IV. Meta-Analysis of Raloxifene for the Prevention and Treatment of Postmenopausal Osteoporosis

ANN CRANNEY, PETER TUGWELL, NICOLE ZYTARUK, VIVIAN ROBINSON, BRUCE WEAVER, JONATHAN ADACHI, GEORGE WELLS, BEVERLEY SHEA, GORDON GUYATT, THE OSTEOPOROSIS METHODOLOGY GROUP, AND THE OSTEOPOROSIS RESEARCH ADVISORY GROUP

A. Abstract

Objective: To review the effect of raloxifene on bone density and fractures in postmenopausal women.

Data Source: We searched MEDLINE from 1966 to 2000 and examined citations of relevant articles and the proceedings of international osteoporosis meetings.

Study Selection: We included seven trials that randomized women to raloxifene or placebo, with both groups receiving similar calcium and vitamin D supplementation, and measured bone density for at least one year.

Data Extraction: For each trial, three independent reviewers abstracted the data and assessed the methodological quality using a validated tool.

Data Synthesis: Data from one large dominating trial suggest a reduction in vertebral fractures with a relative risk (RR) of 0.60 [95% confidence interval (CI) 0.50–0.70, P < 0.01]. The RR of nonvertebral fractures in patients given 60 mg or more of raloxifene in the larger study was 0.92 (95% CI 0.79–1.07, P = 0.27).

Raloxifene resulted in positive effects on the percentage change in bone density, which increased over time and was independent of dose. At the final year, point estimates and 95% CIs for the differences in percent change in bone density (95% CI) between raloxifene and placebo groups were 1.33 (95% CI 0.37–2.30) for total body, 2.51 (95% CI 2.21–2.82) for lumbar spine, 2.05 (95% CI 0.71–3.39) for combined forearm, and 2.11 (95% CI 1.68–2.53) for combined hip (P < 0.01 at all four sites). Results were similar across studies, and formal tests of heterogeneity did not approach conventional statistical significance.

Raloxifene slightly increased rates of withdrawal from therapy as a result of adverse effects (RR 1.15, 95% CI 1.00–1.33, P = 0.05). The pooled RR was significant for hot flashes 1.46 (95% CI 1.23–1.74, P < 0.01) and nonsignificant for leg cramps 1.64 (95% CI 0.84–3.20, P = 0.15).

Conclusion: Raloxifene increases bone density, and the effect increases over 2 yr. The data suggest a positive impact of raloxifene on vertebral fractures. There was little effect of raloxifene on nonvertebral fractures.

B. Background

While many advocate long-term hormone replacement therapy (HRT) to prevent osteoporosis and its sequel, the randomized trials data supporting HRT impact on fracture are very limited. Recent evidence from a large, randomized, controlled trial failed to demonstrate a beneficial effect of HRT on secondary prevention of cardiovascular risk (1). Furthermore, hormone replacement may have associated risks of breast and endometrial cancer, and increases the risk of venous thromboembolism (1).

Raloxifene hydrochloride, a benzothiophene derivative, is a selective estrogen receptor modulator (SERM) (2). SERMs are nonhormonal agents that bind with high affinity to the estrogen receptor and exhibit estrogen-agonist effects on bone and estrogen-antagonistic effects on endometrium and breast (3, 4). Recent evidence from randomized trials suggests that raloxifene prevents bone loss and reduces the risk of vertebral fractures (5, 6), but up to now no one has attempted to pool data across trials.

We therefore conducted a systematic review and meta-analysis of the efficacy of raloxifene on bone density and fractures. We included all published and unpublished randomized control trials (RCTs) that estimated raloxifene effects on bone density or vertebral and nonvertebral fractures.

C. Methods

To identify relevant studies of raloxifene therapy, we used the search strategy outlined in the Section I.

a. Inclusion criteria. Studies satisfied the following inclusion criteria: 1) RCTs of at least 1-yr duration of postmenopausal women, comparing raloxifene to placebo in which treatment and control groups may or may not have received supplementation with calcium and/or vitamin D; and 2) fracture incidence or bone density data available.

b. Search and selection. To identify relevant studies of raloxifene therapy, we used key and text words: raloxifene, SERMs, osteoporosis, postmenopausal. We searched reference lists for other RCTs and obtained additional data from the United States Food and Drug Administration (FDA).

Three reviewers examined all potentially relevant trials for study eligibility. For abstracts consistent with study eligibility, we obtained the full text.

c. Methodological quality. Three reviewers independently evaluated each trial for four characteristics: concealment of randomization, intention-to-treat analysis, blinding, and the extent of loss to follow-up.

d. Outcomes and explanations for variability in raloxifene effect across studies. We examined the effect of raloxifene on fractures, both vertebral and nonvertebral, and bone density at
different sites, as well as adverse effect of the drug. We developed a priori hypotheses for fractures and bone density that might explain the heterogeneity of study results, as outlined in the Section I. Specifically, we compared groups according to prevention vs. treatment, dose of raloxifene, concurrent treatments and individual components of the quality assessment listed in the Section I.F.

e. Data collection. Three reviewers abstracted data regarding study design, patient characteristics, treatment duration, dosage, mean change for bone density, and number of fractures. We did not include data from the estrogen arm of two prevention trials (unpublished reports of the FDA).1,2 We contacted the primary authors to obtain data when important information was missing from the paper. Differences in data extraction were resolved by consensus.

f. Analysis. We chose a random-effects model for all final analyses of bone density (7). We conducted separate analyses for each bone density site. We began by constructing regression models in which the independent variables were year (1st- or 2nd-y data) and dose and the dependent variable was the effect size (Table 1).

We began with a model that included parameters for each year and dose and compared this to a model with no year parameters. We found that removal of year parameters resulted in a statistically significant (P < 0.05) reduction in fit (data for lumbar spine and hip are presented in Table 1), suggesting an increased impact on bone density with longer duration of raloxifene administration. We then compared the full model (with year and dose parameters) to a model without dose parameters and found the reduction in fit failed to reach statistical significance (Table 1), suggesting that dose did not impact on bone density. Thus, we concluded that year (that is, duration of treatment) was an important determinant of effect, but dose was not. The results were similar for all bone density sites. Thus, in subsequent analyses, we pooled across doses but not across duration of therapy. To calculate the weighted mean percent difference in bone density between treatment and control groups, we used the percentage change from baseline in the two groups and associated sd values. A test based on the $\chi^2$ distribution provided an estimate of heterogeneity between studies (7). For the hip, we pooled across different sites using one measurement site from each trial. When there was statistically significant heterogeneity for the combined hip bone density endpoint, we examined the site of measurement as a possible explanation for heterogeneity. When we found statistically

---


---

TABLE 1. Regression analyses results from the raloxifene trials

<table>
<thead>
<tr>
<th>Fuller model</th>
<th>Reduced model</th>
<th>$\chi^2$</th>
<th>P value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lumbar spine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>23.59</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>0.98</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>22.61</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>6.57</td>
</tr>
</tbody>
</table>

Model chosen

Pool all doses

Pool yr 2 and 3 leaving yr 1 separate

Combined hip

<table>
<thead>
<tr>
<th>Fuller model</th>
<th>Reduced model</th>
<th>$\chi^2$</th>
<th>P value</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>35.73</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>1.77</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>33.97</td>
</tr>
<tr>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>30, 60, 120, and 150 mg</td>
<td>[1 2 3]</td>
<td>1.89</td>
</tr>
</tbody>
</table>

Model chosen

Pool all doses

Pool yr 2 and 3 leaving yr 1 separate

---

FIG. 1. Search results for the raloxifene review.
significant heterogeneity between studies, we divided the studies into two groups based on the a priori hypotheses and then tested whether the weighted mean percent changes were different between the two groups (8).

We calculated the RR for fractures as described in Section 1, using only the random-effects model results. Heterogeneity was tested using a $\chi^2$ procedure (7). Intention-to-treat data from the individual clinical trials provided the basis for our analyses.

D. Results

a. Trial characteristics. We identified 202 articles by the search strategy and 10 from hand-searching the reference lists and conference proceedings. We retrieved 13 RCTs for closer examination, of which 6 proved ineligible, 5 because of absence of required outcome measure (4, 9–12) and 1 because of a duration of only 12 wk (13) (Fig. 1).

Table 2 presents the characteristics of the seven trials included. Four trials were treatment trials (6, 14–16), and three were prevention trials (Refs. 17 and 18 and an unpublished report of the FDA1), as defined in Section I. One trial had rates of loss to follow-up of greater than 10% (6, 8, 15, 17, 18). One trial did not report rates of loss to follow-up (16) (Table 2).

b. Fractures. Initially, after our uniform analysis plan, we pooled the results of the one large (6) and one very small trial that lasted only 1 yr (14), both of which evaluated the impact of raloxifene on fractures. Because both the results and the
size of these two trials were very disparate, the random-effects model yielded counter-intuitive point estimates and CIs, particularly for vertebral fractures. Initially, we responded to this situation by presenting both random- and fixed-model estimates for vertebral fractures. However, one reviewer of the manuscript felt very strongly that any pooling of these two very different studies was inappropriate, and we have therefore not presented any pooled results for fractures in this paper.

The 3-yr Multiple Outcome of Raloxifene Evaluation (MORE) trial (6) enrolled 7705 women, in which 6828 women had follow-up x-rays and showed a statistically significant reduction in vertebral fractures RR of 0.60 (95% CI 0.50–0.70, P < 0.01), with a narrow CI. The much smaller 1-yr Lufkin (14) trial, in which 133 of 143 women had follow-up x-rays, showed a trend in favor of the control group, with a much wider CI RR of 1.16 (95% CI 0.77–1.76, P = 0.48) (Table 3).

For nonvertebral fractures, the RR from the larger MORE trial was 0.92 (95% CI, 0.79–1.07, P = 0.27), again very different for the Lufkin trial 0.52 (95% CI 0.12–2.18, P = 0.37) (6, 14) (Table 3).

c. Bone mineral density. Table 4 presents the results of the pooled estimates across the four sites. The units reported are differences in percentage change in bone density. Total body, lumbar spine, and combined hip all demonstrated significant effects at all years examined, and combined forearm after 2-yr treatment with raloxifene (P < 0.01). Figure 2 demonstrates the effects shown at the lumbar spine site after 1 yr of treatment. We did not find statistically significant heterogeneity for the sites examined. None of our a priori hypotheses explained the heterogeneity in study results that was present. Examination of the funnel plots showed no suggestion of publication bias in the 2- and 3-yr data from any site.

d. Adverse effects and withdrawals. Our pooled estimate of the RR of discontinuing medication as a result of adverse effects from three trials (n = 8295) using 30 mg of raloxifene or more was 1.15 (95% CI 1.00–1.33, P = 0.05). For other adverse effects, the pooled RR for hot flashes from four trials (n = 9450) was 1.46 (95% CI 1.23–1.74, P < 0.01). There was a significant increase in deep venous thrombosis in the raloxifene arm, as noted in the results from the MORE trial (Ref. 6) (3.51 95% CI 1.44, 8.56, P < 0.01), as well as influenza syndrome (1.18 95% CI 1.04, 1.34; P = 0.01). For leg cramps, the pooled RR from three trials (n = 8327) was 1.64 (95% CI 0.84–3.20, P = 0.15), and for breast pain, the pooled RR was 0.97 (95% CI 0.75–1.24, P = 0.79), both of which are nonsignificant.
E. Discussion

In this meta-analysis, we performed a comprehensive literature search, specified inclusion and exclusion criteria, and conducted a rigorous data analysis. We made a systematic effort to obtain complete data from all published and unpublished studies. To calculate summary estimates of treatment for bone density, we used a random-effects model that provides a conservative estimate of treatment effect.

We found that raloxifene resulted in significant increases in bone density of the total body, lumbar spine, combined forearm, and combined hip after 2 yr of treatment. We observed larger effects on bone density after 2 rather than 1 yr of treatment, as one would expect with antiresorptive therapy. We found similar results across studies, and formal statistical tests of heterogeneity did not approach statistical significance.

In general, our group believes that pooled results across randomized trials produces the most accurate estimate of treatment effect. The raloxifene fracture data, however, stretched the boundaries of the rationale for pooling. The candidates for pooling are one large trial with 50 times more patients and 3 yr of follow-up (6), and a very small trial with only 1 yr of follow-up (14). The situation becomes more problematic because of the very disparate results, particularly with regard to vertebral fractures. After the strong suggestion of one of the manuscript reviewers, we elected not to pool fracture results across these two trials.

Using the results of the MORE trial (6), the effects of raloxifene on nonvertebral fractures were a RR reduction of only 8%, and the 95% CI included the possibility of harm. The boundary of the 95% CI that provides the largest estimate of effect consistent with the available data is a RR reduction of 21%. Thus, this analysis provides little support for an important effect of raloxifene on nonvertebral fractures.

The increase in bone density with raloxifene is smaller than seen with other antiremodeling therapies, such as HRT and bisphosphonates (19). The relatively large effect of the drug on vertebral fractures may suggest that raloxifene has a positive effect on other aspects of bone, such as bone quality. On the other hand, the very small effect on nonvertebral fractures does not support such an effect.

F. Acknowledgments

The authors thank the following trial authors for their assistance in obtaining additional data requested: Dr. P. Meunier and Dr. D. Black.

G. Bibliography