SPECIAL EDITORIAL ON A HOT TOPIC

Monitoring Osteoporosis Therapy with Bone Densitometry: A Vital Tool or Regression toward Mediocrity?

In a recent issue of the Journal of the American Medical Association, Cummings et al. (1) demonstrated the phenomenon of regression to the mean when osteoporosis therapy was monitored by bone densitometry using data from two major clinical trials, the Fracture Intervention Trial (FIT) and the Multiple Outcomes of Raloxifene Evaluation (MORE) trial. Bone density data from the first 2 yr of each trial in compliant women only was reviewed. In FIT, 2634 women in the intervention group and 2603 women in the control group had the requisite bone density data and pill count to be included in the analysis. In the MORE trial, 3954 women met the criteria for inclusion in this analysis. Briefly, the authors reported that 82% of the compliant women in the intervention group from FIT had an apparent gain in total hip bone mineral density (BMD), whereas only 1.4% had an apparent loss of more than 4% in the first year. In the MORE trial, 75% of the compliant women in the intervention group had an apparent gain in femoral neck BMD, whereas 8% had an apparent loss of more than 4% in the first year. In both cases, those women who lost the most during the first year were observed to be the most likely to gain BMD during the second year. Conversely, those women with the greatest gains in BMD during the first year were most likely to lose BMD or have little, if any, gain during the second year. The authors stated that similar findings were noted at the spine in the intervention group from FIT and at both skeletal sites in the placebo group in FIT.

The authors attributed the gains in bone density in the overwhelming majority of women in the intervention groups of both studies to the efficacy of the drugs used. The finding of gains in bone density in the second year even after an apparent decline in the first year was additional proof of efficacy, according to the authors. That great gains in the first year were followed by losses in the second year and great losses in the first year were followed by gains in the second year was attributed to a phenomenon known as “regression to the mean.” The authors reasoned that monitoring changes in bone density in women on alendronate or raloxifene was of questionable value because virtually all compliant women on these agents gain bone density eventually. They noted that it was likely that the phenomenon of regression to the mean was responsible for the paradoxical swings in bone density in compliant women, rather than drug failure or drug resistance. Regression to the mean, they theorized, likely exists for other quantitative measurements such as lipids, blood pressure, and spirometry values, as well as bone density. Monitoring these and other types of parameters of treatment in clinical practice are likely affected by regression to the mean.

Regression to the mean was first described by Sir Francis Galton in 1886 (2). In a study of the hereditary nature of height using a population of parents and their children, Galton observed that the children of the taller parents had a mean height that was closer to the mean height of all children than the mean height of their parents was to the mean height of all parents. Similarly, if one looked at the taller children, the mean height of their parents was closer to the mean height of all parents than was the mean height of these children to the mean of all the children. Galton concluded, “Each peculiarity in a man is shared by his kinsman, but on the average in a less degree.” Galton called this “regression towards mediocrity.” Today, it is called regression to the mean. Regression to the mean will occur whenever a variable is measured on two separate occasions, when the value of the variable can change in the individual either due to biologic or measurement variability or both, and when a subgroup of the whole group is defined on the basis of a high or low value at the first measurement (3). In these circumstances, the average value for the second measurement of the variable in the subgroup will always be closer to the mean than was the first. In the FIT and MORE trials individuals with low bone density were selected (4, 5). A variable bone density was measured. The bone density in an individual could change, either because of the imperfect precision of bone density measurements, the effect of the therapeutic agent, or both. In the analysis by Cummings et al. (1), the individuals were then divided into subgroups based on the apparent change in bone density during the first year of each study. These conditions met the necessary requirements to create the phenomenon of regression to the mean when the change in bone density in the second year was determined for each subgroup. In these circumstances, regression to the mean occurred with bone density measurements, just as Galton could have predicted it would, so long ago. The change in bone density during the second year of each study regressed toward the mean change for the entire group at the end of the
first year. The study of Cummings et al. (1) does not prove that regression to the mean exists for bone density measurements. It illustrates it.

Understanding regression to the mean is important because if its existence is not appreciated, two mistakes can be made in the interpretation of data derived from quantitative measurements in clinical trials. The first is to mistake efficacy of a therapeutic intervention for regression to the mean (6). The second is to conclude that the difference between two measurements is related to the value of the first measurement.

In the FIT and MORE trials, as has been commonly done in trials designed to evaluate agents that reduce fracture risk, individuals with low bone density were selected for treatment. These are individuals at an extreme of the distribution of values for bone density. Because of regression to the mean, the average value for bone density in such a group of individuals is expected to increase at the second measurement even if no efficacious intervention is given. To avoid mistaking regression to the mean for efficacy, the change in bone density in an intervention group must be compared with the change in bone density in a control group, in which regression to the mean will also occur. Efficacy can then be demonstrated by looking at the difference in bone density between the two groups, allowing regression to the mean in both groups to cancel out. The authors of this paper, indeed, noted that regression to the mean occurred in the compliant women in the control group of FIT. Nevertheless, the authors focused on the changes in bone density in the intervention groups only and concluded that both alendronate and raloxifene were efficacious. In the absence of comparisons to a control group, some authors have suggested using an initial measurement to classify subjects but a subsequent measurement as the baseline value for further assessments. Another approach is to perform replicate measurements of the variable to reduce the random variation. Cummings, et al. (1) did not use the former approach in the analysis, and replicate measurements were not made in the FIT and MORE trials. Thus, although there is ample evidence that both alendronate and raloxifene are, indeed, efficacious drugs, this analysis by Cummings et al. (1) does not provide additional proof of efficacy. One must remove the effects of regression to the mean to prove efficacy. Regression to the mean cannot and should not be used to prove efficacy.

The second mistake that a failure to appreciate regression to the mean can cause is to conclude that the difference between two repeated measurements is related to the first measurement (6). Cummings et al. (1) noted that individuals who lost bone density in the first year of treatment tended to gain bone density in the second year. Similarly, those individuals who made large gains in the first year gained little or even lost in the second year. This is, again, the result of regression to the mean and is a spurious finding caused by the mathematical relationship between the change in bone density in the second year and the value at baseline and the end of the first year. The measured value for bone density at baseline and the end of the first year contributes to the calculation of the percentage change in both the first and second years. Different authors have proposed various methods for overcoming this dilemma. One such method was proposed by Oldham (7) in which the change in a variable is related to the average of the first and second measurements rather than only the first measurement. There is insufficient data presented in the article to independently perform Oldham’s (7) calculation. In the absence of such a correction, the conclusions are suspect and we should not presume to predict the change in bone density in the second year based on what was seen in the first year.

Whereas these discussions may seem esoteric, the authors of the March 8, 2000, Journal of the American Medical Association article state that their findings “raise questions about the value of monitoring BMD during treatment.” They note that the phenomenon of regression to the mean “may influence other measurements that are used to monitor many types of treatments used in clinical practice.” And herein lies the rub. Regression to the mean relates to the experience of the whole group and not to any defined individual. It is a statistical phenomenon created by the circumstances of the type of analysis and not a biological one. Therefore, to suggest that measurement of bone density in an individual to assess biological efficacy of a therapeutic agent is not worthwhile because of regression to the mean is wholly incorrect. In dealing with changes in bone density in an individual under any circumstances, the statistical phenomenon of regression to the mean has no application. The statistical issue to be considered here is the precision of the measurement. Assuming an excellent in vivo precision of 1% at the total hip and using the mean change in BMD during the first year for the various subgroups, 831 of the 2634 women in the intervention group from FIT on an individual basis would have been considered as having no significant change in BMD from baseline at 95% confidence. The change in total hip BMD in 1803 women would have been considered statistically significant. Continuing to use the reported mean changes in BMD for the various subgroups, only 37 women would have been judged as having a significant change in bone density from the end of the first year to the end of the second year. If the change from baseline to the end of the second year is the time frame considered, 1359 of the 2634 women had significant increases in total hip BMD, with the remainder of the women having no significant change. The precision of the measurement, not regression to the mean, is also the paramount issue for measurements of cholesterol, blood pressure, or any other quantitative measurement made in an individual. The bone densitometry literature abounds with articles on precision, its importance, and its effect on the timing and interpretation of repeat measurements. Our colleagues in other fields who use quantitative measurements to monitor their patients would do well to follow our lead in addressing issues of precision. But the misapplication of regression to the mean to justify not repeating a quantitative measurement in an individual in clinical practice would be tragic. Physicians would never measure anything a second time to assess
efficacy. We would simply have to cross our fingers and hope, while we regressed toward mediocrity.

Sydney Lou Bonnick
Institute for Women’s Health
Texas Woman’s University
Denton, Texas 76204

References