.02</td>
</tr>
<tr>
<td>High</td>
<td>2/5,243$^4$</td>
<td>0.77 (0.40, 1.51)******</td>
<td>p=0.01</td>
</tr>
</tbody>
</table>

n/a=not applicable; CI=confidence interval; combined=(low dose=RLX<60 mg/day) plus (high dose=RLX≥120 mg/day); *for relative risk, a CI encompassing a value of 1 indicates statistically non-significant result; **random effects model; ***excluding Meunier trial$^{64-67}$ since each arm of study had zero events; ****fixed effects model; $^1$MORE trial at three years; $^2$MORE trial at three years plus Lufkin trial; $^3$MORE trial at four years; $^4$MORE trial at four years plus Lufkin trial.

The investigators claimed that with a $\geq 30\%$ cutoff, there was a dose-related reduction in vertebral fracture for the raloxifene groups versus placebo. Neither the Lufkin trial reports nor this systematic review team’s contact at the manufacturer provided $\geq 30\%$ cutoff data expressed in a manner (i.e., number of participants with at least one incident vertebral fracture) amenable to pooling with MORE data. The $\geq 30\%$ cutoff seemed to be a post hoc measure of questionable validity, given the lack of difference observed at the predefined $\geq 15\%$ cutoff.

A second factor that likely led to discordance between the vertebral fracture results in the Lufkin and MORE trials was the differences in study populations (Appendix 4). The subjects in the Lufkin trial were older than the women in the largest subgroup enrolled in the MORE trial: women without baseline vertebral fracture (mean of 68 years in the Lufkin trial versus 65 years in this subgroup). There was little difference in age between the patients in the Lufkin trial and those in the MORE subgroup with pre-existing vertebral fractures. Women in the MORE trial with vertebral fractures at baseline were significantly more likely to experience two or more incident vertebral fractures over three years. This supported the view that a history of vertebral fractures was an independent risk factor for future fractures, regardless of BMD.$^{12,17,20,111-114}$ The Lufkin trial did not report on the proportion of women with baseline vertebral fracture, so it was impossible to know whether the populations were similar in this respect.
Figure 2: Risk of vertebral fracture against placebo

Combined=(low dose=RLX≤60 mg/day) plus (high dose=RLX≥120 mg/day); open circle=fixed effects model; solid circle=random effects model; p values for statistical heterogeneity: ¹p=0.00, ²p=0.03, ³p=0.01, ⁴p=0.02.
Both studies used a low dose defined as 60 mg/day and a high dose not exceeding 120 mg/day. The higher doses did not provide an advantage over what was observed at the 60 mg/day dose. Both trials included calcium and vitamin D supplements, but these were provided at different doses (Table 2).

In light of these definitions of clinical heterogeneity and the observation that the MORE trial was larger than and exhibited higher methods-related quality than did the Lufkin trial (Appendix 4), additional attention was paid to MORE results and additional data pertaining to the primary outcome were described, including findings reported by reviewers at the US Food and Drug Administration (FDA).

**Key efficacy outcomes: MORE trial**
The MORE trial was of sufficient size and duration so that it could assess the clinical outcomes of raloxifene therapy, including symptomatic vertebral fracture and safety outcomes, such as mortality and serious adverse events. Table 5 shows the results concerning vertebral fracture.

The MORE trial alone showed a significant relative risk reduction at three years for 60 mg raloxifene per day [0.65 (0.54, 0.79)] and 120 mg raloxifene [0.54 (0.44, 0.65)] compared with placebo; and at four years for 60 mg raloxifene [0.64 (0.54, 0.75)] and 120 mg raloxifene [0.57 (0.48, 0.68)] respectively. The greater risk reduction, compared with placebo, was associated with the higher dose of 120 mg/day. The risk reduction in both groups did not differ significantly. As with the main analyses, all dose-related analyses of pooled results were associated with statistical heterogeneity (p=0.01 to 0.03).

**Table 5: Vertebral effects of raloxifene versus placebo over three years of follow-up**

<table>
<thead>
<tr>
<th></th>
<th>Placebo (N*=2,576)</th>
<th>Raloxifene (pooled) (N=5,129)</th>
<th>ARR</th>
<th>NNT/NNH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptomatic (clinical) vertebral fracture (published trial report)*</td>
<td>35 to 36*(1.4%)</td>
<td>29 to 30 (0.6%)</td>
<td>0.8%</td>
<td>~125 (NNT)</td>
</tr>
<tr>
<td>Symptomatic (clinical) vertebral fracture (FDA medical review)†</td>
<td>3.1%</td>
<td>1.8% 60 mg</td>
<td>1.3%</td>
<td>77 (NNT)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.5% 120 mg</td>
<td>1.6%</td>
<td>63 (NNT)</td>
</tr>
</tbody>
</table>

*Published report states that total number of women with symptomatic fractures was 65, RR=0.4 (95% CI 0.3 to 0.7) for raloxifene users. Using a 2X2 table in Epi Info, this is estimated to be 35 to 36 cases in placebo group; 29 to 30 in pooled raloxifene groups (number per treatment arm not presented in report).
†FDA review reports on vertebral fracture in 2,292 women on placebo, 2,259 on raloxifene 60 mg and 2,277 on raloxifene 120 mg, reflecting number of women with x-rays at baseline and endpoint.

ARR=absolute risk reduction; ARI=absolute risk increase; NN=numbers needed to treat; NNH=numbers needed to harm.

In the MORE trial, women with prevalent vertebral fractures at baseline (study group 2) had higher rates of incident vertebral fracture than women with BMD ≥2.5 SD below the mean. This difference was greater than that observed between placebo and raloxifene users; at three years of follow-up, 4.5% of women with BMD ≥2.5 SD below the mean on placebo had experienced incident (asymptomatic or symptomatic) vertebral fracture, compared with 14.7% of women on raloxifene 60 mg/day with vertebral fracture at baseline. The magnitude of this difference lent weight to the importance of the prevalence of vertebral fracture as a predictor of future incident vertebral fractures.
There were no statistically significant differences between placebo and raloxifene users in the proportion of patients with at least one incident non-vertebral fractures: 9.3% on placebo versus 8.5% on raloxifene.

**Key efficacy outcomes: Lufkin trial**

Raloxifene did not significantly alter the probability of incident morphometric vertebral fracture (defined as ≥15% reduction in height). One or more incident vertebral fractures were observed on x-ray in 40% of women on placebo and 48% of women on all doses of raloxifene (49% at 60 mg/day). The high frequency of incident vertebral fractures in one year, on placebo and on raloxifene, suggested that this patient population also had a high prevalence of vertebral fractures at baseline. There were no wrist fractures; one hip fracture occurred in a woman on raloxifene 120 mg/day.

**a) Subgroup and sensitivity analyses of primary outcome data**

With only two trials providing vertebral fracture data that could be pooled, we could not perform the planned subgroup and sensitivity evaluations of primary outcome data; and we did not pool the Lufkin and MORE data using samples of women who had high and low BMD at baseline.49

**b) Results of analyses of secondary outcome data**

Both primary and secondary prevention studies provided analyzable BMD data. The only comparator was placebo. Raloxifene had a significant and positive impact on BMD, with some variation by degree, for all trials regardless of BMD level at enrolment (Table 6, Figure 3).

### Table 6: Secondary outcome analyses of bone mineral density (BMD) data: against placebo, organized by trial type

<table>
<thead>
<tr>
<th>BMD Sites Outcome (g/cm²)</th>
<th>Trial Type*†</th>
<th>Number of Trials/ Number of Participants</th>
<th>Weighted Mean** (95% CI)</th>
<th>Statistical Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femoral neck BMD</td>
<td>Heterog Tx</td>
<td>3/7,997‡, a</td>
<td>2.26 (1.93, 2.59)</td>
<td>p=0.34</td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>4/1,021‡, b</td>
<td>2.05 (1.51, 2.58)***</td>
<td>p=0.69</td>
</tr>
<tr>
<td></td>
<td>Homog Px</td>
<td>2/729‡</td>
<td>2.23 (1.59, 2.86)***</td>
<td>p=0.66</td>
</tr>
<tr>
<td></td>
<td>Mixed Tx/Px</td>
<td>2/292‡, b</td>
<td>1.62 (0.62, 2.61)***</td>
<td>p=0.63</td>
</tr>
<tr>
<td>Hip total BMD</td>
<td>Heterog Tx</td>
<td>2/272‡</td>
<td>1.72 (0.92, 2.52)***</td>
<td>p=0.43</td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>3/1,402†</td>
<td>2.13 (1.71, 2.56)***</td>
<td>p=0.59</td>
</tr>
<tr>
<td></td>
<td>Homog Px</td>
<td>2/1,273†</td>
<td>2.16 (1.70, 2.62)***</td>
<td>p=0.32</td>
</tr>
<tr>
<td>Lumbar spine BMD</td>
<td>Heterog Tx</td>
<td>3/7,997‡, a</td>
<td>2.62 (2.44, 2.81)***</td>
<td>p=0.49</td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>3/1,437†</td>
<td>2.37 (1.92, 2.82)***</td>
<td>p=0.57</td>
</tr>
</tbody>
</table>

CI=confidence interval; Tx=treatment; Px=prevention; heterog Tx=Tx trials with not all pts diagnosed osteoporotic; homog Tx=Tx trials with all pts diagnosed osteoporotic; heterog Px=Px trials with not all pts diagnosed osteopenic or “healthy”; homog Px=Px trials with all pts diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which when added to trials with homogeneous composition of participants defined via WHO BMD criteria, yields designation of “heterog trials”; combined=all doses=(low dose=RLX<60 mg/day) plus (high dose=RLX≥120 mg/day); **trial type=treatment versus prevention trial, based on WHO’s BMD defined typology (i.e. osteoporotic versus osteopenic versus healthy); ***difference in mean percent change from baseline; †with little statistical heterogeneity, random effects estimate is identical to fixed effects estimate; 1MORE trial at three years plus Meunier64-66 and Johnell68-70 trials; 2Meunier64-66 and Johnell68-70; 3Johnston47,48,76-80 and Pavo82,83 trials; 4Johnston47,48,76-80 and Pavo82,83; 5Lufkin59-63 and Meunier64-67 trials; 6Meunier64-67 and Johnell68-70; 7Johnston47,48,76-80 and Pavo82,83; 8Johnston47,48,76-80 and Pavo82,83; 9variance imputation required for MORE trial at three years and Johnell68-70 trial; bvariance imputation required for Johnell68-70 trial.
Figure 3: Bone mineral density against placebo, by trial type*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Trial type</th>
<th># trials /</th>
<th># participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Femoral neck</td>
<td>Heterog Tx</td>
<td>3 / 7997</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>4 / 1021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homog Px</td>
<td>2 / 729</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mixed Tx/Px</td>
<td>2 / 292</td>
<td></td>
</tr>
<tr>
<td>Hip total</td>
<td>Heterog Tx</td>
<td>2 / 272</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>3 / 1402</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Homog Px</td>
<td>2 / 1273</td>
<td></td>
</tr>
<tr>
<td>Lumbar spine</td>
<td>Heterog Tx</td>
<td>3 / 7997</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>3 / 1437</td>
<td></td>
</tr>
</tbody>
</table>

*All analyses involved combined RLX doses; combined=(low dose=RLX≤60 mg/day) plus (high dose=RLX≥120 mg/day); open circle=fixed effects model; solid circle=random effects model; Tx=treatment; Px=prevention; heterog Tx=Tx trials with not all participants diagnosed osteoporotic; homog Tx=Tx trials with all participants diagnosed osteoporotic; heterog Px=Px trials with not all participants diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which, when added to trials with homogeneous composition of participants defined via WHO BMD criteria, yields designation of “heterog” trials.
A positive effect on femoral neck BMD was observed for all trial types, with the greatest increase seen in populations where most or all women had very low BMD at enrolment [2.26 (1.93, 2.59)]. The next greatest benefit occurred in populations with BMD that was uniformly higher than the 2.5 SD threshold [2.23 (1.59, 2.86)]. Trials with mixed BMD levels at baseline showed an intermediary effect [1.62 (0.62, 2.61)].

For hip BMD, raloxifene showed a statistically significant impact. The greatest increase was observed for patient populations with uniformly very low BMD [2.16 (1.70, 2.62)]. The next greatest increase was for the category of trials with low BMD mixed with trials that include higher BMD levels, Px [2.13 (1.71, 2.56)]. The weakest effect was for the heterogeneous BMD trial types [1.72 (0.92, 2.52)].

For lumbar spine BMD, raloxifene exhibited a significant, positive effect, with the strongest impact associated with the category that mixed very low BMD with higher BMD [2.62 (2.44, 2.81)]. The next greatest impact came from trials in the mixed BMD trial types [2.30 (1.27, 3.34)]. These investigations did not reveal instances of significant (p<0.10) statistical heterogeneity.

c) Results of analyses of mild to moderate adverse events (tolerability)

Meta-analysis was planned to assess data on all adverse events. The most complete data were provided by the MORE trial alone. Given its significant place in the literature as defined by its size, methods-related quality and intervention length, MORE safety results were presented separately, after the results that could be pooled. This strategy established a larger perspective on raloxifene’s efficacy and safety than was observed by looking exclusively at pooled results.

Meta-analysis was used according to the a priori protocol to evaluate adverse event data from the 17 included trials. The analyzable binary (Table 7, Figure 4) and continuous adverse event data (Table 8, Figure 5) involved three types of comparators (placebo, estrogen and combined hormone therapy) and four of five trial types (i.e., no mixed BMD trials). Only studies in populations of women with higher BMD levels at baseline, i.e., less than 2.5 SD below peak BMD, provided analyzable data when estrogen and combined hormone therapy were the comparators.

On three occasions, a subgroup analysis by dose yielded a finding that changed the consideration of a finding as statistically significant. Each involved binary data, a low raloxifene dose and placebo as the comparator.

Results were presented by adverse event type; and organized by comparator and trial type. Observations from analyses of binary data preceded those involving continuous data. Assessed against each comparator (i.e., placebo, estrogen, combined hormone therapy), there was a significantly greater risk of hot flashes associated with raloxifene use (Table 7, Figure 4). For example, when compared with placebo, the greatest relative risk was seen in trials with very low BMD. The relative risk at 40 months was 1.65 (1.40, 1.94) and at 47.4 months was 1.61 (1.38, 1.88).
Table 7: Analyses of binary adverse event data; \(^+\) organized by comparator and trial type

<table>
<thead>
<tr>
<th>Adverse Event Outcome</th>
<th>Trial Type*</th>
<th>Dose</th>
<th>Number of Trials/ Number of Participants</th>
<th>Relative Risk (95% CI)**</th>
<th>Statistical Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Against placebo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Heterog Tx</td>
<td>Combined</td>
<td>2/7,834(^1)</td>
<td>1.65 (1.40, 1.94)*** P=0.57</td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Heterog Tx</td>
<td>Combined</td>
<td>2/7,834(^2)</td>
<td>1.61 (1.38, 1.88)*** P=0.60</td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Heterog Px</td>
<td>Combined</td>
<td>5/2,070(^3)</td>
<td>1.29 (1.03, 1.61)*** P=0.93</td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Heterog Px</td>
<td>Low</td>
<td>4/1,350(^4)</td>
<td>1.14 (0.88, 1.47)*** P=0.98</td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Homog Px</td>
<td>Combined</td>
<td>4/1,941(^5)</td>
<td>1.30 (1.03, 1.63)*** P=0.83</td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Homog Px</td>
<td>Low</td>
<td>3/1,263(^6)</td>
<td>1.15 (0.88, 1.49)*** P=0.96</td>
<td></td>
</tr>
<tr>
<td>Breast pain and Soreness</td>
<td>Homog Px</td>
<td>Combined</td>
<td>5/2,121(^7)</td>
<td>0.64 (0.31, 1.32) P=0.18</td>
<td></td>
</tr>
<tr>
<td>Leukorrhea</td>
<td>Homog Px</td>
<td>Combined</td>
<td>2/1,460(^8)</td>
<td>2.13 (0.84, 5.39)*** P=0.67</td>
<td></td>
</tr>
<tr>
<td>Vaginal bleeding</td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/1,754(^9)</td>
<td>0.57 (0.33, 0.97)*** P=0.69</td>
<td></td>
</tr>
<tr>
<td>Vaginitis</td>
<td>Homog Px</td>
<td>Low</td>
<td>3/1,263(^10)</td>
<td>0.67 (0.37, 1.19)*** P=0.46</td>
<td></td>
</tr>
<tr>
<td>Endometrial proliferation</td>
<td>Homog Px</td>
<td>Combined</td>
<td>2/502(^11)</td>
<td>0.38 (0.19, 0.79) P=0.17</td>
<td></td>
</tr>
<tr>
<td>Retinal vein</td>
<td>Heterog Tx</td>
<td>Combined</td>
<td>2/7,834(^12)</td>
<td>0.61 (0.18, 2.09)*** P=0.36</td>
<td></td>
</tr>
<tr>
<td>Discontinuations</td>
<td>Heterog Tx</td>
<td>Combined</td>
<td>2/7,834(^13)</td>
<td>1.22 (1.05, 1.42)*** P=0.54</td>
<td></td>
</tr>
<tr>
<td>(adverse event)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Against estrogen replacement therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Homog Px</td>
<td>Combined</td>
<td>2/493(^14)</td>
<td>2.96 (1.31, 6.69) P=0.30</td>
<td></td>
</tr>
<tr>
<td>Breast pain and Soreness</td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/660(^15)</td>
<td>0.19 (0.11, 0.32) P=0.24</td>
<td></td>
</tr>
<tr>
<td>Discontinuations</td>
<td>Heterog Px</td>
<td>Combined</td>
<td>2/238(^16)</td>
<td>0.29 (0.12, 0.73)*** P=0.51</td>
<td></td>
</tr>
<tr>
<td>(no reason reported)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Against hormone replacement therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hot flashes</td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/544(^17)</td>
<td>6.77 (3.39, 13.52)*** P=0.71</td>
<td></td>
</tr>
<tr>
<td>Breast pain and Soreness</td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/544(^18)</td>
<td>0.14 (0.07, 0.25) P=0.30</td>
<td></td>
</tr>
<tr>
<td>Vaginal bleeding</td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/544(^19)</td>
<td>0.11 (0.07, 0.15)*** P=0.75</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) includes three subgroup analysis results (each for low dose) where a statistically significant “combined” finding (i.e. without distinguishing by dose magnitude) became a non-significant one when dose magnitude was investigated; CI=confidence interval; pts=participants; Tx=treatment; Px=prevention; heterog Tx=Tx trials with not all pts diagnosed osteoporotic; homog Tx=Tx trials with all pts diagnosed osteoporotic; heterog Px=Px trials with not all pts diagnosed osteopenic or “healthy”; homog Px=Px trials with all pts diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which when added to trials, with homogeneous composition of participants defined via WHO BMD criteria, yields designation of “heterog” trials; combined = all doses = (low dose=RLX\(<60 \text{ mg/day}) \text{ plus (high dose=RLX}\geq120 \text{ mg/day}); *trial type = treatment versus prevention trial, based on WHO’s BMD defined typology (i.e. osteoporotic versus osteopenic versus healthy); **for relative risk, a CI encompassing a value of “1” indicates statistically non-significant result; ***with little statistical heterogeneity, random effects estimate is identical to fixed effects estimate; \(^1\) MORE trial at three years and Meunier\(^{64-67}\) trial; \(^2\) MORE trial at four years and Meunier\(^{64-67}\) trial; \(^3\) Meunier\(^{64-67}\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) Boss\(^{94-96}\) and Walsh\(^{105-107}\) trials; \(^4\) Meunier\(^{64-67}\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) Walsh\(^{105-107}\) trials; \(^5\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) Goldstein\(^{91-93}\) Boss\(^{94-96}\) and Walsh\(^{105-107}\) trials; \(^6\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) Goldstein\(^{91-93}\) Boss\(^{94-96}\) and Walsh\(^{105-107}\) trials; \(^7\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) Goldstein\(^{91-93}\) Boss\(^{94-96}\) and Walsh\(^{105-107}\) trials; \(^8\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials; \(^9\) Johnston\(^{47,48,76-80}\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials; \(^10\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials; \(^11\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials; \(^12\) Prestwood\(^{73-75}\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials; \(^13\) Prestwood\(^{73-75}\) Goldstein\(^{91-93}\) and Boss\(^{94-96}\) and Freedom\(^{97,100}\) and Walsh\(^{105-107}\) trials.
Figure 4: Risk of adverse events (binary), by comparator and trial type

Combined=(low dose=RLX≤60 mg/day) plus (high dose=RLX≥120 mg/day); open circle=fixed effects model; solid circle=random effects model; ERT=estrogen replacement therapy; HRT=hormone replacement therapy; pts=participants; pr=proliferation; thr=thrombosis; heterog Tx=Tx trials with not all participants diagnosed osteoporotic; homog Tx=Tx trials with all participants diagnosed osteoporotic; heterog Px=Px trials with not all participants diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which when added to trials with homogeneous composition of participants defined via WHO BMD criteria, yields designation of “heterog” trials; 1discontinuations due to adverse events; 2discontinuations with no reason reported.
**Table 8:** Analyses of continuous adverse event data; organized by comparator and trial type

<table>
<thead>
<tr>
<th>Adverse Event Outcome (TVU)</th>
<th>Trial Type*</th>
<th>Dose</th>
<th>Number of Trials/ Number of Participants</th>
<th>Weighted Mean** (95% CI)</th>
<th>Statistical Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometrial thickness</td>
<td>Heterog Tx</td>
<td>Combined</td>
<td>2/7,834&lt;sup&gt;1, b&lt;/sup&gt;</td>
<td>0.25 (0.08, 0.43)</td>
<td>p=0.27</td>
</tr>
<tr>
<td></td>
<td>Homog Px</td>
<td>Combined</td>
<td>3/933&lt;sup&gt;2, b&lt;/sup&gt;</td>
<td>0.00 (-0.33, 0.33)</td>
<td>p=0.23</td>
</tr>
<tr>
<td></td>
<td>Heterog Px</td>
<td>Combined</td>
<td>4/1,062&lt;sup&gt;3, b&lt;/sup&gt;</td>
<td>-0.01 (-0.22, 0.20)***</td>
<td>p=0.40</td>
</tr>
</tbody>
</table>

CI=confidence interval; TVU=transvaginal ultrasonography; Tx=treatment; Px=prevention; heterog Tx=Tx trials with not all pts diagnosed osteoporotic; homog Tx=Tx trials with all pts diagnosed osteoporotic; heterog Px=Px trials with not all pts diagnosed osteopenic or “healthy”; homog Px=Px trials with all pts diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which when added to trials with an homogeneous composition of pts defined via WHO BMD criteria, yields a designation of “heterog trials”; combined = all doses = (low dose=RLX<60 mg/day) plus (high dose=RLX>120 mg/day); *trial type=treatment versus prevention trial, based on WHO’s BMD defined typology (i.e. osteoporotic versus osteopenic versus healthy); **change from baseline; ***with little statistical heterogeneity, random effects estimate is identical to fixed effects estimate; MORE trial at three years and Meunier<sup>64-67</sup> trial; Johnston<sup>47,48,76-80</sup>, Goldstein<sup>91-93</sup> and Vardy<sup>101</sup> trials; ‘Meunier, 64-67 Johnston, 47,48,76-80 Goldstein<sup>91-93</sup> and Vardy<sup>101</sup> trials; ‘Goldstein<sup>91-93</sup> and Vardy<sup>101</sup> trials; ‘variance imputation required for MORE trial at three years; ‘variance imputation required for Johnston<sup>47,48,76-80</sup> and Vardy<sup>101</sup> trials; ‘variance imputation required for Vardy<sup>101</sup> trial.

**Figure 5:** Endometrial thickness (continuous) by comparator and trial type*

---

*All analyses involved combined raloxifene (RLX) doses; combined=(low dose=RLX≤60 mg/day) plus (high dose=RLX ≥120 mg/day); open circle=fixed effects model; solid circle=random effects model; Tx=treatment; Px=prevention; heterog Tx=Tx trials with not all participants diagnosed osteoporotic; homog Tx=Tx trials with all participants diagnosed osteoporotic; heterog Px=Px trials with not all participants diagnosed osteopenic or “healthy”; mixed Tx/Px=heterogeneously composed trials, which when added to trials with homogeneous composition of participants defined via WHO BMD criteria, yields designation of “heterog” trials.
The lowest relative risk of adverse events was associated with trials involving women with higher BMD levels at baseline (younger women with fewer years post-menopause). This effect failed to achieve significance for low doses of raloxifene. Overall, the effect regarding hot flashes was greatest when combined hormone therapy was the comparator [6.77 (3.39, 13.52)]. The relative impact of raloxifene versus estrogen [2.96 (1.31, 6.69)] was closer to the relative effect of raloxifene versus placebo.

When compared with placebo, there was an increased risk of discontinuation due to adverse events associated with raloxifene [1.22 (1.05, 1.42)]. There was a decreased risk of endometrial proliferation associated with raloxifene use when compared with placebo [0.38 (0.19, 0.79)] or when fewer than all the postmenopausal women were diagnosed as either osteopenic or healthy [0.06 (0.04, 0.09)]. A risk reduction for hyperplasia was also observed when raloxifene was compared with estrogen alone in women with intact uteri [0.04 (0.00, 0.83)]. As estrogen alone is contraindicated in women with intact uteri because of the associated risk of endometrial cancer, this comparison provides little assistance in a safety assessment of raloxifene. Increases in endometrial thickness and hyperplasia are measured as potential indicators of risk of endometrial cancer. Thus, estrogen alone is also an inappropriate comparator for these outcomes. The results from the MORE trial suggest that when compared to placebo, raloxifene may be protective.

No difference in the incidence of breast pain and soreness events was observed when placebo and raloxifene were compared. A decreased risk of these events was associated with raloxifene compared with estrogen [0.19 (0.11, 0.32)] or combined hormone therapy [0.14 (0.07, 0.25)]. A decreased risk of episodes of vaginal bleeding was associated with raloxifene versus placebo [0.57 (0.33, 0.97)] or combined hormone therapy [0.11 (0.07, 0.15)]. The larger benefit was observed when combined hormone therapy was the control. Where placebo was the comparator, the initial significant effect disappeared for the low raloxifene dose [0.67 (0.37, 1.19)]. No differences were found with respect to episodes of leukorrhea, vaginitis or retinal vein thrombosis when raloxifene was compared to placebo.

Continuous data (Table 8, Figure 5) revealed that, relative to placebo, raloxifene use was associated with significantly greater increase in endometrial thickness [0.25 (0.08, 0.43)]. This finding occurred in trials where not all participants had very low BMD at baseline.

No evidence of significant statistical heterogeneity was observed for meta-analyses of adverse events.

**Adverse events: MORE trial**

Table 9 describes the adverse events that occurred in >2% of patients who participated in the MORE trial.
Table 9: Adverse events experienced by at least 2% of participants in MORE trial*

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>Placebo N=2,576</th>
<th>Raloxifene 60 mg/d N=2,557</th>
<th>Raloxifene 120 mg/d N=2,572</th>
<th>Raloxifene (pooled) versus Placebo p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Influenza syndrome</td>
<td>293 (11.4%)</td>
<td>346 (13.5%)</td>
<td>345 (13.4%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Hot flashes (vasodilation)</td>
<td>165 (6.4%)</td>
<td>249 (9.7%)</td>
<td>269 (11.6%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Leg cramps</td>
<td>96 (3.7%)</td>
<td>178 (7.0%)</td>
<td>178 (6.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Peripheral edema</td>
<td>114 (4.4%)</td>
<td>134 (5.2%)</td>
<td>168 (6.5%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Endometrial cavity fluid§</td>
<td>43 (5.7%)</td>
<td>60 (8.1%)</td>
<td>66 (8.7%)</td>
<td>0.02</td>
</tr>
<tr>
<td>Diabetes</td>
<td>14 (0.5%)</td>
<td>31 (1.2%)</td>
<td>28 (1.1%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Hypertension</td>
<td>231 (9.0%)</td>
<td>177 (6.9%)</td>
<td>194 (7.5%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>121 (4.7%)</td>
<td>55 (2.2%)</td>
<td>50 (1.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hematuria</td>
<td>55 (2.1%)</td>
<td>35 (1.4%)</td>
<td>33 (1.3%)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Bradycardia</td>
<td>30 (1.2%)</td>
<td>13 (0.5%)</td>
<td>17 (0.7%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>4 (0.2%)</td>
<td>4 (0.2%)</td>
<td>2 (&lt;0.1%)</td>
<td>0.67</td>
</tr>
<tr>
<td>Breast pain</td>
<td>65 (2.5%)</td>
<td>61 (2.4%)</td>
<td>70 (2.7%)</td>
<td>0.94</td>
</tr>
<tr>
<td>Vaginal bleeding</td>
<td>62 (3.1%)</td>
<td>67 (3.4%)</td>
<td>56 (2.8%)</td>
<td>0.99</td>
</tr>
</tbody>
</table>

*40 months of follow-up
§among 2,262 women who had transvaginal ultrasonography

**d)** **Results of serious adverse events**
Serious adverse events were reported in the MORE publications and more completely in the medical review by the US FDA (Table 10).

Table 10: Adverse effects of raloxifene versus placebo over three years of follow-up

<table>
<thead>
<tr>
<th></th>
<th>Placebo Absolute Risk (N*=2,576)</th>
<th>Raloxifene (pooled) Absolute Risk (N=5,129)</th>
<th>ARR/ARI</th>
<th>NNT/NNH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deaths: total*</td>
<td>26 (1.0%)</td>
<td>41 (0.8%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Deaths: cardiovascular**</td>
<td>15* (0.6%)</td>
<td>31 (0.6%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Serious adverse events</td>
<td>25%</td>
<td>24%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(FDA medical review)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Venous thromboembolic events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of women</td>
<td>8 (0.31%)</td>
<td>49 (0.96%)</td>
<td>+.65%</td>
<td>154 (NNH)</td>
</tr>
<tr>
<td>Early breast cancer diagnoses ***</td>
<td>27 (1.05%)</td>
<td>15 (0.29%)</td>
<td>-.76%</td>
<td>132 (NNT)</td>
</tr>
<tr>
<td>Cardiovascular events: nonfatal</td>
<td>82 (3.2%)</td>
<td>145 (2.8%)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*13 types of non-vertebral fractures were examined individually and none differed significantly (Bonferroni adjustment for 13 comparisons).
**Cardiovascular deaths included fatal myocardial infarction (MI), sudden death, unwitnessed death in absence of other non-coronary causes, death related to coronary artery procedure, fatal stroke.
Non-fatal cardiovascular outcomes included MI, unstable angina, coronary ischemia and cerebrovascular events (stroke or transient ischemic events).
***All but two cases were classified as invasive; invasiveness of remaining two was unknown.
ARR=absolute risk reduction; ARI=absolute risk increase; NNT=numbers needed to treat; NNH=numbers needed to harm

Cardiovascular event rates have been reported. The rate of early breast cancer incidence is lower in women on raloxifene than on placebo, 27 (1.05%) versus 15 (0.29%) respectively, an absolute risk decrease of -0.76% (NNT=132).
4 DISCUSSION

In this systematic review, a meta-analysis was planned to pool evidence from RCTs regarding raloxifene’s effect on vertebral fracture (primary outcome) and BMD (secondary outcome) in postmenopausal women. We identified 17 RCTs that included vertebral fractures and BMD data. All trials assessing non-vertebral fractures and morphometric and clinical vertebral fractures had placebo as the control intervention. Adverse event data involved placebo, estrogen and combined hormone therapy as comparators.

The planned meta-analysis of vertebral fractures could not be justified for the two trials reporting vertebral fracture data. The central problems were the observed clinical heterogeneity (i.e., definition of vertebral fracture and population characteristics17,20,111-114) and statistical heterogeneity. As a result, there was no justification for using a random effects model to combine MORE and Lufkin data. In addition, vertebral fracture was of clinical importance when it was associated with back pain and disability, but the Lufkin trial did not include any separate reporting of women with symptomatic vertebral fracture.

Given the size, quality and duration of the MORE trial, its results best reflected raloxifene’s effects on vertebral fracture, as compared to placebo, in women with very low BMD or vertebral deformity at baseline. The MORE trial tested two doses of raloxifene: 60 mg/day and 120 mg/day. The 120 mg/day dose provided a greater benefit, but the difference between 120 mg/day and 60 mg/day was not statistically significant. In addition, 120 mg/day exceeds the recommended daily dose in Canada (60 mg).

In the MORE trial, the incidence of vertebral fracture differed in the two substudies and was related to the presence of vertebral fracture at baseline (secondary prevention). Vertebral fracture at baseline was a stronger determinant of subsequent vertebral fracture during the study period, than treatment allocation to raloxifene or placebo.

Pooled analysis was used for BMD findings. Raloxifene had a significant positive impact on BMD compared with placebo at various sites. The most pronounced impacts were observed for the lumbar spine in studies where fewer than all participants had very low BMD at baseline (heterogeneous low BMD) or fewer than all patients had higher BMD at baseline (i.e., heterogeneous high BMD).

In the MORE study, the positive BMD impact did not correlate with clinical benefit as measured by non-vertebral fractures or vertebral fracture. Women with low BMD at baseline but no vertebral fracture had lower rates of vertebral fracture during the trial than women with vertebral fracture. The latter group included women enrolled in this group regardless of initial BMD. In addition, data from the MORE trial (reported by the FDA) indicated that women on placebo with higher (i.e., better) lumbar spine BMD also had a higher incidence of vertebral fracture.

Investigators have tried to resolve the “treatment paradox” seen in the MORE study and elsewhere,117,118 whereby decreases in vertebral fracture rates occur without commensurate increases in bone density.119-121 Other variables may be responsible for the risk reduction in vertebral fracture that is not reflected in BMD. For example, slowing bone resorption may
increase the biomechanical strength of bone and reduce bone turnover, which in turn alter vertebral fracture risk. One explanation for this finding is that vertebral fractures can artifactually increase BMD, which leads to an increase in lumbar spine BMD measures. It is unclear, however, why this would occur in the placebo group but not in women treated with raloxifene.

The MORE trial was of sufficient size and duration to assess the clinical outcomes of raloxifene therapy (Table 11).

Table 11: Non-vertebral fracture incidence: raloxifene versus placebo over three years of follow-up

<table>
<thead>
<tr>
<th>Fracture Type</th>
<th>Placebo Absolute Risk (N*=2,576)</th>
<th>Raloxifene (pooled) Absolute Risk (N=5,129)</th>
<th>ARR/ARI</th>
<th>NNT/NNH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hip fractures*</td>
<td>18 (0.7%)</td>
<td>40 (0.8%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>All non-vertebral fractures*</td>
<td>240 (9.3%)</td>
<td>437 (8.5%)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*13 types of non-vertebral fractures were also examined individually and none differed significantly (Bonferroni adjustment for 13 comparisons).

The Lufkin trial also reported on clinical fractures. Total fracture rate was not reported. Three women on placebo (6.6%) and three on raloxifene (3.4%) experienced non-vertebral fractures. No placebo users and one raloxifene user (1.1%) experienced a hip fracture. The differences were not statistically significant. At a rate of 9% for total non-vertebral fractures in the placebo group, the trial had 80% power to detect an absolute difference of 1.9% in fracture rate with a 95% degree of confidence (post hoc power analysis, Epi Info).

In the pooled analysis, raloxifene decreased the risk of several mild to moderate adverse events compared with estrogen or combined hormone therapy (i.e., breast pain and soreness, endometrial proliferation, vaginal bleeding). Two of the active comparators are unapproved for osteoporosis prevention: tamoxifen and estrogen alone in women without hysterectomy (the latter is contraindicated because of increased endometrial cancer risks).122,123

As compared with placebo, raloxifene increases the rate of hot flashes significantly at a higher than approved dose level (120 mg/day). The comparison of raloxifene to hormone therapy is questionable, however, given that raloxifene has no claimed effects in the symptomatic relief of hot flashes, while this is an approved indication for combined estrogen-progestin therapies and for estrogen alone in women with a previous hysterectomy. The comparison to placebo, on the other hand, indicates whether raloxifene would increase the probability of hot flashes as compared to no treatment.

Discontinuations due to an adverse event were more likely for raloxifene users in placebo-controlled, trials of women with very low BMD (generally older women). Breast pain and soreness were less likely to be a problem for raloxifene users than for recipients of a comparator in trials involving women with higher BMD levels.
The impact of raloxifene on lipid metabolism was beyond the focus of this review, but the MORE trial demonstrated that raloxifene significantly decreased levels of LDL cholesterol and total cholesterol relative to placebo (Table 9).

A post hoc assessment of cardiovascular outcomes in the MORE trial found no evidence of an increased rate of cardiovascular events with raloxifene versus placebo. Raloxifene’s estrogen antagonist effect on breast tissue could not pooled because of inadequate breast cancer data from most trials. The MORE and Johnston-Delmas trials included data, but the data could not be combined. In the MORE trial, there was a reduction in early invasive breast cancer, although it was impossible to establish whether raloxifene had a sustained effect on breast cancer progression or mortality.

Raloxifene’s effect on uterine tissue cannot be illuminated by the results of the pooled analyses. In placebo-controlled RCTs, raloxifene use is associated with increased endometrial thickness. The observations concerning endometrial proliferation and vaginal bleeding, however, are associated with studies of younger women. Unopposed estrogen therapy is a known cause of endometrial cancer and thus, it is an inappropriate comparator to assess the potential risks associated with cancer from another drug. None of the morphological or anatomical changes in the uterus associated with raloxifene are linked to cancer incidence, morbidity or mortality.

Serious adverse event rates could only be determined from the large MORE trial. Raloxifene significantly increased and decreased the risk of venous thromboembolic events and breast cancer respectively.

Overall, 24.2% of participants in the MORE trial experienced a serious adverse event. The incidence did not differ between raloxifene and placebo. Limited information was available on the type of events [cardiovascular, venous thromboembolic events (VTE), breast cancer incidence, mortality]. The reported serious adverse events accounted for approximately 5% of the total serious adverse events in each category. Underreporting of serious adverse events was seen as typical of the RCT literature.124,125

The MORE trial reports limited data regarding VTEs beyond the total events for each therapeutic group. Details are crucial, because these events typically range from less severe venous obstruction to fatal pulmonary embolism.

The manufacturer provided additional details regarding VTEs to the US FDA. They reported this as the main serious adverse event observed in the approximately 10,000 women participating in all clinical trials sponsored by Lilly (Table 12). No dose-specific reports were available, but the manufacturer stated that the frequency of VTE did not increase with (increasing) doses above 60 mg/day.115
**Table 12:** Venous thromboembolic events

| Adverse Events: Estimated annual incidence rate (per 1,000 patients) | Treatment Group |  |
|---|---|---|---|
| Placebo N=3,195 | Raloxifene N=6,681 | RR (95% Confidence Interval) |
| All venous thromboembolic events (VTE) | 1.12 | 3.82 | 3.4 (1.5 to 8.0) |
| All VTE except retinal vein thrombosis | 0.75 | 3.64 | 4.9 (1.8 to 14) |
| Pulmonary emboli | 0.56 | 1.24 | 2.2 (0.6 to 7.8) |

The estimated annual attributable risk of a venous thromboembolic VTE with raloxifene use is 2.7/1,000 women. This is a measure of the incidence in raloxifene users minus the incidence in placebo controls. VTE increases in frequency with age and most women participating in clinical trials are <65. The largest increases in VTE are observed in trials involving older women.\(^{115}\)

The most common complication of VTEs is post-thrombotic syndrome (persistent leg pain and swelling with or without ulceration) after the occurrence of a deep vein thrombosis (DVT). Post-thrombotic syndrome is estimated to occur within five years in 60% to 70% of patients who develop proximal deep vein thrombosis and within two years in 16% of the patients who develop a distal deep vein thrombosis.\(^{126}\) Siragusa *et al.* found that the incidence of this disorder was 24% compared with 4% in controls, after two to four years, in patients who had subclinical DVT demonstrated by contrast venography after hip or knee surgery, despite at least three months of anticoagulation.\(^{127}\)

In the MORE trial, raloxifene decreased the incidence of early invasive breast cancer diagnoses. Whether these represented breast cancer prevention or a delay in diagnosis is unknown, as the number of diagnoses was small and follow-up insufficient. The effect on breast cancer mortality was unknown. There were also more cases of treatment-emergent diabetes in raloxifene users versus placebo in the MORE trial. The longer-term clinical implications of these differences remained unclear.

To summarize observed clinical benefit and harm (beyond tolerability), raloxifene was not found to have a statistically significant effect on mortality or overall serious morbidity, nor was any effect observed on non-vertebral fracture rates, notably hip fractures. The MORE trial was adequately powered to determine the reduction of total non-vertebral fractures and no reduction was found except in a post hoc analysis of individuals with severe vertebral fractures.\(^{128}\) This raised questions about the product’s efficacy in terms of the outcomes that were likely to be most important to patients and clinicians. The MORE trial was inadequately powered to examine hip fractures alone. Larger, long-term studies are needed to determine raloxifene’s use for the prevention of hip fractures. The most important observed clinical outcomes were a reduction in symptomatic vertebral fractures, a reduction in early invasive breast cancer in postmenopausal women and an increase in thrombo-embolic events.
There are several research implications. The published results of the Women’s Health Initiative trial of estrogen-progestin treatment for disease prevention reinforce the need for evidence that benefits outweigh harms before a strategy for osteoporosis prevention can be recommended to healthy populations.129

Other studies are needed to explain why the MORE trial found that raloxifene has a beneficial effect on vertebral fractures, but no effect on non-vertebral fractures. This will help explain the interrelationships among biochemical markers of bone turnover, biomechanical bone strength, BMD, vertebral fractures and clinical fractures.15-17

The STAR trial, now underway, compares raloxifene and tamoxifen for the prevention of breast cancer. Unfortunately, the study lacks a placebo arm, making it impossible to determine the effect of these drugs on serious adverse events and mortality in an initially healthy population.

5 CONCLUSIONS

The best estimate of overall benefit and harm from raloxifene is derived from the MORE trial’s three-year, placebo-controlled results. Raloxifene has no effect on the incidence of non-vertebral fractures or on hip fracture. The absolute risk reduction for symptomatic vertebral fracture is 0.8% (1.4% in placebo and 0.6% in raloxifene groups, number needed to treat=125 over three years). This benefit needs to be balanced by the increase in the number of serious adverse events due to venous thrombo-embolic disease of 0.65% (0.31% in placebo and 0.96% in raloxifene, number needed to harm=154 over three years).

The largest observed incidence of vertebral fracture occurs among older women with vertebral deformity at baseline. This is also the group in which raloxifene has the greatest impact. The group includes women with a range of BMD levels. The impact of raloxifene on vertebral fracture in younger postmenopausal women or women with higher BMD levels has not been established.

The impact of raloxifene on morbidity and mortality due to cardiovascular disease or cancer (breast or uterine) has not been established, although the evidence suggests a small but significant decrease in the incidence of early breast cancer. A subgroup analysis of cardiovascular adverse events reported in the MORE trial indicates no significant increase in morbidity and mortality in raloxifene users as compared to placebo, after three years.
6 REFERENCES


78. Delmas PD. Raloxifene reduces the risk of vertebral fracture and improves lipid profiles of postmenopausal women after 3 years [abstract]. *Calcif Tissue Int* 1999;64(Suppl 1):S120.


86. Shah AS, Scheele WH, Glant MD, Fugère P. Raloxifene HCl is not stimulatory in the endometrium as assessed by the blustein criteria and an estrogenicity scoring system [abstract]. *Prim Care Update Ob/Gyns* 1998;5(4):167.


Appendix 1: Literature Search Strategy

**Dialog® OneSearch® and Alerts®**

<table>
<thead>
<tr>
<th>DATABASES</th>
<th>SEARCH TERMS</th>
</tr>
</thead>
</table>
| Dialog® OneSearch® | 1. raloxifene/ti,ab  
| | 3. keoxifene/ti,ab  
| | 4. tn=keoxifene  
| | 5. Set 1: Set 4  
| | 6. osteoporosis/de  
| | 7. osteoporosis!  
| | 8. fractures/de  
| | 9. fracture/de  
| | 10 fracture reduction/de  
| | 11. bone demineralization, pathologic/de  
| | 12. bone demineralization, pathologic!/de  
| | 13. bone atrophy/de  
| | 14. bone density/de  
| | 15. bone regeneration/de  
| | 16. bone regeneration!/de  
| | 17. bone demineralization!/de  
| | 18. calcification, physiologic/de  
| | 19. bone mineralization/de  
| | 20. osteoporosis/ti, ab  
| | 21. (facture OR fractures)/ti,ab  
| | 22. fracture(reduction?)/ti,ab  
| | 23. bone()demineralization/ti,ab  
| | 24. bone()atrophy/ti,ab  
| | 25. bone()density/ti,ab  
| | 26. bone()regeneration/ti,ab  
| | 27. bone()demineralization/ti,ab  
| | 28. physiologic()calcification/ti,ab  
| | 29. bone()mineralization/ti,ab  
| | 30. Set 7 : Set 29  
| | 31. dt=clinical trial  
| | 32. clinical trial!/de  
| | 33. clinical trials/de  
| | 34. clinical trials!/de  

**Notes:**
de = descriptor, e.g. in MEDLINE®, HealthSTAR a Medical Subject Heading or MeSH, a controlled, thesaurus term; in EMBASE®, etc., a thesaurus term as well.  
ti = title (i.e. word has to occur in title field of the bibliographic record)  
ab = abstract (i.e. word has to occur in abstract field of bibliographic record)  
! = explode; picks up narrower terms as well, i.e. terms which are conceptually subsets of a broader term  
() = words must be adjacent  
? = truncation symbol  
rd = reduce duplicates, i.e. duplicate references are removed
35. dt=randomized controlled trial
36. dt=controlled clinical trial
37. random allocation/de
38. double-blind method/de
39. random?/ti,ab
40. placebo?/ti,ab
41. controlled ()trial?/ti,ab
42. double()blind?/ti,ab
43. dt=meta-analysis
44. meta-analysis
45. meta()analy?/ti,ab
46. metaanaly?/ti,ab
47. quantitative()?review?/ti,ab
48. quantitative()?overview?/ti,ab
49. quantitative()?review?/ti,ab
50. evidence based medicine/de
51. multicenter study/de
52. randomized controlled trial/de
53. drug comparison/de
54. comparative study/de
55. toxicity/de
56. adverse effect/de
57. contraindications/de
58. side effects/de
59. adverse reactions/de
60. adverse reaction/de
61. drug toxicity/de
62. controlled()clinical()trial?/ti,ab
63. multicenter()stud?/ti,ab
64. multicentre()stud?/ti,ab
65. toxic?/ti,ab
66. adverse/ti,ab
67. effect?/ti,ab
68. adverse()effect?/ti,ab
69. adverse()effect?/ti,ab
70. contraindicat?/ti,ab
71. side()effect?/ti,ab
72. adverse()reaction?/ti,ab
73. Set 31: Set 72
74. Set 5 AND Set 30 AND Set 73 = References
75. estrogen replacement therapy/de
76. hormone replacement therapy/de
77. hormone substitution/de
78. hormonal therapy/de
79. estrogen therapy/de
80. Set 75 : Set 79
81. Set 74 AND Set 80 = References (subset of Set 74)
Appendix 2: Additional Details Concerning Methods

Selection Process: Citations were entered into a Reference Manager database, with duplicates removed manually. Two individuals (HMS, AC) independently “broad screened” the title, abstract and key words for each citation by liberally applying the eligibility criteria to determine whether to retain it. It was retained if it contained pertinent information. If one broad screener was at best uncertain as to its potential relevance, it was entered into the next phase of the review and deemed to be “potentially relevant.”

Hard copies of the full reports were retrieved and assessed for relevance independently by two reviewers (HMS, TC) who applied the eligibility criteria strictly. All disagreements were resolved by forced consensus. The reasons for excluding studies were noted. Before completing the relevance assessment, the reviewers undertook a reliability study, which yielded a 95% agreement with 20 randomly selected, full reports. The reason for the one disagreement was discussed to prevent its recurrence.

Assessment of Methodological Quality of Trials from Reports: The quality of each included trial was assessed independently by two reviewers (HMS, TC) who were familiar with the validated, three-item Jadad et al. scale, which assesses randomization, double-blinding and the inclusion of data for dropouts and withdrawals. Total scores range from zero to five, with scores <3 indicating poor quality. The concealment of each trial’s allocation to treatment was also evaluated by the reviewers (grade A=adequate; B=unclear; C= inadequate) (Appendix 3). Scoring differences were settled by forced consensus.

Data Synthesis and Analysis of Efficacy and Safety Data: If appropriate, binary data from trials were pooled using the fixed effects Mantel-Haenszel estimate of the relative risk. Statistical heterogeneity of the relative risks across studies was assessed using the chi-square Q statistic, but the power of this test may be low and clinical insight may be more relevant in understanding heterogeneity. The DerSimonian-Laird random effects estimate of the relative risk was used to pool trials if there was between-trial statistical heterogeneity. When there was little statistical heterogeneity, the DerSimonian-Laird estimate was identical to the fixed effects estimate. As an index of precision, a 95% confidence interval was calculated for each estimate.

When percentages of patients responding to treatment were reported, the numbers of patients were computed by multiplication with rounding. When the goal of a MA was to compare all doses of raloxifene to a comparator, different doses were pooled in trials by summing the number of patients. Where appropriate, it was planned to derive the numbers needed to treat (NNT) and harm (NNH) for primary outcome and adverse event outcome data respectively.

Continuous (i.e., BMD, one adverse event) data were used to derive the weighted mean difference. While “percent change from baseline” (i.e., BMD) was a poor outcome measure, it could be synthesized in a fixed effects MA by computing the difference in mean percent change from baseline between treatment and control; and weighting by the reciprocal of the squared standard error. The chi-square Q statistic and the DerSimonian-Laird approach were used as required.
In any trial where multiple doses were compared to a comparator, the mean percent change from baseline for treatment was computed by taking a weighted average of the mean percent change from baseline for the different doses, with weights proportional to the inverse of the squared standard error of each mean percent change from baseline. Not all variances were available, so variance imputation was required. For studies in which the standard error (SE) was available, the standard deviation (SD) was computed as:

$$SD = \sqrt{n} \times SE,$$
where $n$=number of participants.

The median standard deviance was computed for outcomes (Tables 6 and 8). SEs were computed for the studies where they were not reported as:

$$SE = \text{median } SD / \sqrt{n}.$$

The difference in mean percent difference from baseline, between treatment and control, was computed. The SE of the difference was computed as:

$$SE = \sqrt{SE^2(\text{treatment}) + SE^2(\text{control})}.$$

**Sensitivity and Subgroup Analyses:** Where baseline population data were unavailable for a given study, the inclusion criteria were used in Appendix 4 and to derive summary statistics (Tables 1 to 3). If the inclusion criterion information was expressed as a range (e.g., two to eight years post-menopause) or as a minimum requirement (e.g., $\geq 2$ years post-menopause), the midpoint and the lower bounds respectively were entered into calculations.

**Publication Bias:** Funnel plots (e.g., effect size versus precision) involving primary outcomes were to be inspected, while tests of publication bias were to be used where appropriate. The latter yielded quantitative views of a funnel plot’s degree of asymmetry (rank correlation test, graphical test), the number of “missing” or unobserved trials given the dispersion of trial estimates (trim and fill method) and the pooled estimate of efficacy or safety when adjusted for the impact of publication bias (trim and fill method).
Appendix 3: Tools to Assess Quality of Randomized Controlled Trials

1. Randomization:
Is the study described as randomized (i.e., including words such as randomly, random, randomization)?

Yes = 1
No = 0

A trial reporting that it is “randomized” receives one point. Trials describing an appropriate method of randomization (table of random numbers, computer generated) receive an additional point.

Appropriate = 1
Not appropriate = 0

If the report describes the trial as randomized and uses an inappropriate method of randomization (e.g., date of birth, hospital numbers), a point is deducted.

Total points: 0 1 2

Score = ___

2. Double-blinding:
Is the study described as double-blind?

Yes = 1
No = 0

A trial reporting that it is double-blind receives one point. Trials that describe an appropriate method of double-blinding (identical placebo: colour, shape, taste) receive an additional point.

Yes = 1
No = 0

If the report describes the trial as double-blind and uses an inappropriate method (e.g., comparison of tablets versus injection with no dummy), a point is deducted.

Total points: 0 1 2

Score = ___

3. Withdrawals and dropouts:
Is there a description of withdrawals and dropouts?

Yes = 1
No = 0

A trial reporting the number of and reasons for withdrawals or dropouts receives one point. If there is no description, no point is given.

Score = ___

Overall score: ___
Low = 0 to 2 points
Moderate = 3 to 4 points
High = 5 points (maximum)

4. Adequacy of Allocation Concealment: (circle one)

- Central randomization; numbered or coded bottles or containers; drugs prepared by a pharmacy; serially numbered, opaque, sealed envelopes = Adequate

- Alternation; reference to case record number or date of birth = Inadequate

- Allocation concealment is not reported or fits neither category = Unclear
Appendix 4: Raloxifene for Primary and Secondary Prevention of Osteoporosis in Postmenopausal Women

<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD; Range)</th>
<th>Years Postmenopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delmas Eastell</td>
<td>2000</td>
<td>7,705</td>
<td>66.5 (6.8) yrs; 31 to 80 yrs</td>
<td>19.3 (8.0) yrs (&gt;2 yrs*)</td>
<td>Tx: substudy 1: &lt;2.5 SD below peak femoral neck or lumbar BMD;* substudy 2: low BMD (undefined) and ≥1 moderate or severe VFx or ≥2 mild VFxs or ≥2 moderate VFxs*</td>
<td>4 yrs (mean=39.4 mos; median=47.4 mos)</td>
<td>RLX 60, RLX 120</td>
<td>Placebo</td>
<td>Abstract</td>
<td>180 (25 countries)</td>
<td>4/adequate</td>
</tr>
<tr>
<td>Ettinger</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensrud</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ettinger</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cauley</td>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cummings</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cummings</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maricic</td>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MORE trial</strong></td>
<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>(GGGGG)</td>
<td></td>
<td>143</td>
<td>68.4 (5.0) yrs; 51 to 76 yrs; 45 to 75 yrs*</td>
<td>22.6 (1.1) yrs (NR)</td>
<td>Tx: BMD &lt;10th percentile and ≥1 non-traumatic vertebral Fx*</td>
<td>52 wks</td>
<td>RLX 60, RLX 120</td>
<td>(Ca 500 mg/day; vitamin D 400 to 600 IU/day)</td>
<td>Placebo</td>
<td>Journal</td>
<td>2 (US)</td>
</tr>
<tr>
<td>Lufkin</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lufkin</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lufkin</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lufkin</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lufkin</td>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nickelsen</td>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abstract</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Id=identification; yr=year; N or n=sample size; RLX=raloxifene; SD=standard deviation; SE=standard error; wks=weeks; mos=months; BMD=bone mineral density; Fx=fracture; VFx=vertebral Fx; NR=not reported; Lilly=Eli Lilly Inc.; Px=prevention; Tx=treatment; CEE=conjugated equine estrogen; *inclusion criteria; Ca=calcium; FSH=follicle stimulating hormone; *according to World Health Organization’s (WHO) BMD definition of osteoporosis, osteopenia and normal; two identical trials [Delmas/GGGF (n=601) and GGGG (n=544)]; a crossover at 3 mos for topical vaginal interventions; likely a Lilly trial=authors’ associations suggest Lilly sponsorship yet unconfirmed via trial code information provided by Canadian representative of Eli Lilly.
<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD); Range</th>
<th>Years Post-menopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meunier Meunier Meunier Sarkar (GGGP)</td>
<td>1999</td>
<td>129</td>
<td>60.2 (6.7) yrs; 50 to 75 yrs*</td>
<td>12.5 (9.0) yrs (NR)</td>
<td>≥1 yr* or estradiol and FSH levels</td>
<td>Tx &lt;2.8 below BMD peak with ≤3 previous Fx; 60% were &lt;2.5 SD below</td>
<td>104 wks 52 wks 52 wks</td>
<td>RLX 60, RLX 150 (Ca 1,000 mg/day; vitamin D 300 IU/day)</td>
<td>Placebo</td>
<td>Journal Abstract Abstract Abstract</td>
<td>8 (France)</td>
</tr>
<tr>
<td>Johnell Johnell Stock (GGHV)</td>
<td>2000 1999 1999</td>
<td>330</td>
<td>NR (NR); NR ≤75 yrs*</td>
<td>NR (NR) ≥2 yrs*</td>
<td>Tx and Px: &lt;2 SD below peak femoral neck BMD*</td>
<td>52 wks</td>
<td>RLX 60; RLX 60 + ALN 10 mg/day (Ca 0.5 g/day; vitamin D 400 IU/day to 600 IU/day)</td>
<td>Placebo; ALN 10 mg/day</td>
<td>Abstract Abstract Abstract</td>
<td>NR</td>
<td>2/unclear</td>
</tr>
<tr>
<td>Morii Morii (likely a Lilly trial)</td>
<td>1997 1996</td>
<td>93</td>
<td>56 (NR) yrs; 38 to 65 yrs</td>
<td>Median= 7 (NR) yrs (NR) (serum estradiol and FSH levels confirmed)</td>
<td>Tx and Px: &lt;2 SD below peak BMD*</td>
<td>12 wks of 24 wk trial</td>
<td>RLX 30, RLX 90 (Ca 800 mg/day)</td>
<td>Placebo</td>
<td>Abstract Abstract Abstract</td>
<td>NR</td>
<td>1/unclear</td>
</tr>
</tbody>
</table>

Primary and secondary prevention trials (i.e., osteoporotic and osteopenic women): *efficacy and safety data

<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD); Range</th>
<th>Years Post-menopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestwood Gunness Prestwood (GGGM)</td>
<td>2000 1997 1997</td>
<td>51</td>
<td>64.4 (5.8) yrs; 55 to 85 yrs*</td>
<td>18.2 (8.5) yrs ≥5 yrs*</td>
<td>Tx and Px: 3.0 SD below peak BMD to 1 SD above*</td>
<td>26 wks</td>
<td>RLX 60 (NR)</td>
<td>CEE 0.625 mg/day (women with uterus given Provera in last 2 wks)</td>
<td>Journal Abstract Abstract</td>
<td>NR?</td>
<td>3/unclear</td>
</tr>
</tbody>
</table>

Id=identification; yr=year; N or n=sample size; RLX=raloxifene; SD=standard deviation; SE=standard error; wks=weeks; mos=months; BMD=bone mineral density; Fx=fracture; VFx=vertebral Fx; NR=not reported; Lilly=Eli Lilly Inc.; Px=prevention; Tx=treatment; CEE=conjugated equine estrogen; *inclusion criteria; Ca=calcium; FSH=follicle stimulating hormone; ALN=alendronate sodium; aaccording to World Health Organization’s (WHO) BMD definition of osteoporosis, osteopenia and normal; btwo identical trials [Delmas/GGGF (n= 601) and GGGG (n=544)]; acrossover at 3 mos for topical vaginal interventions; likely a Lilly trial=authors’ associations suggest Lilly sponsorship yet unconfirmed via trial code information provided by Canadian representative of Eli Lilly.
<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD; Range)</th>
<th>Years Post-menopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnston (GGGF)</td>
<td>2000</td>
<td>1,145</td>
<td>54.6 (SE=0.01) yrs; 45 to 60 yrs*</td>
<td>NR (NR)b</td>
<td>Px: 2.5 SD below peak BMD to 2.0 SD above peak* =55% osteopenic and 45% “normal” participants (for Delmas data only)</td>
<td>36 mos</td>
<td>RLX 30, RLX 60, RLX 150 (Ca 400 to 600 mg/day)</td>
<td>Placebo</td>
<td>Journal Abstract</td>
<td>11</td>
<td>4/unclear</td>
</tr>
<tr>
<td>Shahb</td>
<td>1998</td>
<td></td>
<td></td>
<td>4.8 (2.0) yrs= Delmas trial only (2 to 8 yrs*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delmas</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bjarnason</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delmas</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GGGF)</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsai</td>
<td>2001</td>
<td>116</td>
<td>56.9 (4.2) yrs; 47 to 66 yrs</td>
<td>NR (NR)</td>
<td>Px: 2.5 SD below, to 1 SD above, peak bone mass*</td>
<td>12 mos</td>
<td>RLX 60 (Ca 500 mg/day)</td>
<td>CEE 0.625 mg/day + 5 mg/day medroxy-progesterone acetate</td>
<td>Journal</td>
<td>2</td>
<td>3/unclear</td>
</tr>
<tr>
<td>(a Lilly trial)</td>
<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>Pavo</td>
<td>1999</td>
<td>128</td>
<td>59 (NR) yrs; NR</td>
<td>NR (NR) (&gt;2 yrs*)</td>
<td>Px: no selection criteria or baseline data reported</td>
<td>12 mos</td>
<td>RLX 60 (no supplements or vitamins allowed)</td>
<td>Placebo</td>
<td>Abstract</td>
<td>≥2</td>
<td>2/unclear</td>
</tr>
<tr>
<td>Stock</td>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Lilly trial but no ID provided)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Id=identification; yr=year; N or n=sample size; SD=standard deviation; RLX=raloxifene; SE=standard error; wks=weeks; mos=months; BMD=bone mineral density; Fx=fracture; VFx=vertebral Fx; NR=not reported; Lilly=Eli Lilly Inc.; Px=prevention; Tx=treatment; CEE=conjugated equine estrogen; *inclusion criteria; Ca=calcium; FSH=follicle stimulating hormone; ‘according to World Health Organization’s (WHO) BMD definition of osteoporosis, osteopenia and normal; ’two identical trials [Delmas/GGGF (n= 601) and GGG (n=544)]; a crossover at 3 mos for topical vaginal interventions; likely a Lilly trial=authors’ associations suggest Lilly sponsorship yet unconfirmed via trial code information provided by Canadian representative of Eli Lilly.
<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD); Range</th>
<th>Years Post-menopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fugere</td>
<td>2000</td>
<td>136</td>
<td>56 (5.1) yrs; 45 to 66 yrs</td>
<td>5.9 (4.5) yrs (1 to 17 yrs)</td>
<td>Px Healthy*</td>
<td>24 mos</td>
<td>RALX 150 (NR)</td>
<td>CEE 0.625 mg/day + medroxyprogesterone acetate 2.5 mg/day</td>
<td>Journal Abstract</td>
<td>11</td>
<td>3/unclear</td>
</tr>
<tr>
<td>Shah</td>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jolly</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glant</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheele</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheele Prestwood</td>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GGGZ)</td>
<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>Goldstein</td>
<td>2000</td>
<td>415</td>
<td>54.7 (3.5) yrs; NR 47 to 60 yrs*</td>
<td>4.6 (1.9) yrs (2 to 8 yrs*)</td>
<td>Px Healthy* 2.5 SD below to 2 SD above peak BMD* (i.e. osteopenic and &quot;normal&quot;)</td>
<td>12 mos</td>
<td>RALX 60, RALX 150 (Ca 520 mg/day)</td>
<td>Placebo: CEE 0.625 mg/day</td>
<td>Journal Abstract</td>
<td>33</td>
<td>5/adequate</td>
</tr>
<tr>
<td>Thiebaud</td>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldstein</td>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GGHG)</td>
<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>Boss</td>
<td>1997</td>
<td>251</td>
<td>53.1 (3.4) yrs; 45.2 to 60.9 yrs</td>
<td>Median=35.6 (NR) mos; (6 to 87 mos)</td>
<td>Px Healthy*</td>
<td>8 weeks</td>
<td>RALX 200, RALX 600 (Ca 520 mg/day)</td>
<td>Placebo: CEE 0.625 mg/day</td>
<td>Journal Journal Abstract</td>
<td>11</td>
<td>4/unclear</td>
</tr>
<tr>
<td>Draper</td>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huster</td>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(GGGB/US)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Id=identification; yr=year; N or n=sample size; SD=standard deviation; RLX=raloxifene; SE=standard error; wks=weeks; mos=months; BMD=bone mineral density; Fx=fracture; VFx=vertebral Fx; NR=not reported; Lilly=Eli Lilly Inc.; Px=prevention; Tx=treatment; CEE=conjugated equine estrogen; *inclusion criteria; Ca=calcium; FSH=follicle stimulating hormone; *according to World Health Organization’s (WHO) BMD definition of osteoporosis, osteopenia and normal; †two identical trials [Delmas/GGGF (n= 601) and GGGG (n=544)]; ‡a crossover at 3 mos for topical vaginal interventions; likely a Lilly trial=authors’ associations suggest Lilly sponsorship yet unconfirmed via trial code information provided by Canadian representative of Eli Lilly.
<table>
<thead>
<tr>
<th>Author (Lilly Trial Id)</th>
<th>Yr</th>
<th>N</th>
<th>Mean Age (SD); Range</th>
<th>Years Post-menopause (Mean; SD) (Range)</th>
<th>Trialist’s Stated Goal of Osteoporosis Intervention</th>
<th>Intervention Length</th>
<th>Raloxifene Intervention (mg/day) (Supplements)</th>
<th>Comparator Intervention Dose</th>
<th>Publication Status</th>
<th>Number of Centres</th>
<th>Trial Quality: Total Jadad Score/Allocation Concealment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freedman San Martin Cohen Laksh-manan (GGGH)</td>
<td>2001 2000 1998 1999</td>
<td>619</td>
<td>52.9 (4.7) yrs (based on n=168); 45 to 60 yrs*</td>
<td>6.1 (4.2) yrs based on only n=168; (NR)</td>
<td>Px healthy* no breast cancer</td>
<td>24 months (breast data; n= 168) 36 mos (vaginal data; n= 619) (5 yr trial) (hysterectomy &lt;15 yrs before study*)</td>
<td>RLX 60, RLX 150 (NR) low dose vaginal estrogens (except estradiol) up to 3 times per wk during trial</td>
<td>Placebo; CEE 0.625 mg/day</td>
<td>Journal Abstract Abstract Abstract</td>
<td>38 (10 countries in North America, Europe, South Africa, New Zealand)</td>
<td>3/unclear</td>
</tr>
<tr>
<td>Vardy (not a Lilly trial?)</td>
<td>1998</td>
<td>33</td>
<td>NR (NR)</td>
<td>NR (NR)</td>
<td>Px: healthy*</td>
<td>18 wks</td>
<td>RLX 60 (NR)</td>
<td>Placebo; tamoxifen 20 mg/day; CEE 0.625 mg/day</td>
<td>Abstract</td>
<td>NR</td>
<td>2/unclear</td>
</tr>
<tr>
<td>Parsons Parsons Parsons (GGIK)</td>
<td>1999 1999 1999</td>
<td>187</td>
<td>59 (NR) yrs; NR</td>
<td>9.5 (NR) yrs (NR)</td>
<td>Px: naturally post-menopausal;* genito-urinary atrophy*</td>
<td>3 mos*</td>
<td>RLX 60 (NR)</td>
<td>Placebo; open label vaginal Premarin 0.5 g/day or Replens*</td>
<td>Abstract Abstract Abstract</td>
<td>NR</td>
<td>2/unclear</td>
</tr>
<tr>
<td>Walsh Walsh Walsh (GGGY)</td>
<td>1998 1997 2000</td>
<td>390</td>
<td>59.3 (6.3) yrs; 45 to 72 yrs*</td>
<td>11.0 (7.5) yrs; (≥1 yr*)</td>
<td>Px: Healthy* (if hysterectomy, then 50 to 72 yrs*)</td>
<td>6 mos</td>
<td>RLX 60, RLX 120 (NR)</td>
<td>Placebo: CEE 0.625 mg/day + medroxyprogesterone acetate 2.5 mg/day</td>
<td>Journal Abstract Journal Journal</td>
<td>8 (US)</td>
<td>5/unclear</td>
</tr>
<tr>
<td>Mijatovic Mijatovic Valk deRoo (likely a Lilly trial)</td>
<td>1999 1998 1999</td>
<td>60</td>
<td>54.8 (3.5) yrs; NR</td>
<td>NR (NR) (NR)</td>
<td>Px: healthy*</td>
<td>24 months (5 yr trial) (all hysterectomy*)</td>
<td>RLX 60, RLX 150 (Ca 500 mg/day)</td>
<td>Placebo: CEE 0.625 mg/day</td>
<td>Journal Journal Journal</td>
<td>1 (Netherlands)</td>
<td>3/unclear</td>
</tr>
</tbody>
</table>

*Ident=identification; yr=year; N or n=sample size; SD=standard deviation; RLX=raloxifene; SE=standard error; wks=weeks; mos=months; BMD=bone mineral density; Fx=fracture; VFx=vertebral Fx; NR=not reported; Lilly=Eli Lilly Inc.; Px=prevention; Tx=treatment; CEE=conjugated equine estrogen; *inclusion criteria; Ca=calcium; FSH=follicle stimulating hormone; *according to World Health Organization’s (WHO) BMD definition of osteoporosis, osteopenia and normal; *two identical trials [Delmas/GGGF (n= 601) and GGGG (n=544)]; *a crossover at 3 mos for topical vaginal interventions; likely a Lilly trial=authors’ associations suggest Lilly sponsorship yet unconfirmed via trial code information provided by Canadian representative of Eli Lilly.
Appendix 5: Classification of Trials by Bone Mineral Density of Participants at Baseline

In each qualitative or quantitative synthesis, the goal was to use trial populations’ respective baseline BMD data to categorize the studies (e.g., low versus high BMD). Some studies did not select populations according to this scheme. Thus, several trials contained a heterogeneous collection of BMD levels among the enrolled population (i.e., category 5 below) (Appendix 4). To avoid combining data from homogeneously defined populations with very low BMD and homogeneously defined studies with BMD levels higher than the 2.5 SD threshold, it was acknowledged a priori that there were five possible trial types, each of whose data could be synthesized separately:

1. homogeneous (i.e., homog) very low BMD: trials where all participants had BMD <2.5 SD below the young adult peak bone mass
2. heterogeneous (i.e., heterog) very low BMD: category 1 very low BMD trials and category 5 (i.e. mixed very low and higher BMD level) trials
3. homogeneous (i.e., homog) higher BMD: trials where all participants were diagnosed as having BMD higher than the 2.5 SD threshold (2.5 SD below the young adult peak bone mass)
4. heterogeneous (i.e., heterog) higher BMD: category 3 trials and category 5 trials
5. mixed low and high BMD: in each trial, there were participants who had a full range of BMD levels and who would all be receiving raloxifene (this is not a mix of treatment and prevention).

Adding category 5 trials to each of category 1 and 3 trial collections of pooled participants yielded two heterogeneous groups of women wherein fewer than all (i.e., <100%) of the participants had very low BMD and some had BMD levels greater than the 2.5 SD threshold. Qualitative summaries (section 3.2.1, Tables 1 to 3) and a meta-analysis (section 3.2.2, Tables 4, 6 to 8, Figures 2 to 5) were planned to evaluate the possible impact of varying BMD definitions of “population” on the empirical picture of raloxifene’s efficacy and safety.

These categories were not mutually exclusive; one trial may be included in several categories. For example, all trials in category 1 were also listed in category 2. In the MORE trial, the two subgroups fitted into different categories: subgroup 1 was low BMD; subgroup 2 was mixed, as women with prevalent fractures were enrolled regardless of their BMD.