We believe that the points raised are critical. If they are not dealt with, patients needing advance therapies may suffer as important advances are delayed or prevented from ever reaching them.

We anticipate that the Division and the Committee will continue to encourage constructive dialogue with industry today and in the future. Additionally, it will be important to keep communications open with the CPMP and with public health agencies such as the NIH to address these critical questions and to provide recommendations for workable new guidelines for developing osteoporosis therapies. Thank you very much.

CHAIRMAN BRAUNSTEIN: Thank you, Dr. Dere. The next speaker is Dr. Thomas Marriott, Vice-President, Development Research, NPS Pharmaceuticals.

DR. MARRIOTT: Good morning, Mr. Chairman. Thank you very much for giving us the opportunity to make some comments. We certainly appreciate the work the Committee is doing in tackling this difficult area.

This morning's speakers have done a good job of outlining many of the points and much of the data that need to be considered in the design of clinical trials and the clinical evidence necessary for the approval of new osteoporotic agents.

The summaries of the Guidance documents also make it clear that there's an urgent need for an ICH-like harmonization of the meaning of the terms "prevention" and "treatment" as they relate to osteoporosis and of the regulatory requirements for the approval of new agents.

At this point NPS still believes that the randomized double-blind calcium and vitamin D controlled trials are the best way to evaluate the safety and efficacy of new osteoporotic agents. However, our recent experience suggests that it is becoming increasingly difficult to conduct calcium and vitamin D controlled trials.

We are currently in the middle of a 2600 patient randomized calcium and vitamin D controlled trial in nine countries. In 2000, when we were initiating the study in the U.S. and Canada, several

IRBs refused to approve the study because we included women with severe osteoporosis; that is, women with a BMD of less than minus 2.5 and a prevalent fracture.

In 2001, as we expanded the study worldwide, two of the multiple research ethics committees, the MREX in the UK and the Central Ethics Committee in Denmark would not approve the study because they considered it placebo controlled and requested that we add an approved agent to the calcium and vitamin D control group.

Thus, if we're to continue to employ calcium and vitamin D controlled studies, the scientific and regulatory communities must clearly describe why this study design is appropriate and better than alternative study designs.

We must also demonstrate that we have reduced the risk to our patients as much as possible. We suggest that there are at least three ways to minimize the risk to our patients. First, it's possible, we believe, to reduce the number of clinical studies.

Harmonization of the definitions of

treatment and prevention may allow both indications to be investigated in a single trial. An obvious example would be the study of a true anabolic agent where a reduction in fracture incidents is demonstrated and virtually all patients show an increase in BMD, beginning in the osteoporotic range and increasing through the osteopenic range.

A second study to specifically investigate prevention divined by an increase in BMD should not be necessary in this case. Secondly, we would believe that it should be possible to reduce the recommendation -- the recommended duration of clinical trials.

It is clear that it is possible to demonstrate statistically significant increases in BMD in short periods of time with many agents, and statistically significant decreases in the vertebral fracture, incidence, for example, in less than three years.

The recommended duration of efficacy studies required for approval, we believe, should therefore be considered and in fact should be less in

fact than two years. We believe it's also possible to reduce the number of patients in clinical studies.

There are two ways, at least, to reduce the number of patients participating in clinical trials from the osteoporotic agents. The first is to use the one-sided test to determine efficacy when the control group is calcium and vitamin D.

For example, it is obvious that a treatment would need to demonstrate better efficacy than that of calcium and D. Therefore, the null hypothesis is whether the incidence of fractures in patients receiving the experimental treatment is lower than the incidence of fractures for patients receiving calcium and vitamin D, not whether there is a difference in the incidence.

This question can be answered using a one-sided T-test or one-sided test, and the use of the one-sided test should in fact reduce the number of patients in the trial by 15 to 20 percent.

A second way to reduce patient numbers is to accept a lower level of confidence, for example, 80 percent, for the reduction in fracture incidence at a

second fracture site once reduction in the fracture incidence at the first site has been demonstrated.

Since the significance level is the risk of concluding a difference exists when in fact there is no difference, the level of significance is chosen based on the consequences of this decision. Therefore, if a treatment has been demonstrated to reduce the incidence of vertebral fractures, for example, the question of whether it also reduces the incidence of fractures at another site, for example, the hip, should addressed using a lower level of confidence.

This does not substantially increase the risk of concluding that an agent with a deleterious effect at the second site is better than the control, but will require fewer patients and fewer fractures at the second site to reach the appropriate conclusion. Thank you for your time, and again, I thank you for your efforts in taking on this task.

CHAIRMAN BRAUNSTEIN: Thank you, Dr. Marriott.

Our last speaker is Ms. Amy Alina, from

the National Women's health Network.

MS. ALINA: Hi. I'm speaking here on behalf of the National Women's Health Network, which is a nonprofit organization that advocates for national policies that protect and promote all women's health.

We also provide evidence based, independent information to empower women to make fully informed healthcare decisions, and the network does not accept financial support from pharmaceutical or medical device companies.

We're supported by a national membership of about 8,000 individuals and 300 organizations. We're here today representing the concerns shared not just by our members, but also by millions of women who, particularly in the wake of the news this summer about the Women's Health Initiative results, are really struggling with questions about the safety, effectiveness and the need for drugs prescribed at menopause.

And while the topic of this meeting is clinical trials for new osteoporosis treatments, it

touches on issues that go far beyond clinical trials and affect the way that women are educated about bone health, screened for bone density loss, counseled on prevention strategies and finally treated for osteoporosis.

We recognize that this Committee and the FDA do not control all those aspects of women's healthcare, but we address them in our comments because the way that clinical trials for osteoporosis drugs, particularly prevention trials, are designed will have consequences for women's health education and care.

In the 1980s and earlier we were among the women's health advocates who agreed that the problem of bone fractures and their effect on elderly women's quality of life was being overlooked by the medical community and needed to be addressed.

And today, we think the pendulum has swung to another extreme for those women who do have access to healthcare and insurance coverage. Now, we believe it's the case that women who are in the healthcare system are commonly over-treated -- over-screened,

over-diagnosed and over-treated for problems relating to their bones.

At the same time it's still true that there are women who would benefit from screening and treatment who don't get the care they need as a result of economic and other barriers to health services. But bone density screening has become a rite of passage for women approaching and entering menopause, and this means that women are being screened in their '40s and '50s, which we believe is far too early to use a test that hasn't been shown to be a reliable predictor of fractures that typically occur 20 to 30 years later.

And this is a problem, though many people might ask what the harm is in taking a measure of bone density. The assumption that osteoporosis screening must be a good thing fails to recognize its limitations or how it plays into the medicalization of menopause.

So the problem is that over-screening leads to over-treatment, and many of you who see patients must hear, just as we do in our office, from

women who tell story after story about how they've been told that they have the disease of osteopenia, that they need a prescription for their borderline osteopenia.

And once women are diagnosed in this way they may be much less likely to do many of the things that could help them maintain their "borderline bone health," like staying physically active. So we're very concerned about that.

And we're also concerned about the fact that many of these women are given prescriptions and told that they have to take drugs to prevent their osteopenia from developing into osteoporosis, and then leading to bone fractures, the slippery slope.

Some of them may need help from a drug to prevent serious bone loss and debilitating fractures, but some of them don't, and the bone density test is not a sufficiently reliable predictor of fractures to support that use of it.

The experience of hormone replacement therapy should serve as a warning, we think, as an example of a drug that was prescribed to millions of

women based on false assumptions about unproven benefits and inadequately tested safety.

So how does this relate to the discussion issues that you have to address? Clinical trial design, as I said, doesn't control clinical practice, but it does have an affect on it.

And as you think about the answers to the questions that the FDA staff has posed to you, we urge you to put them in the context of how the clinical trials of new osteoporosis treatments will affect the way that the drugs tested will be put into use in clinical practice and the way they will therefore, affect women's lives.

I'm going to respond to a couple of the specific questions, first on efficacy. When is bone mineral density an adequate primary endpoint? Well, my guess -- our answer to this question is never, and we recognize that this puts us somewhat outside the mainstream of discussion.

But we're not alone in questioning the value of bone mineral density measures, and it's already been pointed out this morning that the NIH

Consensus Conference two years ago, the report itself raised questions about the accuracy of bone mineral density testing and recommended that more comprehensive ways of assessing risk for fracture should be studied, and so we're echoing that.

On the question of duration of study, what duration of study is appropriate for assessment of effectiveness, we link this very much with the age of women included in the trials.

If study durations are going to continue to fall in the two- to three-year range, or certainly, if they were to be shortened we think it's important that the prevention trials not be conducted on women who are younger than 65, unless those women are at particularly high risk for bone fractures, because of, you know, early removal of ovaries or long-term steroid use.

It isn't possible to determine in the short term of a two- to three-year trial whether a drug has effectively prevented bone fracture in a younger woman who hasn't yet reached the age at which she's likely to experience bone fractures.

On the question of using a placebo or an active control, the new understanding of the risks posed by hormone therapy significantly changes the terms for discussing this issue.

While the Women's Health Initiative showed us that estrogen plus progestin is highly effective for osteoporosis and will likely show the same for estrogen alone, it also demonstrated that the combined hormone regimen poses serious health risks which outweigh its benefits for healthy women.

So we would say that we can't hold out HRT to be the standard comparison for a trial of the new osteoporosis. And in prevention trials we still believe it's appropriate to use a placebo control by which we mean vitamin D and calcium.

In treatment trials, however, where participants have experienced a fracture prior to beginning in the trial, we think an active control is both ethical and appropriate, and we would also say it's desirable, because it will provide more valuable and useful results showing whether a new drug offers a benefit over existing options in terms of either

efficacy or safety.

On the safety question specifically, the instance of osteoporotic fractures being used as a safety rather than an efficacy endpoint, well, I guess we would say that it should be used as both a safety and efficacy endpoint in prevention and treatment trials, and reiterate our statement that the intermediate endpoint of bone mineral density isn't an adequate measure.

And in terms of duration of study needed for assessment of safety, we recognize it's not practical to require sponsors to conduct trials that last ten or more years. I'm sure the speakers who went before me would agree with that.

But women who are prescribed drugs for osteoporosis are likely to be taking them for decades. And so we would say it's necessary to gather data on safety of such long-term use and that FDA should recognize this need by making long-term follow-up studies on these products a condition of approval, and by putting in place active systems for monitoring adverse reactions to the drugs. Thank you.

CHAIRMAN BRAUNSTEIN: Thank you.

(End of this portion of proceedings; 12:40 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:32 p.m.)

CHAIRMAN BRAUNSTEIN: Our next speaker is Dr. Steven Cummings. He's going to discuss the size, scope, and implications of placebo versus active control trials.

DR. CUMMINGS: Thank you. I'd like to thank the committee, particularly Eric Colman for his invitation to come and talk to you about this set of issues.

I've probably got the most diffuse task of the day, and that's to cover a lot of issues and hopefully stimulate some discussion about a few important issues and principles that I'd like to propose.

To start off with, I'd like to acknowledge a lot of help from a couple of other people in putting this together. As I am not a biostatistician, I'll be showing a lot of numbers that were generated with the help of two statisticians: Dr. Charles McCulloch, who is Chief of Biostatistics at UCSF, as well as my colleague Dr. Dennis Black, who has been involved in

the design and implementation of a number of trials in the area of osteoporosis and fracture prevention.

What I'm going to cover first will be just a few comments about the rates of fractures and risks in placebo-controlled trials as a background and a sample size that we are typically using for such trials.

And then we'll move on to the alternative that I've been asked to discuss at more length, which is the non-inferiority trials designs, the samples sizes for trials that have bone densities as an endpoint and vertebral fracture outcomes, or vertebral deformity outcomes as I'll call them as an endpoint. And, we'll propose a principle for how to set non-inferiority margins, which are such a critical assumption in that.

And then finally, I'd like to make a couple of comments and a proposal regarding the duration of trials and why consideration of the duration of trials for fracture prevention may be an important issue for this panel to consider in reframing guidelines.

There have been a number of discussions about whether or not placebo-controlled trials remain an alternative for the testing of new drugs for the prevention of fractures. It really in many ways boils down to both the risk and your perception of the risk of being in a placebo group instead of receiving standard therapy.

And when we talk about osteoporosis patients, it's very important to realize that we're not dealing with just osteoporosis as a single group. It's a very heterogeneous group of people with heterogeneous risks. I would like to point out how we might in designing trials begin to draw finer distinctions between people that may allow us to do placebo-controlled trials in a more ethical fashion.

For example, a woman who has a vertebral fracture has about a four-fold greater risk of suffering another fracture, vertebral fracture and other fractures than a woman with just low bone density or osteoporosis defined by her densitometry measurement at the hip or the spine.

It's a very important principle, but we've

also been able now to distinguish even finer gradations of risk among women who have vertebral fractures. And I'll point out that it makes a difference whether it's recent or whether it's an indeterminate age. It makes a difference whether there are multiple.

And although there is less data on this last point, and perhaps some of my colleagues can help me about some of the unpublished data, it makes a difference about how severe these fractures are in terms of the risk to the individual patient. So, that database is just starting to develop and I think it will be very important for designing trials.

The annual risk of having a new vertebral deformity on which we base our sample size estimates in trials -- and I'll return to in discussing alternative designs -- are in this order for new vertebral deformities. That means new radiologic events rather than clinically apparent painful events. Defined as Ken Faulkner described earlier, those with a hip density that's in the osteoporotic range have about a one to two -- in one trial up to about three

percent per year risk of having those events, depending in part how they're defined.

Patients who have a vertebral fracture in an undifferentiated way have between a five and ten percent risk of suffering a vertebral fracture per year in the existing trials that have been done to date. Now because these drugs have generally reduced the risk by about 35 to 50 percent, then you can go through those and find out what the comparative risks are for someone in the placebo group compared to someone in the treatment group.

I'll return to the limited activity days at the end. But, the risk of having a spine fracture during the course of a trial on an annual basis, again a radiographic vertebral fracture depends on whether you have a spine fracture or you just have osteoporosis according to the densitometry machine. So if it's just a densitometric osteoporosis, your risk per year of suffering a vertebral fracture is on the order of one percent. And if you start off with a spine fracture, it's on the order of two to three percent per year.

I'll return to limited activity days later. However, as I was pointing out earlier, not all fractures are equal. Bob Lindsay pointed out a couple of years ago in a very nice article that defining fractures somewhat more liberally so you get a somewhat higher incidence -- in the VERT trial, there's a 15 percent reduction in risk. He pointed out that a woman who'd had a fracture in the last year had about a 20 percent fracture risk in the following year.

We've gone back to two other databases that we have, the FIT trial and the MORE trial, and have confirmed that these women have about a four to five-fold greater risk of a subsequent fracture than women whose fracture is old or of an indeterminate age.

That means to me, and roughly estimating this, that the women who have a recent fracture have about a ten percent per year. Just a rough estimate, about a five to ten percent per year are at risk of suffering another radiographic event. About a third of those will be clinically evident, diagnosed as

clinical vertebral fractures.

These limited activity days on the right point out the number of days that we estimate from the FIT trial that an individual is disabled as a consequence of having a fracture or is a result of back pain.

And we've estimated that a woman who has a spine fracture, for example -- not differentiating the recently acute or multiple ones from the indeterminate ones -- those who have a spine fracture on average have about seven days of limited activity per year. In the placebo group, it would've been prevented by treatment. That's about seven days.

But for recent fractures, this probably amounts to on the order of weeks, two, three, four weeks that would be preventable by taking standard treatment instead of being in the placebo group of a trial.

One of the other issues that comes up is the risk of death. I have in many epidemiologic talks, as all my colleagues have, have said over and over that hip fractures are associated with 12 to 20

percent risk of dying in the first year. And those who have vertebral fractures have an increased risk of mortality from trials.

It turns out, of course, that those are epidemiologic associations. The reason, by close chart review or other kinds of methods, the reason that most people die after hip fractures is because they have other diseases like cancer that lead to the fracture and then cause the death. And when you get down to actually estimating how many deaths are attributable to the fracture, it's a much smaller number but very hard to figure out.

So, I've gone back to the trial databases to try to figure out whether prevention of fractures prevents death. I think that's an important thing to know.

In the fracture intervention trial, we went back to all the 6,459 women who were in that trial and suffered. Nine hundred and seven women suffered 1149 fractures, and there were 122 deaths. And, we couldn't find in the database a single death due to the fracture.

Now these are healthy women. They don't have other comorbid conditions because they'd been screened out for such as you would normally do in a trial. But I think that's an interesting statistic.

We've also then gone back to all of the trials and pooled all the mortality rates in the existing trials that you've seen, the major pivotal trials, to see if we could find evidence that reducing fractures in some way reduced overall mortality.

Again, this is not fracture related mortality because those are not reported in the papers. But in no single trial was there a reduction in risk of mortality that was statistically significant in the dozens of thousands included in our overall poolings. Again, it fails to find a statistically significant reduction of risk of mortality due to participating in the placebo group of a placebo-controlled fracture trial.

One other area is quality of life. We have relatively insensitive methods of measuring that, just questionnaires given once every six to twelve months in a couple, not all of the trials. We're

fortunate that Merck has allowed us to have access to the database from the FIT trial to begin to look at some of these issues.

One of the things I did was take a look at the SF-12 quality of life instrument that was measured at baseline and then at the end. It has six subscales of functional status. What we found is that the change in functional status, that the pain in other domains in fact did not differ significantly between the placebo and the alendronate group on any measure from beginning to end of trial.

So it's not to say that osteoporosis is not an important condition, but that, in terms of global changes and quality of life for an individual participating in the placebo group, there doesn't appear to be a substantial risk of deterioration in overall quality measured this way.

We've measured it in other more sensitive ways, and that is to count the number of days, again, of limited activity or back pain that sent you to bed. Amongst women with vertebral fracture from the FIT trial, we've estimated that again you can see an

aggregate here that they spent on average an extra day in bed compared to those in the alendronate group. And, they have additional days of disability that is limited activity that add up in total to about a week of limited activity or bed rest as a consequence if they're in the vertebral fracture group.

But remember, not all vertebral fractures are equal. This gamushes together all of those with recent and multiple and single indeterminate aged fractures. I haven't been able to distinguish those amongst who have just osteoporosis by density of the hip. We have a reduction of days in disability due to a fracture, but we haven't been able to find statistically significant reduction in other measures of days of limitation.

Let me just summarize. The risks of being in a placebo group in past trials of effective drugs, on average have produced limitations of activity due to the fracture in the bisphosphonate trial, the alendronate trial -- I don't know the data from the Risedronate trials -- has not measurably reduced quality of life as measured by questionnaires and

doesn't significantly increase mortality. Most importantly, however, even the risk of disability depends very much on the degree of severity of osteoporosis.

So, placebo-controlled trials are becoming much more difficult to do. And many of us involved in those know that it's hard to recruit people from the United States because so many of them want to be on alternative drugs, and their doctors resist, and IRBs are difficult at times.

I want to make one other point. The placebo-controlled trials and women with osteoporosis have sample sizes. The ones that are out in the field now that I know about had sample sizes between 2,000 and 8,000. Towards the end, the 2,000 end, if the only endpoint is vertebral deformities in very high-risk women. Towards the right end, more towards 8,000, if they have more ambitious goals such as reducing risk of fracture in the first year or eventually trying to find a risk of reduction of hip fracture.

So with that as background of the current

state of placebo-controlled trials, let me move on to non-inferiority trials. We're testing new chemical entities versus actual comparators, and I'm going to divide this into two parts. As we go through, I'll focus on bone density and then outcomes of vertebral deformity.

I'm not going to talk about hip fracture endpoints because the numbers will get really big. But BMD and vertebral deformity will be the two outcomes I cover. And, I'm going to address bisphosphonates for which we at least have adequate data to do some of this.

The way I develop some of the assumptions here were to avoid looking at anybody else's sample sizes and published articles first, but to try to develop all the assumptions from discussions with a number of investigators at ASBMR at recent bone meetings. Then, I had statistical models provided to me by Dr. McCulloch. And then once I had the assumptions in place, I went ahead and calculated the sample sizes so that I wasn't iterating back and forth to try to make this look good or bad for you. So, I

hope that this is an unbiased estimate.

The assumptions that underlie all the models that I have given you, we're going to use confidence limits. One sided, essentially, we're going to just look at the bottom part of the confidence limit and we'll use an alpha of 0.025. If you wanted to use an alpha 0.05 for that lower confidence limit, I'd be more liberal. The sample sizes would go down by approximately 20 to 25 percent, as noted by an earlier speaker.

I'll show a couple of examples of 0.9 power, but in general, we'll stick with 0.8 power, or 80 percent power. I'm assuming that the bone density trials will last two years throughout, so I won't repeat that figure, and the fracture trials will last three. Towards the end, when we begin to talk about duration of studies, I'll also look at a one-year fracture outcome trial with a comparator.

I did not inflate any of the numbers here to take count of loss to follow-up since that will vary a lot from drug to drug and how it's done. But in general, our experience is that there should be

about a 20 to 25 percent inflation of the numbers I give you to account for loss to follow-up the way most trials are done.

The key specification in all of these models is the non-inferiority margin. I have to say what this meant was still a little fuzzy to me. I hope it's not still fuzzy to me as I'm presenting this. But I would sympathize if some of you are still a little unclear about what the non-inferiority margin is. So, I was asked to try to define that a bit.

It's really kind of an accrued streetwise fashion of how much inferiority are you going to allow in a new drug and still let them get away with approving it. And so, it's a margin of difference below the existing drug that you allow in order to say that the result is comparable, sufficiently comparable. You say that you have confidence it's not inferior.

I'll just use an example to illustrate this. If this is an old drug "A" that's been around in the markets, it's proved. We know that it has approximately a four and a half percent improvement in

bone density over the course of two years. This is what that would look like.

And we may decide then that to set a non-inferiority margin, that's like a confidence limit that extends down. You'll see it also extends up, but I'm just going to ignore the top part because I don't care about superiority for this. I'm interested in non-inferiority so I'm looking at the bottom part of that confidence limit in order to establish the margin by which this other drug has to perform.

So in a sense, the way I've set this up that non-inferiority means that there is less than a two to five percent change that the effect of "B" will be at least one percent worse than the effect of "A". Here's an example of where it worked. "A" and "B" have about the same mean effect on bone density. "B" may be a little bit worse, but it's still within the margin. You'll call this non-inferior.

And here's a situation essentially where it doesn't work. The difference between "A" and "B" exceeds the non-inferiority margin. You reject the non-inferiority assumption. Again, for those of us

who think in other common terms, it's probably inferior by the criteria that were established.

So now, some of the assumptions that go into the bone density calculations come from the trials that we've done and that others have done in looking at the literature and from a consensus when there's been a difference of opinion, we're looking at two years for bone density. So over the course of two years, we're assuming that alendronate, which is the example I'll use first, has an improvement in spine bone density of roughly 4.5 percent over the course of two years compared to the placebo. And at the hip, it's approximately 3.0 percent at the end of two years. The numbers are bigger at the end of the trials that go on to three to four years.

We going to make the assumption in all of these that the new drug you're bringing to market is estimated to have the same effect. It's changes our calculations to assume that it's better or assume that it's worse to start with. We could do that, but I'm not going to. If you want to, we'll do it in private.

The new drug has essentially the same

estimated effect. That's why you're bringing it to market. You're not trying to bring something that's inferior. And we're also making the assumption that the non-inferiority margin is one percent lower.

Again, this comes from a consensus or sort of a median of people that I've spoke to about this. That's roughly about a 20 percent difference between the placebo effect and the mean effect. But I test a range in this from 0.5 percent margin to 2 percent.

A very important assumption in all of this, which is surprising perhaps to some of you, but is the standard deviation over two years of change in bone density measurement in the whole population. Small changes in that actually make a big change in the sample size.

So, what I've done for this is I've used real data. There are lots of data points, and Dennis Black and I have tried to find what looks like the modal or the median value of the various groups that we looked at.

So for the standard deviations of change in bone density, I've used five percent for the spine,

which is a pretty good estimate for what we saw in the fracture trial treatment groups. Since these are comparative trials, everybody has got treatment. So, we're using just the treatment groups for those estimates. And, a total hip of about four percent.

Now some of you may say that's surprising. Spine should be better than total hip. In the fracture intervention trial database, this is what we found. This is what's in that database for a standard deviation of change, so that's what I've used.

I don't think it makes qualitative differences in the results. But for those specifications of non-inferiority at 0.8, it looks as though you need hundreds of course. And if you want a very, very narrow margin, that actually is not 0.01, that's 0.005 due to a glitch in PowerPoint. But 2002, you want very stringent non-inferiority margins. But for the ones we assumed were reasonable, it's on the order of hundreds. Five hundred and six is the exact number for the total number in the trial that we specified.

To use total hip BMD, it actually was

somewhat greater because of the smaller change in total hip BMD. If you were using that as an endpoint, the numbers are somewhat larger. And the one that fit our specified assumptions is about 788-person trial. You can see that if you're more liberal with the non-inferiority margin, you can get down to numbers that are in the 100 to 200 range.

Yes?

PARTICIPANT: (Speaking from unmic'ed location).

DR. CUMMINGS: Yes. I'm sorry. The PowerPoint, this actually is 0.03. I don't why, but I just couldn't get my PowerPoint program to read out the right numbers. So, 0.025 and that's 0.005. In other words, that's a half of a percent non-inferiority margin. I apologize that I'm clumsy with PowerPoint. But, it's fixed on later graphs.

If you use a power of 0.9, it doesn't really make a whole lot of difference to the magnitude of the trial you're planning. Again, it's in the hundreds of patients that you need for non-inferiority for BMD.

Okay, now vertebral deformities is a more challenging task. Vertebral deformities -- again, I want to make the assumptions clear. For me, this was a change in measurement or semi-quantitative grade on lateral spine films.

For the morphometry part, we defined this as a 20 percent decrease in any vertebral height. And that's been the primary outcome of trials that we've done, and then confirmed by semi-quantitative gradings or radiologic readings. We used this criteria for defining the rates of fracture primarily.

Some trials have found much higher rates like the VERT trial, but they've used more liberal criteria like 15 percent. So therefore, their rates of fractures tend to look higher. But the numbers I've used might be slightly smaller than you're used to because I tried to standardize it around 20 percent.

Now the assumptions used are that in these non-inferiority vertebral deformity trials, we'll be using a high-risk group. And those are women with a vertebral fracture, probably of indeterminate age.

It's been our experience that we're tending to get somewhat lesser of lower risk patients from the United States than we used to.

The placebo rate that we started with, that Dennis and I came to for this presentation, was a placebo rate of about five percent per year for vertebral fractures. It can be done for higher as you'll see later. That means that there is a 15 percent cumulative incidence of having a vertebral fracture at the end of three years under these assumptions. That makes the numbers turn out to also look neat and easy to remember.

Now I'm going to use for the active comparator the approved drug alendronate. And I going to assume, because the numbers worked out exactly and it's very close, that that drug reduces the risk of vertebral deformities by 47 percent over the course of the three-year trials. That's almost exactly what it is.

That means that if you were to do a trial with an Alendronate group over three years, their rate of fractures, if we had the same kind of placebo group

as before, would be eight percent. An eight percent cumulative rate over three years is a 47 percent reduction.

So, this is the number to which we are going to be comparing the new drugs, that eight percent cumulative rate over three years. That's the estimated effectiveness in the right-hand column. That's what we're assuming that eight percent means.

One of the problems with doing active comparator trials is you're really not sure of that. That estimated effectiveness could be 55 or it could be 60 or it could be 35 percent. But, that's one of the assumptions you have to use going into this. And this is the best estimate I've got now.

If you then accept rates for the comparison of nine percent, ten percent, eleven percent, twelve percent margins that go up by you see a percent each, the non-inferiority margin for nine percent would be one percent essentially. If you're going to accept those, then the estimated effectiveness associated with each one of those allowable rates of fractures in the new drug group are

on the right-hand column.

I have trouble using figures like 20 percent reduction in the non-inferiority margin in order to understand the value of a comparator trial. It makes a great deal more sense to me as a clinical investigator and clinician to be concerned about the effectiveness of the drug I'm going to be approving.

This represents about a two-thirds preservation of effect. This is about half of the benefit or half of the effect of the comparator drug. I think we chose from the survey of people I did, there seemed to be a consensus that about a 33 to 35 percent reduction in risk was a clinically important reduction.

And if it were below that, people began to have concerns about whether or not it's worth using such a drug in practice and whether or not it's sufficiently different from placebo to be acceptable for approval. So, that's the 33 percent or the 10 percent three-year rate is what I will use as the principle assumption, the number that I'll highlight.

I think that a 20 percent estimated

effectiveness or a 12 percent three-year rate in the comparator trial is just -- personally, I don't believe that this is an acceptable degree of efficacy, sufficiently different from placebo to make it worthwhile, even estimating numbers about that alternative. So, the range of numbers I've used is from eight to eleven percent.

Those are assumptions I've already given you. The approved bisphosphonate has a 47 percent reduction risk. The new drug has the same effect, and we set that non-inferiority margin at a two percent different. In other words, we'll accept it if, under this scenario, it reduces the risk of fractures by about a 33 percent.

And, I've tested the range I told you about that includes 11 and 9 percent, a 40 percent reduction, 27 percent estimated reduction, and these are the sample sizes. So if the non-inferiority margin is as we guessed is about two percent, under these sample sizes, it's about a 6,000 person trial not accounting for dropouts under a power of 0.8.

Now if you're using a more strict, a one

percent margin that goes way up -- and you can see the effect of a non-inferiority margin is huge on these assumption, so we'll return to how you go about setting those. The panel will have copies of those slides.

Now if you raise the power to 0.9 instead of 0.8, it has a modest effect on the sample size estimates that we came up with. Again, about a 7,000 to 8,000 person non-inferiority trial with a two percent margin. If you're more liberal, you get down to about a 3,000 person trial.

It's possible, I mean I've been a consultant at a couple of meetings where people say we want to compare this to another bisphosphonate that is presumed to have lesser effects or weaker effects, sort of lowering the bar.

And so, let's assume that an alternative, in this case risedronate -- it's best unable to tell from the VERT trial -- has about a 40 percent reduction of risk of fractures in the populations in which it's been tested. It's hard to tell whether it's as effective or less effective than alendronate

because they're in different populations. But, there's about a 40 percent risk.

So if you use that, maybe then if you get a two percent inferiority margin and you get all the way down to 20 percent, you can actually get a drug that's weaker on to market by choosing as it were a lower target. Now that's one alternative we'll talk about briefly.

But I think that's where this analysis helps. I really don't want to go below this line. In fact, I'd say that if you're going to use a drug that has weaker effects, you set a narrower inferiority margin in order to test the non-inferiority of the drug.

If that's the comparator, and this is two-thirds of the effect, although that's a half of effect, under this circumstance, I will draw the line in exactly the same place about the estimated effectiveness. I'll look to test that ten percent. In other words, a one percent margin for this particular drug instead of a two percent for the other. And not test the drug that's estimated to have

-- I wouldn't be interested in just a 20 percent reduction in risk.

The sample size for a non-inferiority under these circumstances would be at 23,100. That would be the preferred sample size. In a sense, you're penalized by choosing the lower bar. But if you went ahead with the two percent non-inferiority margin, you're in the 6,000-person trial arena.

So if the issue is aiming low, if a fixed margin is allowed, you know two percent regardless of which drug you're choosing as a comparator, a 20 percent reduction, then choosing the weakest comparator will tend to produce an easier, a smaller sample size, which may lead to approval of drugs that are less and less and less, eventually as you go through time, distinguishable from placebo.

I would argue that the basis for choosing non-inferiority margins should be the estimated effectiveness of the new drug as a principle. So the base margins on the minimal estimated effectiveness that you're aiming at, and amongst colleagues, that seemed to be somewhere around 30 percent. Again, a

very informal convenient sample of friends has a lower limit.

Now SERMS. SERMS are interesting in this kind of analysis. I'm a little uncomfortable doing this for SERMS because they have effects on multiple organs and conditions. Non-inferiority tests just for bones to allow a new SERM on the market is problematic to me. I mean I'm a bonehead. Despite that, I'm not sure that the bone is the most important organ in the body. My colleagues I hope will forgive me for that.

We've only had one in the class that's been tested for fracture effects, Raloxifene. So, there's not a great deal of data here, which to make these comparisons. This approval on the basis of non-inferiority on bone alone to other SERM agents seems to me to be premature. Having said that, I will nonetheless go through with my assigned task of giving the assumptions and the results of this.

The assumptions are essentially here that the approved Raloxifene effect is about a two percent change in spine bone density. The new drug has the same effect. And we set the non-inferiority margin

here because it's such a small difference from placebo. Setting it at one percent is reducing that potential benefit by half. It doesn't make much sense, so we tested it at a 0.5 percent lower bone density effect as the non-inferiority.

It just turns out in the MORE database that the standard deviation of two-year change in the spine bone density looks better than in the FIT trial. Perhaps, because the effect of the drug is smaller, you get a lower range of changes within the population. It's 3.5 percent for the standard deviation change in our database.

Anyway, the sample size for non-inferiority spine bone density, in part because of the better standard deviation, is modest. At 0.5 percent, it's around 800.

And for vertebral deformities, again I'm going to assume for convenience that it reduces fracture risk by about 40 percent. Therefore, in a sense the analysis we did for Risedronate really applies quite directly, and I can just skip to that. Depending on the inferiority margin you'll use, it's

between 6,000 to 23,000. It's a very critical assumption.

Again, because the fracture effects are perhaps more modest with the SERM class than with the most powerful bisphosphonates, you would tend to choose somewhat larger sample sizes.

DR. COLMAN: Steve?

DR. CUMMINGS: Yes?

DR. COLMAN: For those sample sizes, those are total or --

DR. CUMMINGS: These are all total, not per group. These are total, not per group.

DR. COLMAN: Okay.

DR. CUMMINGS: Now, these are a lot of numbers. And when you come down to actually calculating this, there will be different assumptions. People will come to the table, manufacturers will come with different assumptions that will make different numbers.

So, I think that the summary really is that for vertebral fracture comparator trials, the number that you need is in the thousands. And for

bone density comparator trials, the numbers will be in the hundreds. The non-inferiority margin that you accept makes a huge difference. Probably next on the list would be the standard deviation of change over time for bone density studies.

Now, what non-inferiority margin makes sense? I've suggested something for vertebral fractures as a minimum, sort of a bottom floor. But I just wanted to offer an opinion that in a sense, using non-inferiority margins of the sort that I talked to you about that reduced or allowed 20 to 30 to 40 percent inferiority makes some sense to me if the new drug has other benefits. You can accept something that's inferior if it's got other health benefits, it's safer, it's more convenient, and therefore better adherence and it's less expensive.

I know that this is not usually done in setting non-inferiority margins, but it seems to me that there would be a compelling pace for being a little bit more liberal for something that has other benefits. For new drugs with no other advantages, they should either prove superiority or the margins of

non-inferiority should be very narrow, probably narrower than I've rehearsed for you earlier in the talk.

PARTICIPANT: What is standard? DR. CUMMINGS: What is what?

PARTICIPANT: What is standard?

DR. CUMMINGS: No, I'm sorry. What I meant is the ones I just used. Standard came about because Eric suggested some percentages that we used that were within the range that I talked about. In other words, 20 to 30 percent differences between the effect of the drug and the placebo group.

And so, I think that those are acceptable if there is something new being brought to the table. But if not, then I would propose that the criteria be more stringent.

Now let's go to duration of trials, and this will be the last. We'll start with one-year duration trials. I think that there is some sense nowadays for some compounds and maybe even for bisphosphonates, where they worked so dramatically in the first year, to consider a shorter duration of

trials for initial registration of drugs.

Let's first look at the one-year study. One year, a 60 percent reduction in risk of vertebral fracture can actually be seen in relatively small trials compared to placebo. So, I'll just focus here on the non-inferiority margin for a new drug and assume that the estimated reduction that we're interested in would be no worse than a 40 percent reduction in one year.

This happened in the first year, but I'll show you that in later years, the effect is not quite that strong. And so if you're going to do it on the basis of just the first year, then I don't think you allow as low as a 20 percent reduction in risk, a big range. So, I'm going to assume a 20 percent non-inferiority margin, which means a 40 percent effectiveness, and test a range from 10 to 30 percent with the assumptions you've seen before.

Two-thirds effect is a 40 percent reduction. Half of that effect is a 30 percent reduction. These are the non-inferiority margins that Chuck calculated. I'm sorry, the sample sizes for

those non-inferiority margins. In other words, for that first year comparison of a drug that reduces risk by 60 percent with another new drug assumed to have the same effect, but you're willing to accept lower. That 6,154 sample size would be my preferred number.

What if the risks were higher? In a 12 to 18 month study, you've managed to get a very, very high-risk group. If we do the same thing with much higher risk groups, double the rate of events, there is a modest effect. It reduces the preferred sample size by about half to double the incident rate. So now, that's the year one-year effect.

I'd like to now talk a bit about longer durations and why there might be a rationale for being concerned about longer duration studies. It's been pretty clear, and certainly if you've been at any meetings or read any medical journals, very well advertised that all drugs, antiresorptive drugs, dramatically reduce the risk of fractures in the first year. You've heard 65, 68, 60 percent -- I saw an article recently claiming a 70 percent reduction in risk of fractures in just the first year. Again, I

think that's true and there are biological reasons why that might be happening.

There is concern among some members of the osteoporosis community about the long-term effects of antiresorptive drugs. Right now, we can't do much with that concern because the placebo-controlled trials have lasted only three to four years.

There have also been long-term trials, particularly with the alendronate, with the long-term extensions going out to seven to ten years from which we've had to try to draw very indirect inferences about how long the drug continues to work. And there is some reason, I think, to be concerned that we should go a little bit longer than three to four years, coming from the data that we've seen so far.

Risedronate, for example, we know continues to improve bone density. It dramatically reduces risk of vertebral fractures in the first year. In the second to third year, we really don't know what it does separately because the sponsor hasn't released that data from the VERT trials to allow us to do those kinds of calculations.

But fortunately the sponsor for Alendronate has allowed us use of the databases and we are able to show that a similar phenomenon of dramatic improvement in the reduction of risk in the first is seen with alendronate, and there is a 40 percent reduction in the third year. It's a little bit more difficult to figure this out, exactly what that means because there was also a dosage change from five milligrams to ten milligrams between those time points.

This change or this apparent waning of effect could be due to something called depletion of the susceptibles. That is, if there is susceptible people in the placebo group that all fracture in the first year and they're gone. Then there are fewer of them around in the next year to have fractures and the drug won't look as good.

So, we've done some extensive modeling about this statistical artifact. Models with very extreme assumptions that basically assume that all the fractures happen in the susceptibles in the first year, and high rates of susceptibles with extreme

assumptions really can't account for the observed declines in effects that we've seen with these drugs. In particular, because such an effect, a statistical artifact would require that you also see substantial declines in the rate of fractures in the placebo group because the people who are going to have them are gone.

And, you don't see that in the existing trials. You see pretty constant rates in the placebo group. Although it's been said that this is statistical artifact, it is not. So that means that it is a biological effect and there are two possibilities.

I think the most likely is just the dramatic first year effect on bone resorption that gives you, that preserves architecture. Then that, on top of a long-term sustained effect, results in a 30 to 40 percent long-term reduction in risk because of the improvement in bone density.

But I can't tell on the basis of the current data we have available, I can't tell that from the second alternative, which is that inhibition to

bone resorption initially strengthens, and then after five, six, seven years longer than the existing data from trials or from other types of studies, longer than it's gone. So that remains a concern in some quarters of the community. So I would like to suggest something.

I would like to suggest that you might consider approving drugs for use even as early as one year or 18 months, fracture data, but then approve the drugs for use equal to the duration of the trials that you are provided as evidence. And that to get an extension in the label of how long patients be allowed to use that would be contingent on providing data that the drug remains safe and that the fracture risk is durable, that the reduction of fracture risk is durable.

Now there's a challenge with this of course, and that's that it's not feasible to continue placebo-controlled trials beyond about three years. It's just too difficult because of the environment about placebo-controlled trials.

There are a couple of alternatives to

this, however, that I think are a reasonable comprise. And one is to continue your placebo-controlled trial for three years and then stop the placebo group if it works and continue the treatment for longer. Companies are doing this, but they're often doing it without adequately powering the study or planning it in this fashion in advance.

It looks like this. There's the rate of the placebo group for the first three years, and here is the year-by-year rate in the treatment group as it continues out to ten. It looks like it's continuing to work, and that's great.

Again, this is just one possibility. There are other variations on this, but this is the simplest one. You could compare slopes of lines or just year-by-year effects. For the purposes of just this one example, I'm going to suggest comparison of the rates in the treatment group at seven to ten years with the rate of the placebo group for the first three, then needing to adjust for the advancing age of the patients who are going out ten years.

And that's the comparison one. It would

draw then the placebo group to the rate in the treatment group carried out longer. In this particular case, it's a statistically significant difference. But it's possible that with time we would see a loss of effectiveness, a loss of durability and that it would no longer differ from placebo. It might even cross this line and would lead to the conclusion that you should stop after three or four years rather than continue it 20, 30, or 40, or lifelong.

These kinds of trials are feasible. I won't go through the details of the sample size estimates, but we've done several such estimates. And in general, trials that had more than 1,000 per group at the baseline with less than 20 percent loss to follow-up during the placebo period and are able to retain at least 50 percent of participants out ten years will have over an 80 power to confirm a 30 percent lower risk in the last three years versus the placebo rates.

There is another alternative. I think Merck has done a very innovative thing with the FIT trials in taking the treatment group after four years

and then re-randomizing them to continuing out ten, or stopping the drug to test whether or not there is benefit from continuing as opposed to stopping. I like that design very much if it's adequately powered.

With that, let me summarize by saying that placebo-controlled trials with women who have densitometric osteoporosis entail low risk to the participants and are feasible. And I think that that might also include women who have a single vertebral fracture of indeterminate age. Non-inferiority trials on bone density require hundreds, and thousands are needed for vertebral deformity comparator trials on non-inferiority, but that depends very heavily on the non-inferiority margin.

I would suggest or I would like the panel to consider at least initial registration of drugs for the duration of the evidence that you're presented with, and extend that duration with subsequent demonstration, that there is durability of effectiveness as well as safety. And, consider setting the non-inferiority margins, such as an important determinant based on some judgment about

advantages, if any, of the new agent. But at the least when you're dealing with comparator trials for vertebral fractures, set those margins that preserve a minimal estimated effectiveness of treatment over placebo.

With that, with those modest suggestions, I'll stop and say thank you.

(APPLAUSE.)DR. CUMMINGS: Henry, am I allowed to take questions?

CHAIRMAN BRAUNSTEIN: Dr. Bone?

DR. BONE: Steve, thank you for a nice review. I had two specific questions, and you may even want to stay up there with your computer to answer these two questions.

DR. CUMMINGS: I don't think I'll be able to re-project, but go ahead.

DR. BONE: Okay. These had to do with some estimates that you didn't mention. One is, if you use the figure 47 percent, which is just about exactly the reported relatively risk reduction for alendronate in the trials you're referring to, what is the confidence interval around that estimate for the

relative risk?

DR. CUMMINGS: For alendronate?

DR. BONE: Yes.

DR. CUMMINGS: Although I'm an author on those trials, I don't remember the confidence limit. It was relatively narrow, particularly when the two trials were pooled, those FIT-1 and FIT-2 for patients with osteoporosis.

Someone else might be able to help me with the confidence limit, but I think that it goes down no further than the high 30 percents. About 37, 38 percent is the lower limit of that confidence limit.

DR. BONE: And the other question was since it's important to know whether drugs actually reduce the risk of hip fracture -- and we've had examples of drugs which did resist the risk of vertebral fracture with no relative risk reduction at all, even not a significant one. Just no change in the risk for hip fracture. What's the sample size calculation for doing the active controlled trial for a hip fracture endpoint?

DR. CUMMINGS: That's also been done by

John Kanis, and I'd refer to those. I did it one pass at it, and it's in the 20 to 50 margin.

DR. BONE: With what power?

DR. CUMMINGS: I think we used a 20 percent difference from a 50 percent reduction. Just because it's such a rare event, it's impossible to do a comparator trial for hip fractures.

CHAIRMAN BRAUNSTEIN: Dr. Watts.

DR. WATTS: You chose a 40 percent reduction in fracture for your example with Risedronate. In the two vertebral fracture trials, one showed a 41 percent reduction and the other showed a 49 percent reduction.

DR. CUMMINGS: Yes.

DR. WATTS: And that raises a dilemma as to which of those numbers you would choose if you were powering a trial. It might've been cleaner had you chosen those two numbers rather than two different agents.

DR. CUMMINGS: Yes, that's true. I could've done it the other way. What I was trying to do was not pin 40 percent just on Risedronate. But

say, what if you had a drug, which you thought was less effective, and therefore you chose as a company to choose that as a lower mark to hit with your comparator.

DR. WATTS: Another trial design that you didn't mention was superiority trials.

DR. CUMMINGS: Yes, I wasn't asked to do that.

DR. WATTS: It's possible that a new agent might come out that looks like it's a lot better. And by my calculations, it takes a far smaller sample size to show superiority.

DR. CUMMINGS: Well if it's really superior, it takes a different, a larger sample size. But if it's not, if it's on the same, then that changes your sample size. Remember, throughout this I assumed that the new agent that was coming on, in fact, had the same reduction in fracture risk and the same change in bone density.

And if you change that assumption, I can show you what it does. If your drug is better and you use a non-inferiority margin, then it's an easy time.

You can do it with just a few hundred patients.

CHAIRMAN BRAUNSTEIN: Yes, Dr. Temple.

DR. TEMPLE: Choosing the margin for non-inferiority trials is the subject of infinite quantities of discussion because there are a bunch of cardiovascular diseases where no one would debate the possibility of whether you can still do placebo. So, this becomes a very important issue.

It's very important in doing that to distinguish between a non-inferiority margin whose purpose is to show that your new drug has some effect compared to placebo, any, and one in which you're designing it so that you show you preserved some fraction of it.

It should be obvious, but if the situation is such that it's unethical to use placebos anymore, it's obviously important to preserve a fair fraction of the effect of the control agent, otherwise what's the point? But, those two things need to be kept in mind.

If, for example, you thought a reliable meta-analysis of the effect of some positive control

agent was a 40 percent reduction, you could document superiority of the drug to placebo by showing that the difference between your new drug and the controlled drug is not more than 40 percent. Then, it would be better than placebo, which is approximately equal to what you do when you discover that something is significant 0.05.

If you don't like that well enough, if you have a mixed feeling that you want to preserve some fraction of it, then you have to do what Steve was doing, preserve 50 percent of it or something like that. Of course, the implications for sample size are spectacular.

One problem that we encounter repeatedly is where you only have one trial of something. For example, in most lipid settings, there's only one trial and it a particular setting because no one will let you do another trial once a benefit has been established. So how on earth do you pick a non-inferiority margin based on a single trial?

Well, taking the mean doesn't seem good enough because half the time the effect is going to be

smaller than that so you wouldn't know what the effect in your new trial is. One thing that people have done is take the 95 percent lower bound, which if there's only one trial means the difference that you have to rule out is considerably smaller than the mean - sorry, than the point estimate of the effect, and trials get very large.

So in these situations, we've encouraged people to, either through pooling a lot of data or looking at the one drug with the most data, use that as the active control and then at least you have a number that you can rely on. You don't have to be entirely conservative.

But, we would never think that the mean effect of a single trial would be the right non-inferiority margin. It has to be, as somebody over there suggested, a 95 percent lower bound or something. Anyway, there's a great debate about exactly how to do that.

It does seem very important to distinguish between trying to show through a non-inferiority study that you're better than nothing, which might be good

enough in some cases in trying to show that you preserve some clinically meaningful effect of the drug, which then enters into major debates.

For what it's worth for thrombolytics, where the endpoint is death, CBER, the Biologics people accepted a non-inferiority margin that represented retention of half of the effect of the thrombolytics based on the 95 percent lower bound of a meta-analysis. That turns out to be a little more conservative than one might do. But that's one living, breathing illustration.

Of course, there the consequence of being wrong is death. Here, as people have been saying, it's a fracture. That's not as bad as death, but it might be bad anyway.

CHAIRMAN BRAUNSTEIN: Dr. Marcus.

DR. MARCUS: A box containing Steve's slides is going around. I'd like to go out beyond the box if you don't mind of a minute.

There are two possibilities that haven't been described in ways to approach some of these issues related to fracture, particularly hip fracture.

One has been proposed by Nelson Watts, and I'm going to ask that he discuss it since he is the father of that. That has to do with pooling of groups across various published trials.

The second one has to do with using as a control group, published data from ongoing, contemporary, very large-scale public health databases such as NHANES-3, where we have a very good indication of what true hip fracture rates are in this country. We can isolate the data by age, by ethnic group, by gender, and I just raised the possibility that one might be able to do a trial in which the control group could be valid public health data.

I'd like to hear the agency's response to that, if they would automatically exclude that or if they'd be willing to think about that approach. I'd like to hear Steve and some of the other epidemiologists discuss that.

As the second model, I yield to my distinguished colleague, Nelson Watts, to raise his idea.

DR. WATTS: What I've done is to look at

the fracture experience in a trail in which everyone received active treatment: daily dosing, weekly dosing that had equivalent effects on bone density and bone turnover markers, and extract from another large database of a placebo-controlled trial, subject to or matched to the entry criteria for the trial that lacked simultaneous controls. I matched for key characteristics, age, bone density, years since menopause, and the percentage of subjects with prevalent fractures.

In doing that, we not only had the historical control group, but a historical active treatment group as a way of internal validation. At least in that one, the rates of fracture in the historical treatment group were indistinguishable from the rates of fracture in the study that had no control group. The difference in fractures between the historical controls and the active treatment was statistically significant.

Though it should be possible, given the large trials, to create a huge database against which to judge -- extract a control group to judge

antifracture efficacy. That doesn't necessarily allow you to compare bone density or bone turnover markers or establish safety, but at least to get an antifracture efficacy.

CHAIRMAN BRAUNSTEIN: Dr. Cummings.

DR. CUMMINGS: The problem with doing that for hip fractures from databases is that there's a universal experience in trials and observational studies -- very marked in our studies and also in the WHI, that for reasons that we can't understand and that are not explained by bone density, age, or even estimates of health, there is a healthy volunteer effect on hip fractures.

So, their rates in the first few years of any study are in the order of one-tenth to no better than a quarter of the rates that you would expect from those patients matched by characteristic to databases. It could very well be that the hip fractures, of course, happen in people who are frail and don't come in to trials or to observational studies.

I don't know how to adjust for that healthy volunteer effect, so it's not a really

credible, tenable way to develop controls for a hip fracture study.

CHAIRMAN BRAUNSTEIN: Dr. Temple.

DR. TEMPLE: Well, the Women's Health Initiative gives you reason for caution. I mean why was it done? It was because every epidemiologic study ever done showed a 50 percent reduction in users of cardiovascular events. It turned out that those rates didn't really represent the truth for reasons that remain inexplicable.

It's almost surely true that people who enter trials are not the same as people picked up by NHANES or something like that. There are too many examples to enumerate, but there are numerous.

CHAIRMAN BRAUNSTEIN: Okay. We'll turn it over to Dr. Orloff, who is going to give the group their charge.

DR. ORLOFF: And charge it is. The first thing I can think of is ladies and gentlemen, start your engines because I think there is some discussion to ensue.

Let me thank everybody for a number of

very informative and careful presentations. Clearly, I don't hear any definitive answers yet. Let me start this by saying that we did not expect a consensus to come out of this meeting. I'm not hiding behind this, but I guess in part I'm saying this so that we can all leave here as friends as I think we entered the room.

We're here to frame the issues. This meeting is the first step we would hope in developing a guidance for industry for the development of drugs for U.S. marketing in the treatment of osteoporosis.

I just want to make sure that as we do deliberate -- and this again is aimed at trying to make sure we're all toned appropriately -- that guidance is just that. I think it's not universally understood that FDA guidance is guidance. It's not law and it's not regulation. It is, we hope, representative of the agency's best thinking. That's obviously with the input of our advisors and consultants on the subject issue.

So on the one hand, notwithstanding guidance, sponsors are at liberty to plot their own courses in drug development to meet the stated or

implied needs or concerns of the agency. Though to the extent that the guidance is not followed, we expect sponsors to justify the alternative approach. On the other hand, adherence to guidance does not guarantee approval or otherwise mandate a particular regulatory action by the agency.

Anyway, I'm going to ramble a little bit with some thoughts that I have before I get to a specific discussion of the questions that we'd like to have you address.

From sitting here, I think we got conflicting messages in the discussions on the acceptability of bone mineral density as a predictor of fracture benefit. I think it was also clear that we need to know more about the impact of animal findings in any given specific instance as confirmatory of potential salutory effects on bone for different mechanistic classes.

The meta-analyses that were presented in brief support BMD as a good predictor for amino bisphosphonates and parathyroid hormones, to the extent that there are data for that drug, with the

qualifier of course that animal data be positive or favorable even if the bone mineral density in those studies does not explain the whole fracture effect. I think we're all in agreement that that point was made.

It does sound to me as though BMD generally may be a reliable positive predictor of efficacy. It's just not a useful negative predictor. Thus, specifically it's probably not to be used to compare efficacy across different mechanistic classes for the purposes of placement in the armamentarium, which was a subject that's been raised. Whether or not it can be used for the purposes of approval, comparisons of BMD effects across classes, is perhaps another question.

I want to recognize that we did hear Dr. Marcus' comments on the need for additional histomorphometric characterization on the effects of new and existing drugs. Though the validity of some of these endpoints, as independent predictors of bone quality, remain to be demonstrated.

So, we well recognize the need perhaps to

understand more about the differences between one drug class and next. Although just saying that they have different effects on a marker, doesn't necessarily mean that they are different in terms of absolute clinical efficacy.

As we, you and we, that is to say go about thinking about the problems before us, I just offer that we need to distinguish between two important needs of the system, the healthcare system with regards to osteoporosis.

On the one hand, we need to come up with standards for approval as safe and effective. We need to be thinking about placement in the armamentarium. Again this theme comes up, the latter, the placement in the armamentarium is very important but not unfortunately the driving force for standards of approval. Perhaps it should be.

Also I want people to understand that we do not, or we are not, FDA, considering new drugs in this -- or for that matter, any other therapeutic area about which we know something -- we are not considering new drugs in a vacuum. For post-

menopausal osteoporosis, for example, we have tremendous or a great deal in the way of priors that allow us to put new data and development problems in perspective.

And so, for example, one might ask the simple and blunt question, which I do think we need to toss about a bit: Why wouldn't bone mineral density in an active controlled trial suffice for approval of a new bisphosphonate, or a new SERM, or estrogen or a new agonist at the PTH receptor?

It's also apparent that the ethical questions, that I guess at some level I'd hoped we wouldn't spend too much time haggling about today, have not been resolved. Two general approaches have been advanced around the table here and by others.

One is the idea of short-term placebo-controlled trials in high-risk subjects in which the fewest patients would be placed at risk for fracture for the shortest time. This is a burden-to-society argument, if you will, that does not address the irreversible morbidity and mortality standard down by

the Declaration of Helsinki.

The other approach is larger, long-term placebo-controlled trials in low-risk subjects with theoretical advantages of overall low frequency of events, and the low risk in a carefully chosen population of serious morbidity or mortality in a presumably generally healthy population.

I would venture that we need to be careful hearkening to Dr. Cummings' presentation with calculations of average morbidity experienced in the trials to date. For example, average days of pain or average days of work loss or average days of bed rest.

Since this measure of central tendency, if you will, masks the fact that clearly more patients had significant morbidity in association with placebo than drug, than there must be some patients in there on placebo who had serious or significant morbidity.

Finally, on the subject of trial designs, Dr. Cummings has left us with the conclusions that hundreds of patients would be required over several years for non-inferiority BMD trials and thousands for a fracture non-inferiority trial looking at

morphometric vertebral fractures.

All of these issues that I've skipped over bear further discussion. I'm not sure actually how much discussion we need to have here about placebos, but I leave it obviously to the Chair. Whether the issue of placebo versus active controls should be a question related to post approval broadening of claims, that is to say assuming that there is some consensus that bone mineral density might in many instances, studies may be sufficient for approval.

Anyway, the way we structured the discussion at least in our planning was to ask about the nature and extent of evidence from approval centered around four hypotheticals. You've seen these in your agenda. I said earlier, a new bisphosphonate, a new estrogen agonist bone, a new mechanistic class antiresorptive, and a new anabolic agent.

I just want to go through with you some of our thinking as we put together the structure of the questions that we'd like you to consider. For the establishment of efficacy, we asked you to focus on three main questions.

The first, when is bone mineral density an adequate primary endpoint, is really about how far clinical studies need to go to bone mineral density versus fractures in order to support approval. We understand full well that this decision or this judgment must take into consideration the specific results of preclinical studies. That's given. But clearly, also the confidence that should be placed in such studies. And that's why we asked for the presentations that were given to today.

So for example, going back to Dr. Rodan's presentation and Dr. Rozzoli's, such studies are all important for loose dating mechanisms of actions as well as such things as the potential for toxic mineralization effects of drugs or the existence of an apparent unfavorable relationship between BMD and bone strength indicative of poor bone quality.

I should say that as an important caveat or kind of a reverse catch-22, we also have to ask why would anyone in this day and age pursue a drug that demonstrated a poor efficacy or safety profile in animals.

The issue raised in discussion earlier, relative to extrapolation of animal studies and models of osteoporosis to other forms of bone disease associated with fracture risk and/or osteoporosis, I think is an important one. This presents a perplexing problem, particularly since the number of such patients, that is with other forms of disease associated with fracture risk will often be much smaller than those of patients with postmenopausal osteoporosis.

The reality is though, as pointed out by Drs. Bone and Rizzoli, that the other role for animal studies after pharmacology is bone toxicology. I suppose the judgment of whether some unique toxicity of bone could be anticipated, and these are others, must be made on a case-by-case basis.

With regard to the second question we then ask: If BMD is deemed sufficient, how long should trials be to establish durability of that effect -- and probably not separable, although it will be asked separately for assurance to say what duration of trial is necessary for assurance of bone and extraskeletal

safety. If fractures are deemed necessary in a specific instance, do the same questions apply?

And now with regard to the choice of placebo versus active control, the question must be addressed separately if BMD are fractures that are required. Notwithstanding the numbers involved, if BMD is the endpoint, we must ask what constraints exist against the use of placebo, whether they can be addressed by escape criteria for BMD or fracture on trial, what risk categories are appropriate for such trials -- as we said before, low-risk prevention versus treatment -- and of course, whether extrapolation of efficacy from prevention to treatment populations is possible.

We must also address advantages of active versus placebo with regard to safety assessment, or disadvantages, and with regard to again, placing the drug in the armamentarium. If fractures are required, what are the opportunities for add-on, what are the possibilities with regard to active controls, and are there indeed hypotheticals at least in the list that we've given or any that you can think of for which

placebo is really the only option in order to assess efficacy.

With regard to safety -- again in many instances, not separable -- I do have a question that I think has been of some confusion, at least it was Dr. Temple before I got to him, in number one, which asks about whether fractures can be used as a safety rather than as a efficacy endpoint. Although the answer may be simple, we thought we wanted to hear people's thought about it.

If a trial is examining BMD as the primary endpoint of efficacy, fracture, rather than being the measure of effectiveness, becomes a safety outcome. The question is: How should it be evaluated? Should it be evaluated based solely on ascertainment with regard to clinically apparent fractures or should there be active ascertainment as a way of monitoring patients in the trial? And, what issues does active ascertainment raise with regard to escape criteria on the one hand, but also on the other hand with regard to thoughts about essentially powering the trial for safety.

We also ask about other safety monitoring in the study, and this gets back to Dr. Grady's question at the beginning. I just want to make sure our position on this is well understood. We do not propose that the choice of efficacy endpoint necessarily impacts the scope or duration of the trials in order to assess safety. So, use of non-fracture endpoints does not necessarily imply shorter, smaller, or narrower scope trials.

With regard to other safety, we would say that those assessments of other safety issues is driven by the usual mechanisms of action, preclinical signals, early phase findings, plausibility of risks.

We also ask you about duration of trials for bone and extraskeletal safety that I mentioned earlier. And finally, what are the theoretical or real advantages and disadvantages to active versus placebo versus add-on trials for safety assessment?

With that, I hope I haven't confused matters. I'll turn it back over to Dr. Braunstein.

CHAIRMAN BRAUNSTEIN: Thank you. There's a lot of subquestions stuck in there.

I think what we'll do is ask the committee and the guests, of course, to consider and keep in mind the four hypothetical osteoporosis drugs, and we'll go through the questions even those there's obviously a lot of cross over from one question to the other.

We'll go through the questions sequentially and ask the members to ask questions of each other, to make their comments, and then when there's a lull, maybe what we'll do is go around the room and ask everybody to take their best stab at answering each of the questions.

So, we'll start off with the question about efficacy and when is bone mineral density an adequate primary endpoint. I would say in discussing this question, anybody who wants to indicate that they'd rather have fracture endpoints in place of bone mineral density for specific compounds in specific issues should mention that. So, let me open up that question to the group.

Dr. Watts?

DR. WATTS: There are three or four drugs

on the market for which a relationship has been shown for gains in bone density and reduction in fracture risk. That's Alendronate, Risedronate -- you choose that as a class or agent -- Raloxifene and maybe estrogen if you take the recent bounds health initiative.

There have also been some trials that seemed adequately powered to show an antifracture effect in which a bone density change was noted, and yet the antifracture effect was not seen. And I've already posed this question to Henry, so don't answer it please Henry. But, one of these trials was with intravenous Ibandronate, which produced about a five percent gain in bone density over three years, and did not show a reduction in fracture rates.

So, I'd be interested from those of you who know the trial or those of you on the FDA side to tell me if gain in bone density in a clinical trial were the endpoint for approval of these drugs, would intravenous Ibandronate meet that standard?

CHAIRMAN BRAUNSTEIN: Yes, Dr. Khosla.

DR. KHOSLA: I guess I just caught into

the Ibandronate because I remember seeing the data. And if it's the data that I've seen, the problem as I understood it was that the bone turnover markers were coming back up before the next dose of Ibandronate was given. That suggested that that particular trial didn't have a sustained reduction in bone turnover.

DR. WATTS: That was a post hoc analysis. And if you look at the marker data in the trial, the markers were suppressed. It's only by looking at marker data from other trials that it's possible to see they were much more suppressed within a week or two of the dose, and they headed back towards baseline. But, they were still 50 percent below baseline.

DR. KHOSLA: I guess the only comment I'm making is that with most other antiresorptives, there's been a sustained and consistent reduction in bone turnover. We've already heard about the importance of bone turnover as a potential additional factor that contributes to the antifracture efficacy.

I guess my only comment to this is that for classes such as "A" and "B", as I mentioned

earlier, if anything, bone density is perhaps a conservative estimate of the reduction in fracture risk. Provided that changes in bone turnover are consistent with what is otherwise seen with these classes, you could argue that it may not be an unreasonable surrogate.

CHAIRMAN BRAUNSTEIN: Dr. Sampson, you had a question?

DR. SAMPSON: Actually, I just wanted clarification from Dr. Orloff for my understanding. If one came to the conclusion that bone mineral density is a primary endpoint, how would that be reflected in the indication for the compound, and would you anticipate a fracture claim being allowed if one were able to show BMD as a adequate surrogate?

DR. ORLOFF: Well in my rather naive world, as I mentioned back at the beginning, the question we're asking is whether one can rely on bone mineral density in some instances as an adequate surrogate for a reduction in fracture risk even if we can't say exactly how much of a risk it involves.

So, analogous to the approval of statins

based upon LDL lowering in the absence of an effect demonstrated in a large endpoint trial with regard to reduction in the risk for heart disease. Although admittedly there is an implied claim of fracture benefit, it is not so stated in the labeling and promotion for the drug.

CHAIRMAN BRAUNSTEIN: Dr. McClung?

DR. MCCLUNG: To come back again to the issue about bone density as the alternative to fracture -- with drugs, in which that relationship, that there maybe a reasonable time when that would be appropriate, particularly if we're simply looking at other groups of patients with the same drugs we've studied, or with drugs in the same class in which the mechanism of action has been shown to be very similar, and when the dosing regimen is the same.

And, there are at least examples with bisphosphonates where a whole variety of alternative dosing regimens of different durations, which reflects on the Ibandronate data, provides a different pattern of suppression of bone turnover. And until that has been evaluated, that would be a restraint I think to

not allow bone density to be the only endpoint.

And certainly to go across classes, you've already got an example of that in the lipid field. You've already used that as your example. If you are comfortable with stating the changes in serum lipid levels reflected across the class of statins, you maybe comfortable with that, but you haven't approved the use of hormone replacement therapy for the reduction of heart disease on the basis of the reduction in lipid levels that are seen with that agent.

Using bone density as the surrogate across different classes of drugs would be analogous to that circumstance.

DR. ORLOFF: Well, let me just address that for a brief moment.

The willingness to accept any surrogate -- and a surrogate by definition is imperfect because it falls short of the ultimate endpoint of interest. We concede that. The willingness to accept it is based upon not only a robust predictable repeated effect on the marker, but also on the absence of any apparent

countervailing risk in the same of different body systems that you would estimate might adversely impact overall outcomes.

So, there is a judgment call always in reliance on the surrogate. We rely on them in the context of sort of a reasonable assurance of safety based upon, in many instances, very large exposures, long-term exposures, multiple -- a lot of experience, for example, with the class of drugs.

So in a case of the absence of labeling for cardiovascular risk reduction for estrogens based upon a lipid altering effect -- incidentally, they're not labeled as lipid altering drugs either. There's an appropriateness there. I can't say there's a reason, there's an appropriateness there.

Because of long standing, there has actually been some concern, doubt about either the possibility of countervailing general cardiovascular adverse effects like deep venous thromboses and/or more recently the possibility that there might indeed be coronary adverse effects. So, I don't think it makes it a non-starter, the whole issue of the

surrogate.

CHAIRMAN BRAUNSTEIN: Dr. Lukert?

DR. LUKERT: One thing that would seem helpful would be to combine markers of bone resorption or bone turnover with the bone density results. Particularly when we get into intermittent dosing, I think it might be helpful to observe whether or not you're getting good -- particularly with antiresorptive drugs, to see if you're getting consistent reduction in the resorptive markers that would add to maybe the validity of bone density measurements.

CHAIRMAN BRAUNSTEIN: Dr. Bone?

DR. BONE: I've asked Dr. Watts privately, and he said now I can respond to the question.

(Laughter.)DR. BONE: I think that the particular example that he raised is a vexing one because it is a member of the class that we've been talking about. It showed a substantial and statistically significant increase in bone density in the trial that was cited.

The problem is that we have explained this

ex post facto. We say now, "Well, maybe it was the markers". But, there was evidence that it also wasn't the optimal dose for bone density. So which is it? Or, is it both or is it something else?

The problem we have here is an example of a large trial with a drug in our best characterized class in which there was an increase in bone density and where there was a trend toward a reduction in the fracture rate, but it did not reach statistical significance according to the test supplied. And, I think this is very annoying.

(Laughter.)DR. BONE: I really wish that this trial had been done a little differently because I think we would be talking about this whole issue in a different way.

At the same time, it may be a lucky thing. Because, if we didn't have this trial to vex us, we would probably be much happier about accepting the idea that simply seeing a bone density increase with no fracture data and no marker data or no any other data but just the bone density increase for a drug in this class would be just fine. There was nothing to

suggest this drug was making the -- the implication of that trial was that the effect was simply insufficient to have a robust clinical effect. But, it puts us in a position when we want to generalize.

DR. COLMAN: Henry, has anyone seen the actual statistics on the fracture rate data for the Ibandronate trials?

I mean because if we're talking about a p-value of 0.5 versus a p-value of 0.06, there's a huge difference there. And I for one would not be willing to say that it was a complete disaster and that the BMD fracture relationship has been permanently smeared because of that. I think if the p-value was 0.06 or 0.07 -- or if they would've added 100 patients, it would've been 0.03.

CHAIRMAN BRAUNSTEIN: Dr. Aoki?

DR. AOKI: It seems to me that we're paying a lot of attention to the numbers in the human study, and I was wondering if the same attention shouldn't be directed at the preclinical or animal studies.

The issue with the bone mineral density is

a good surrogate. I think it can probably be best investigated if you looked at animals, looked at the relationship of bone volume, bone strength, and bone density in those animals.

I'd be kind of curious to see in the studies that you were referring to Dr. Bone, if you went backwards and when you found a clinical outcome in humans that was somewhat perplexing, to go back and see if those same problems were present in the animal studies. The problem may be that we don't have standard animal studies were we give a certain dose that will give rise to a certain degree of increase in bone mineral density and then extrapolate that in a stepwise fashion, both in animals and in humans.

CHAIRMAN BRAUNSTEIN: Dr. Temple?

DR. TEMPLE: Certainly in talking about surrogate endpoints generally, it's usually thought to be a bad thing if there's a well-done negative example. Now if the study was too small and other things were wrong with it, that doesn't count so much. But one of the reasons we still use blood pressure, to my knowledge, there's never been a negative placebo-

controlled study of any blood pressure agent even though it may be the drugs differ from one another. I had a related question and that is: If you establish somehow that a drug has a fracture effect in one setting, does that then settle the issue for all settings? In other words, is there a proof-of-principle thing here where let's say you do make a persuasive active control case in a very high-risk setting, does that then make everybody comfortable about the lower-risk setting? There have been people who've said changing dosage forms, changing regimens, things like that, that that's okay for BMD if it's the same drug. That sort of implies that nobody is too worried about the specific way it's used or the dose and a lot of other things, once you've established that this is the kind of drug that has not only a good effect on bone mineral density but also has a fracture effect. That could make a big difference. I have to tell you part of what I have in mind is that there seem to be circumstances in which one could ethically do an add-on study which might show a fracture effect where you might have difficulty doing a placebo-

controlled study against no treatment. So my question -- and I'm obviously very interested in the answer -- is: Is this something you have to show once and then it works for all of them, or do you have to sort of show it in each setting?

CHAIRMAN BRAUNSTEIN: Dr. Cummings.

DR. CUMMINGS: I don't know that we have a large enough database to answer it across things like women and men and steroid-treated patients and other treated patients.

There is, in cardiovascular disease as you know, pretty consistent effects across classes of patients for the relative risk. So the relative risk tends to remain constant for an intervention across various classes of patients: men, women, ages, and such like that. That allows you to generalize from those to the cardiovascular drugs you're used to.

There have been at least two or three trials now finding an interaction, however. That is that women who are more severely affected seem to have a greater relative risk of reduction for non-spine fractures, and women who are in low-risk populations

in fact don't have a reduction in relative risk of non-spine fractures. So with the bisphosphonate class, it may not be generalizable from high-risk to low-risk patients that it works.

And those have been reinforced by interaction terms. We don't know where that threshold is, but, no, you can't generalize from high-risk populations to low-risk populations for that one outcome. Otherwise, the database is not large enough.

DR. TEMPLE: Can you distinguish though between qualitative interactions and quantitative interactions? I mean, most people believe that in a lot of settings, yes, one group might be somewhat better affected than the other. But it would be a big surprise if it went the wrong way, which has major implications for how many of these studies you have to do.

DR. CUMMINGS: There are not enough events in the low-risk patients to say that it goes either no effect, or the wrong way, or the right way. The estimates are close to one for the low-risk populations.

In one trial that I know of, the Fracture Intervention Trial No. 2, it was done. And there are a couple of other examples where there is a much stronger effect in those with very low bone density than those without. So that proposition that the relative risk remains the same regardless of the risk of the population in the bone density, it does not seem to hold within osteoporosis for bisphosphonates.

But the major statement is that we don't have the same volume of data that you do in cardiovascular disease to be able to generalize.

DR. TEMPLE: That could imply you believe that you may have studies in severely -- in very high-risk people and that won't really tell you anything about the lower-risk people at all. So you have to do another study --

DR. CUMMINGS: The efficacy in fracture reduction.

DR. TEMPLE: Right.

DR. CUMMINGS: Again, the major point here is that the database is very limited for the other kinds of extrapolations.

CHAIRMAN BRAUNSTEIN: Dr. Grady is next.

DR. GRADY: I just want to bring us back to the bigger issue here, and I'll just give you my opinion. The question is: Is bone density an adequate outcome? And I think the question is: Is it an adequate outcome to register drugs for prevention of fracture?

So let me just first say my thinking about these things is quite different if we're talking about prevention than if we're talking treatment. So in terms of talking prevention, we're talking essentially about treating mostly women who are not symptomatic, particularly if we're just talking about low BMD.

I personally have never understood the difference between an indication for treatment of osteoporosis and management or treatment of osteoporosis. Osteoporosis basically is low bone density. It's really another surrogate outcome for risk of fracture. That's just an aside.

But let me just say when we're talking about prevention, I think we need to be much more careful that the benefit outweighs the risk than when

we're talking about treatment of symptomatic conditions. I heard Dr. Orloff say that for a good surrogate, it should represent the outcome in a way that's robust, repeatable, and reliable, which nobody has convinced me that BMD is today. And, secondly, there shouldn't be much of a possibility that this surrogate could have harmful effects, which, we've been discussing that there may be in certain situations.

So I would just say I'm not convinced that BMD is an adequate surrogate for fracture prevention. And I personally think that new drugs, which are going to be registered for prevention of fracture, should be shown to reduce the risk of fractures.

Going beyond that, we get into some more difficult issues like once a drug has shown fracture prevention, should we then approve it in a different risk group. It's also in my mind a more difficult question as to whether or not, once we have several drugs in a class that all have shown fracture prevention, whether or not we need to continue requiring fractures as an outcome. I think those are

questions that are much more difficult to discuss.

But in terms of new drugs -- and I guess for right now I would include the bisphosphonates because I don't think two drugs in the class is enough. I think eight statins is quite enough, but I personally don't think two bisphosphonates is. So I'll just register my opinion that I don't think BMD is an adequate outcome.

CHAIRMAN BRAUNSTEIN: Dr. Bone was next.

DR. BONE: Thank you. Dr. Cummings --

DR. ORLOFF: Make that a little clearer. No, I'm kidding.

(Laughter.)

DR. BONE: Don't be shy, Deb.

One of points that Dr. Cummings was just addressing was the generalized ability from high-risk to low-risk groups. But let's turn that around.

Steve, what would you say about our ability to generalize from relative risk reduction in the low bone density category to the patients at greater risk for postmenopausal osteoporosis?

DR. CUMMINGS: We have very little

experience. I mean, I'm more optimistic about from what I've seen in the data. I think that that would probably apply, but we just don't have very many trials to generalize from.

We know, for example, we can't generalize from a reduction risk of vertebral fractures to a reduction risk of non-spine fractures because they're a lot of exceptions to that rule. So seeing a vertebral fracture reduction in a low-risk population doesn't necessarily mean that we'll see reductions of other kinds of fractures.

DR. BONE: But within the category of vertebral versus vertebral or non-vertebral versus non-vertebral?

DR. CUMMINGS: If something reduces the risk of vertebral fractures in one population, I think we've got enough consistency across these databases that I would probably believe that would work for everyone. But we just have too many exceptions to the rule for other kinds of generalizability.

So if you're asking, Henry, we see a vertebral fracture reduction risk in low-risk