

## Letter of Response

### Relationship of Early Changes in Bone Resorption to the Reduction in Fracture Risk With Risedronate: Review of Statistical Analysis

Richard Eastell,<sup>1</sup> Rosemary A Hannon,<sup>1</sup> Patrick Garnero,<sup>2</sup> Michael J Campbell,<sup>3</sup> and Pierre D Delmas<sup>4</sup>

#### INTRODUCTION

A REPORT WAS PUBLISHED in 2003 entitled "Relationship of Early Changes in Bone Resorption to the Reduction in Fracture Risk With Risedronate."<sup>(1)</sup> Articles appeared in the popular press in 2005 raising concerns about results shown in the two figures in the paper, particularly whether there was a threshold in the relationship between bone turnover markers and vertebral fracture risk and about whether the authors had access to the data.<sup>(2,3)</sup> The *JBMR* journal editor and ASBMR chairman of publications requested in December 2005 that a reanalysis be performed that was independent of the sponsoring company.

We discussed the need for the independent analysis with Procter & Gamble (P&G), the sponsoring company. We asked for the raw data and proposed that the two authors of the original report used by the company would not be involved in the preparation of this letter. The data were provided to us by P&G in May 2006, together with a report detailing the statistical methodology and the SAS codes used. We confirmed that the bone turnover marker results we had sent the company were the same as those they returned to us. We approached a statistician (MJC) who had no prior collaboration on this project with the original authors and no links to the sponsoring company. Together we conducted a reanalysis of the data. This statistician wrote a report from which we prepared the first draft of this letter. Both documents were sent with the raw data to a second senior medical statistician in August 2006. This independent statistician read through the report, repeated the analyses, and made some valuable comments that were incorporated into this letter.

We are attaching as an appendix the checking the original analysis and figures, the description of an alternative approach to the analysis, and the listing the limitations we identified in the original analysis. In this letter, we will address three specific questions with the help of the analysis in the Appendix.

Drs Eastell and Delmas have received research grants and consulting fees from Procter & Gamble and sanofi-aventis. Dr Hannon has been employed on research grants to Dr Eastell received from Procter & Gamble. All other authors state that they have no conflicts of interest.

#### WHAT ACCESS DID THE AUTHORS HAVE TO THE DATA?

In the original paper,<sup>(1)</sup> one of the authors, a statistician working for P&G (IB), had full access to all data. P&G (like most pharmaceutical companies we contacted over this issue) used the PhRMA guidelines in relation to publication of clinical trial data, and these restrict the release of original data to investigators (<http://www.phrma.org/>). He worked closely with all of the authors of the original report on the data analysis by preparing a publication brief and responded to all requests for further analyses. Thus, the authors had full access to the analyses they had requested based on data held by one of the authors but not all had direct access to the raw data.

At the time of writing (2002/03), not all the original authors were given access to the raw data. In 2006, the American Association of Medical Colleges published recommendations regarding access to raw data (<http://www.aamc.org/research/clinicaltrialsreporting/clinicaltrialsreporting.pdf>). These proposed that the sponsor may conduct all the analyses but that the investigators should be able to conduct their own analysis if they deem it to be necessary, and we endorse these recommendations.

#### WERE THE DATA SHOWN IN THE ORIGINAL ANALYSIS THE SAME AS THOSE SHOWN IN THIS LETTER?

The original figures were based on all of the data available. However, the smoothing lines were cropped. In the redrawn figures, presented in this letter, there is no cropping of these lines.

#### HOW DO THE NEW ANALYSES AFFECT THE EVIDENCE OF A THRESHOLD IN THE RELATIONSHIP BETWEEN BONE TURNOVER MARKERS AND FRACTURE RISK?

The authors used the term "threshold" in the original paper based on visual inspection of the figures and the statistical analysis of the bone resorption marker T-score and

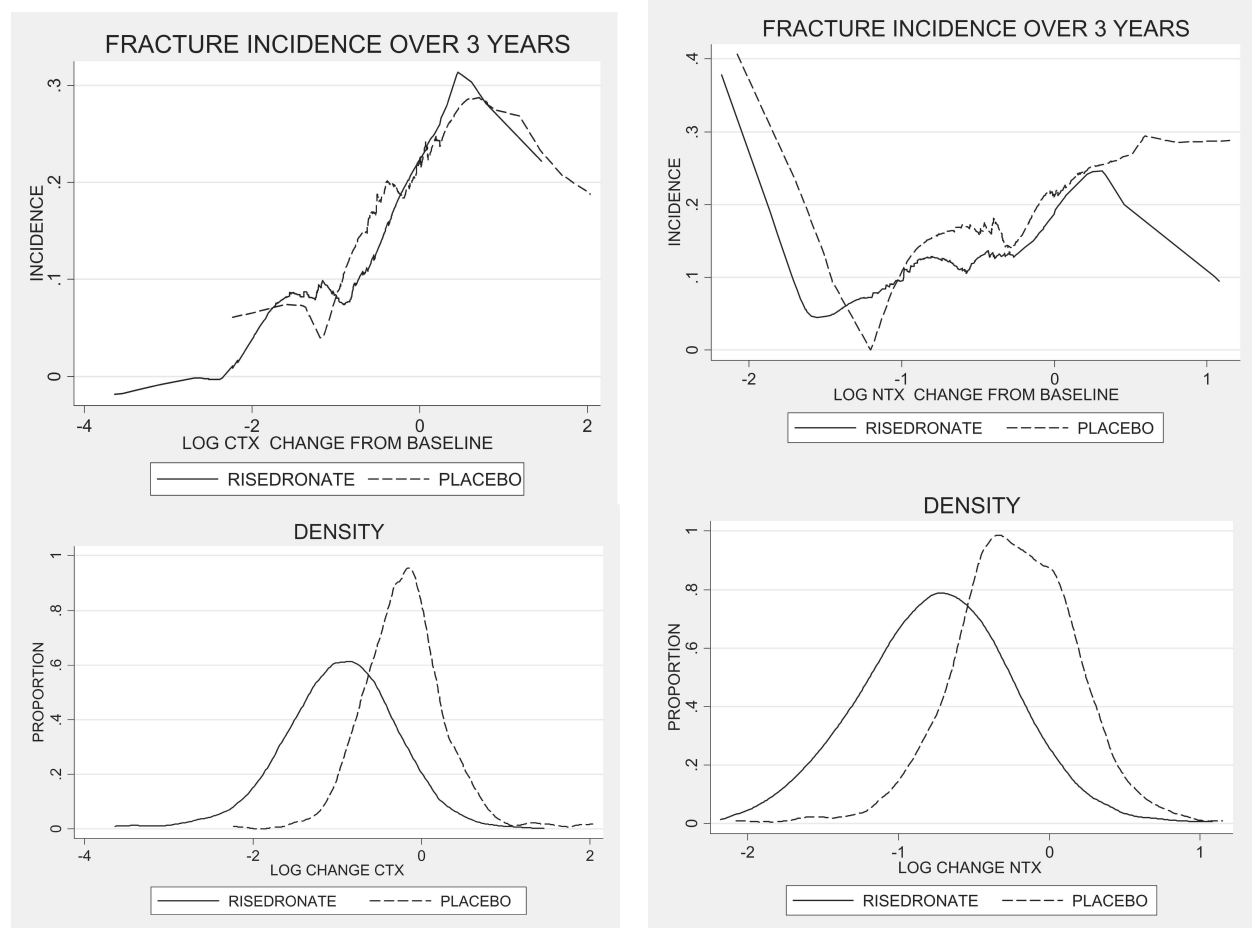
<sup>1</sup>Academic Unit of Bone Metabolism, University of Sheffield, Sheffield, United Kingdom; <sup>2</sup>SYNARC, Lyon, France; <sup>3</sup>School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom; <sup>4</sup>INSERM Research Unit 831 and Université of Lyon, Lyon, France.

TABLE 1. PERCENTAGE OF VERTEBRAL FRACTURES BY DECILE OF PERCENTAGE CHANGE IN NTX OR CTX FROM BASELINE TO 3–6 Mo (EACH CELL CONTAINS ~66 OBSERVATIONS)

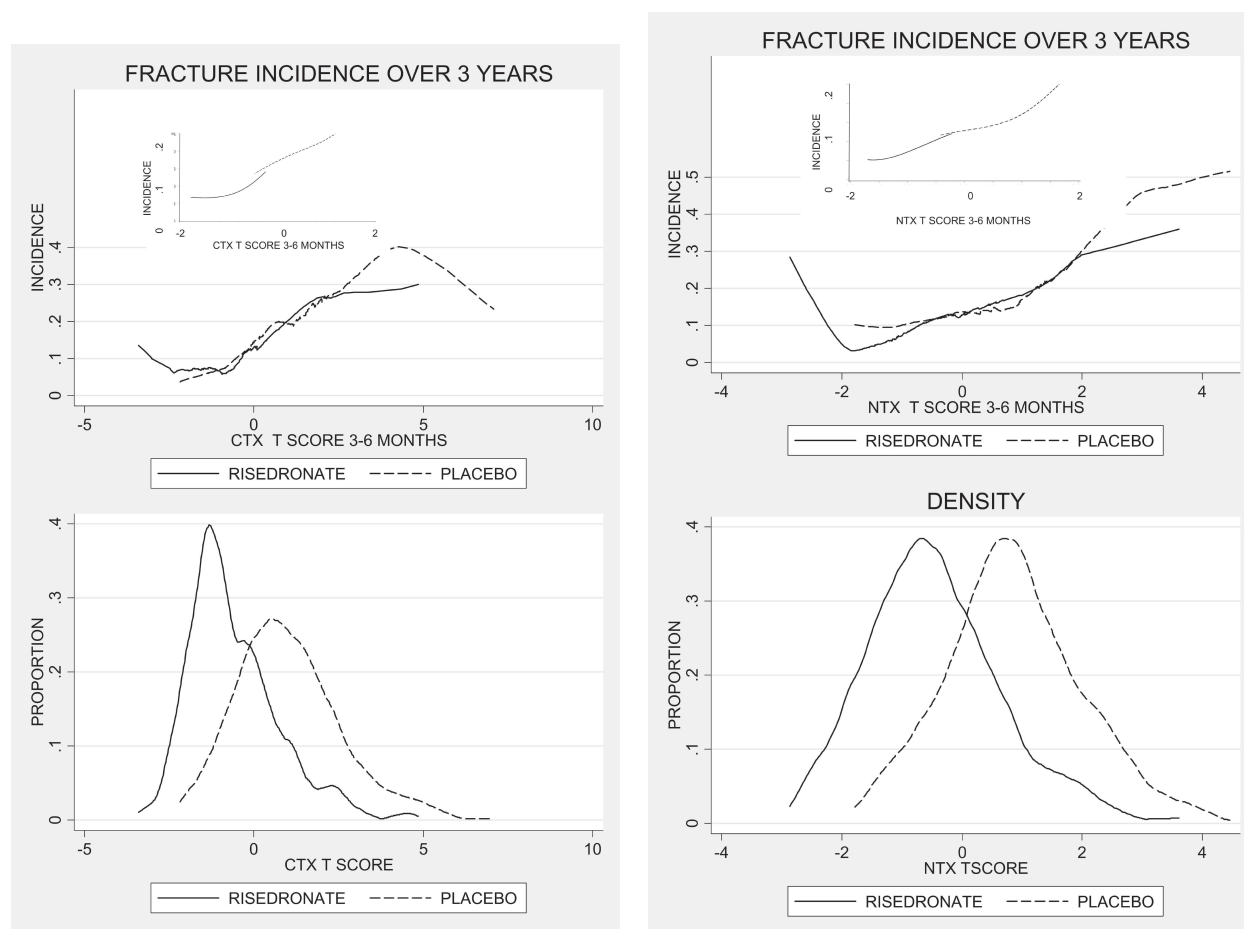
	<i>Tenth</i>									
	1	2	3	4	5	6	7	8	9	10
Percent change in NTX										
Range	<-69	-69 to -60	-60 to -52	-52 to -45	-45 to -37	-37 to -30	-30 to -22	-22 to -6	-6 to 14	>14%
Percent fractures	6	11	11	21	21	5	14	18	21	25
Percent change in CTX										
Range	<-77	-77 to -68	-68 to -59	-59 to -51	-51 to -41	-41 to -30	-30 to -20	-20 to -4	-4 TO 17	>17%
Percent fractures	3	15	4	7	16	19	18	18	22	28

TABLE 2. PERCENTAGE OF VERTEBRAL FRACTURES BY DECILE OF NTX OR CTX T-SCORE AT 3–6 Mo (EACH CELL CONTAINS ~66 OBSERVATIONS)

	<i>Tenth</i>									
	1	2	3	4	5	6	7	8	9	10
NTX T-score										
Range	<-1.4	-1.4 to -0.97	-0.97 to -0.60	-0.60 to -0.23	-0.23 to 0.13	0.13–0.49	0.49–0.82	0.82–1.3	1.3–2.0	>2.0
Percent fractures	4	7	9	10	17	13	16	15	21	37
CTX T-score										
Range	<-1.7	-1.7 to -1.2	-1.2 to -0.9	-0.9 to -0.4	-0.4 to 0.0	-0.0 to 0.4	0.4–0.9	0.9–1.5	1.5–2.4	>2.4
Percent fractures	6	7	8	9	4	20	16	26	22	33



**FIG. 1.** The relationship between the incidence of vertebral fractures (as a fraction) over 3 yr and the change in bone resorption marker (the difference between log marker on treatment and log marker at baseline, equivalent to  $\log [\text{on treatment/baseline}]$ ) was examined using lowess plots (using STATA statistics package) with 0.6 bandwidth CTX (A), and NTX (B). The bottom panels show the density distribution for the risedronate and controls groups, respectively.



**FIG. 2.** The relationship between the incidence of vertebral fractures (as a fraction) over 3 yr and the bone resorption level at 3–6 mo on treatment (expressed as a T-score) was examined using lowess plots (using STATA statistics package) with 0.6 bandwidth for CTX (A), and NTX (B). The bottom panels show the density distribution for the risedronate and controls groups, respectively. The insets are from Fig. 2 from the original article, with the incidence given as a fraction rather than a percentage.<sup>(1)</sup>

fracture risk; the latter showed that the relationship was best fit by including a quadratic term (in addition to a linear term). We have been able to confirm this finding in our reanalysis.

We went on to divide bone resorption marker percent changes into deciles (Table 1 of the appendix) and we concluded for CTX (but not NTX) that there was a critical point after which no further decrease in vertebral fracture risk associated with further decreases in bone resorption markers. The critical value of this percent decrease was 51%.

We divided bone resorption T-scores into deciles (Table 2 of the appendix), we went on to fit a Cox regression to the data, and we concluded that the decile approach supports the hypothesis that there is a threshold level for CTX T-scores (but not for NTX), below which a further reduction in bone resorption is not associated with a further reduction in fracture risk. The critical value of this T-score for CTX was 0. Clearly, this hypothesis needs testing in other studies, and the *p* value should be treated with caution because it was calculated posthoc.

### DO THE RESULTS OF THE SECOND ANALYSIS CONFIRM THE CONCLUSIONS MADE IN THE ORIGINAL PAPER?

We stated three conclusions at the end of our paper.<sup>(1)</sup> The first was “The baseline level of bone resorption is related to subsequent fracture risk.” This conclusion can still be supported. The second was “The reduction in bone resorption explains, in part, the reduction in the risk of fractures with risedronate.” This conclusion has been confirmed using the method of Li et al.<sup>(4)</sup>

The third conclusion was “There is a level of bone turnover reduction below which there no further fracture benefit is observed.” This third conclusion was supported in the original paper by fitting linear, quadratic, and cubic functions to the relationship between vertebral fracture risk and bone resorption level, and we have confirmed this relationship is not linear for both NTX and CTX. It was also supported by the appearance of this relationship as shown in Figs. 1 and 2. This is where the two analyses differ. The original smoothing curves were cropped extensively and in an asymmetric manner. When we redrew the graphs, we

noted that there was, in fact, an apparent increase in fracture risk at high percent reductions (Fig. 1) and low levels of markers (Fig. 2), particularly for NTX. We pointed out five reasons why smoothing can be misleading (Appendix, New Analysis, Decile Approach) and not the best method to use. We examined the data using a decile approach and found a reduction in vertebral fractures at high percent reductions (Table 1) or low levels of markers (Table 2). We went on to use the decile approach to examine the relationship between bone turnover and vertebral fracture risk. For percentage change in CTX (not NTX), we observed that there was no further decrease in vertebral fracture risk associated with further decreases in bone resorption markers. For T-score, we found evidence for a threshold for CTX at a T-score of 0; for NTX, the data can support a threshold or a slowly declining risk. Thus, the third conclusion can still be supported based on the new analysis.'

## ACKNOWLEDGMENTS

We thank Prof Stephen Evans from the London School of Hygiene and Tropical Medicine for providing independent statistical analysis and confirmation of the full data.

## REFERENCES

1. Eastell R, Barton I, Hannon RA, Chines A, Garnero P, Delmas PD 2003 Relationship of early changes in bone resorption to the reduction in fracture risk with risedronate. *J Bone Miner Res* **18**:1051–1056.
2. Revill J. 2005 How the drugs giant and a lone academic went to war. *The Observer* **4**:10–11.
3. Baty P. 2005 Journal warned about P&G data. Available at <http://www.thes.co.uk/search/story.aspx>. Accessed December 9, 2005.
4. Li Z, Meredith MP, Hoseyni MS 2002 A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Stat Med* **20**:3175–3188.

Address reprint requests to:

Richard Eastell, MD  
Metabolic Bone Centre  
Sorby Wing  
Northern General Hospital  
Herries Road  
Sheffield, South Yorkshire S5 7AU, UK  
E-mail: [r.eastell@sheffield.ac.uk](mailto:r.eastell@sheffield.ac.uk)

## APPENDIX

### Confirmatory analysis

We confirmed as correct the numbers of subjects and their description in Table 1 of Eastell et al.<sup>(1)</sup> We found the same reduction in bone resorption markers (BRMs) and the same *p* value, although we used a Wilcoxon rank sum test rather than a signed rank test. We examined the relationship between baseline characteristics and the subsequent risk of incident fractures over three years using the same approach (Cox regression) and obtained the same *p* values as in Table 2 of Eastell et al.<sup>(1)</sup> and obtained the same *p* values with one exception; the *p* value for the coefficient relating to prevalent vertebral fractures was 0.022 and not 0.001.

Paragraphs 3 and 4 of the Results section of Eastell et al.<sup>(1)</sup> refer to the relationship between changes in NTX and CTX (at 3–6 mo) and fracture risk. From the P&G statistical report, it was

apparent that the analysis was conducted on the 3- to 6-mo values alone, in line with the concept of using bone turnover markers as surrogate endpoints. Using the 3- to 6-mo values, we found similar *p* values to those in paragraph 3 of Eastell et al.<sup>(1)</sup> and evidence that combined quadratic and cubic terms of the BRM improved the fit of the model over a linear fit.

We used the method of Li et al.<sup>(4)</sup> to show that the NTX and CTX values at 3–6 mo for NTX and CTX explain 66% and 67% of the reduction in vertebral fracture risk with treatment at 3 yr, respectively, the same figure as given in paragraph 4 of the Results section in Eastell et al.<sup>(1)</sup> The *p* value for the relationship between T-scores and vertebral fracture incidence refer to the 1-yr risk, and the *p* value for NTX should be 0.47 rather than 0.047.

### Reanalysis of Fig. 1 of Eastell et al.

In Eastell et al.,<sup>(1)</sup> Figs. 1 and 2 were described as being “constructed using a smoothing curve.” From the P&G statistical summary, this was described as a cubic spline and was calculated using the computer package SAS. However, when we examined the distribution of percent changes in NTX and CTX, we found them to be highly skewed. The range of the percentage change of NTX was –88% to 215% and that of CTX was –97% to 659%.

We decided to use an alternative measure with less skewness, defined by subtracting the logarithm of the BRM at 3–6 mo from the logarithm of the BRM at baseline. Instead of a cubic spline, we used a LOWESS plot with 0.6 bandwidth (STATA version 9; Statcorp, College Station, TX, USA), an alternative method of smoothing that is easier to implement. The 3-yr vertebral fracture incidence data are shown in Figs. 1A and 1B for CTX and NTX, respectively. The data show generally that low values of BRMs are associated with fewer fractures and also how outliers of the BRM affect the shape of the plots. This plot is not directly comparable to Fig. 1 of Eastell et al.<sup>(1)</sup>

### Reanalysis of Fig. 2 of Eastell et al.

We also used a LOWESS smoother to examine the relationship between the T-scores at 3–6 mo of the two BRMs and the incidence of vertebral fractures at 3 yr (Fig. 2). These also show that generally the lower the T-scores, the lower the risk, but also show how outliers affect the shape of the relationship at the extremes. Figure 2 is comparable to the original Fig. 2 of Eastell et al.<sup>(1)</sup> Within the same range of the T-scores, the figures are similar. However, it is apparent that the original graph shows only a restricted range of the T-scores. This is discussed later.

### New analysis: decile approach

The application of smoothing to these data is challenging for the following reasons: (1) the outcome variable is 0/1, and so heavy smoothing is needed; (2) the data are quite sparse at the extremes; (3) the predictor variables have skewed distributions, meaning that outliers may have undue influence; (4) time exposed is not accounted for; and (5) it is difficult to display levels of uncertainty on these figures.

We wanted to apply a method that would not impose too many restrictions on the shape of the relationship. To this end, we divided the distribution of the percent change in BRMs into tenths. The percentage of fractures in each tenth and the corresponding range of percent change in NTX or CTX is given for the treatment and control groups combined in Table 1. There is no simple relationship between percentage change in NTX and risk of fracture, and the risk is not flat below a 35–40% decrease as claimed.<sup>(1)</sup> There is also no simple relationship between percentage change in CTX and risk of fracture, and there is an unusual increase to 15% (10 fractures from 65 patients) between –77% and –68%. Although the claim for a threshold appears untenable, the statement in the original Results section would be supported by this table for CTX: “there was no further decrease in vertebral fracture risk associated with further decreases in bone resorption markers.”

The percentage of fractures by decile of NTX or CTX T-score is given for the treatment and control group combined in Table 2.

In general NTX T-score and risk of fracture decreased together. For CTX T-score, there seems to be a considerably greater risk of fracture above the median (T-score > 0), and it is plausible that, below the median, the population risk remains constant. We cannot draw the same conclusion for NTX.

To investigate this further, a Cox regression was fitted to the risk of fracture at 3 yr with indicator variables (numbered 1–10) in the model to denote the deciles of the T-scores. Drug and placebo groups were modeled separately and stratified by trial. For both CTX T-scores and NTX T-scores, there was a highly significant ( $p < 0.01$ ) relationship between the predictors and fracture risk in the placebo arm and significant relationships ( $p < 0.05$ ) in the treatment arm. Whether the coefficients associated with the lower deciles differed from one another was analyzed using a posthoc approach testing procedure (STATA postestimation). This decile approach was explored further with the CTX T-score only because there was no simple relationship with the NTX T-score. For both the placebo and the risedronate groups, there was evidence for similar coefficients below the median (T-score of 0, equivalent to a urinary CTX of 4.2 nmol/mmol creatinine;  $p = 0.57$  for test of the null hypothesis that coefficients attached to the contrasts comparing the lowest tenth to the next four higher tenths are equal to each other.)

Thus, the decile approach supports the hypothesis that there is a threshold level for CTX T-scores (but not for NTX), below which a further reduction in bone resorption is not associated with a further reduction in fracture risk. Clearly, this hypothesis needs testing in other studies, and the  $p$  value should be treated with caution because it was calculated posthoc.

#### *Limitations of the original analyses*

We were able to confirm most of the analyses in the original report. However, the two independent statisticians identified some errors and some poor practice.

The Wilcoxon signed rank test (for paired data) had been used to compare changes in BRMs from baseline to 3–6 mo; the correct test would have been a test for unpaired data, the Wilcoxon rank sum test.

The cubic spline approach is sensitive to outlying points and so it is usual to crop such lines, especially when the distributions are extremely skewed. In the original paper, we should have given a rationale for the approach used in cropping these lines and stated how much data were cropped. However, we were unaware that the cropping procedure was carried out in an asymmetric way at the time we wrote the original manuscript and therefore that was not indicated in the legend of Figs. 1 and 2. The axes for Fig. 1 in Eastell et al.<sup>(1)</sup> are –65 to 0 for NTX percent change and –70 to +5 for CTX percent change. We calculate (based on Tables 1 and 2) that this range excludes between 34% and 49% of the more extreme values.

In the analysis, the 3- to 6-mo values were inadvertently referred to as changes in NTX and CTX (at 3–6 mo). This was an authors' error. Using the 3- to 6-mo values rather than the change from baseline is better because the former reflect the post-treatment levels and thus gives a better measure of future risk.

One issue is that the average value for the 3- and 6-mo BRMs was used in the analysis, and this resulted in a value for each individual that might have been based on only one measurement if the other was missing. It would have been better to have weighted the values for imputed measurements. However, there were relatively few measurements based on only one measurement, so this was not an important criticism. Table 2 in Eastell et al.<sup>(1)</sup> reported  $p$  values from the Cox regression analysis; it is better practice to report the regression coefficients and confidence intervals. The discrepancy in  $p$  values for baseline vertebral fracture in this table resulted from the use of the entire VERT cohort in the original analysis (in error), rather than just the bone turnover marker subset that was used in the current report.

There were some other discrepancies in  $p$  values for the quadratic and cubic analysis, but such findings can occur when different methods used (we used the exact partial likelihood approach) or if terms are entered into the statistical model as “strata” rather than “terms,” because these allow for different or common underlying hazards, respectively.