FNIH
Foundation for the
National Institutes of Health

August 9, 2018

Beth Walton MBA, PMP, RAC
Regulatory Project Manager
Office of New Drugs (OND)
Immediate Office (IO)
Biomarker Development and Regulatory Science
U. S. Food and Drug Administration

Re: Legacy Biomarker Qualification Project: Status Update for DDTBMQ000054

Dear Ms. Walton,

Attached to this email is a status update on the Letter of Intent submitted by the Foundation for the National Institutes of Health (application DDTBMQ000054) to the FDA in 2016. The team has since made significant progress based on the feedback received from the FDA. We are happy to share our project update per the guidelines requested by you for the legacy biomarker project status. Attached also is the original LOI and specific responses to FDA's comments, wherever possible. Additionally, the team has attached several relevant manuscripts that may aid in the review of this project.

Upon your review of the updated documents, the team would greatly appreciate advice on a few issues highlighted in the project update. Our goal is to address these concerns to FDA's satisfaction before we submit a qualification plan (QP) and full qualification package (FQP) per the new section 507 process.

Please do not hesitate to contact me or Tania Kamphaus (tkamphaus@fnih.org) should you need additional documents or wish to set up a meeting to discuss this updated document.

Sincerely,

Joseph Menetski, PhD
Associate Vice President, Research Partnerships
Director, Biomarkers Consortium
Foundation for the National Institutes of Health

Legacy Biomarker Qualification Project Status Update[1]

Administrative Information

Requesting Organization

Name: Biomarkers Consortium, Foundation for the National Institutes of Health (FNIH)

Address: 11400 Rockville Pike, Suite 600, North Bethesda, MD 20852

Phone: (301) 402-5311

Email: foundation@fnih.org

Website: https://fnih.org/what-we-do/biomarkers-consortium

Primary Contact

Name: Tania Kamphaus, PhD – Scientific Program Manager, Metabolic Disorders; and Joseph Menetski, PhD – Associate Vice President of Research Partnerships

Address: 11400 Rockville Pike, Suite 600, North Bethesda, MD 20852

Phone (301) 435-6247; (301) 594-6596

Email: tkamphaus@fnih.org; jmenetski@fnih.org

Alternate Contact

Name: Dennis Black, PhD

Address: Department of Epidemiology and Biostatistics, UCSF; 550 16th St., San Francisco, CA 94158

Phone: (415) 378-7144

Email: dblack@psg.ucsf.edu

Submission Date (MM/DD/YYYY): _____**4/26/16**_____

[1] The content you provide in this completed Status Update will be publicly posted as part of the section 507 transparency provisions.

I. Context of Use

A. Biomarker Category

Proportional change in dual-energy x-ray absorptiometry (DXA) bone mineral density (BMD) can be used to predict outcomes of hip fracture. We propose that results from DXA BMD scans in subjects at risk for hip and non-vertebral fracture can be a quantitative surrogate endpoint for response to investigational anti-osteoporosis drug treatments.

B. Intended Use in Drug Development

In order to facilitate development of additional osteoporosis medications, the current project aims to establish change in DXA BMD, as a surrogate endpoint for the established clinical efficacy endpoints of non-vertebral and hip fracture. Additionally, if qualified as a surrogate endpoint marker for non-vertebral and hip fracture, use of DXA BMD will allow pivotal trials to establish efficacy that can be accomplished with many fewer participants with abbreviated study durations of 1 to 2 years, as it will obviate the need for hip or non-vertebral fracture as the clinical endpoint. This can be followed by confirmation in phase IV studies after accelerated approval registration.

C. Context of Use Statement

Primary COU:

A minimum threshold increase in DXA BMD at the total hip over 24 months (for example 3%) as assessed in a randomized, controlled trial can serve as a surrogate endpoint for the clinical endpoints of hip and non-vertebral fracture risk reduction. The mean percent difference will be calculated within each treatment group and the increase calculated as the difference between the two treatment groups.

The target population of interest will be postmenopausal women with osteoporosis. We propose to more precisely define osteoporosis based on the treatment threshold guidelines of the National Osteoporosis Foundation (NOF) [1,2], which define treatment thresholds generally consistent with entry criteria used in past clinical trials which will serve as the evidence base for this qualification effort.

The future approval of a new agent based on a qualified BMD biomarker would also require preclinical studies to show proof of normal or improved quality bone and maintain the expected BMD to bone strength relationship. In addition, there must be evidence from a placebo- or active-controlled clinical randomized trial in humans,
of a reduction in vertebral fracture incidence.

II. Drug Development Need

Osteoporosis is responsible for more than 2 million fractures annually, including hip, vertebral, wrist, and other fractures. These fractures often lead to decreased quality of life and increased disability, and may contribute to premature death. The costs to society in the US are enormous: by 2025, annual direct costs from osteoporosis are expected to reach >$25 billion. Since 1996, FDA has approved at least eight new treatments that are highly effective in reducing osteoporotic fracture risk, including hip fracture risk

reductions of 40–50% [3] (see Fig. 1 below), non-vertebral fracture risk reductions of 20-25% (and vertebral fracture relative risk reduction by as much as 86% (abaloparatide).

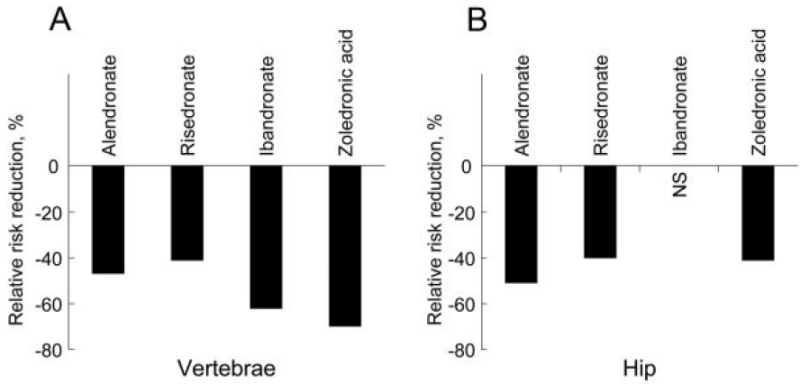


Figure 1 Summary of Bisphosphonate Fracture Reductions (up to 5 years)*.
Also ~25% reductions in non-vertebral fractures [3].

These treatments have the potential to greatly decrease the cost, morbidity, and mortality associated with osteoporosis. However, rare but troubling side effects have created concern among patients and clinicians. These concerns are likely the primary cause of the fact that usage of these highly effective medications has dramatically decreased over the last 10 years.

The decrease in osteoporosis treatment rates supports the urgent need for new medications, with improved safety profiles to increase patient and clinician choice and acceptability. However, the randomized trials supporting these approvals have all had fracture endpoints. The number of subjects required for these trials has grown increasingly large for a variety of reasons, including the need to enroll subjects with relatively low baseline fracture risk due to ethical issues associated with possible randomization to placebo when effective treatments are available. Most recently, a trial of a new anti-osteoporosis agent with a novel mechanism of action required a sample size of more than 16,000 patients for 5 years [4]. Such massive trials are no longer financially or logistically feasible due to high cost and complexity, precluding development of any new drugs for osteoporosis. Consequently, with only current medications, it would seem unlikely that we can significantly alleviate the huge costs, morbidity, and mortality associated with osteoporosis-related fracture in the US.

If the results of the proposed analyses enable the development of a pathway to qualification of DXA BMD change as a surrogate endpoint for fracture, the project will revolutionize both anti-fracture drug development and testing in future trials as well as better inform the clinical use of anti-osteoporosis treatments. For patients being treated for osteoporosis, this will result in more treatment options, improved therapies, and better guidance for their clinicians regarding how to use these treatments. Overall, results of this project will improve public health by reducing the occurrence of fractures and by allowing the more efficient and safe use of osteoporosis drugs.

III. Biomarker Information

A. Biomarker Name, Source, Type, and Description

Percent (%) change in DXA total hip bone mineral density (BMD) as a surrogate endpoint for hip and non-vertebral fracture risk reduction by measuring with DXA over a period of 24 months in a patient population of osteoporotic postmenopausal women DDT #: (DDTBMQ000054)

Type of Biomarker (Check relevant type[s])			
	Molecular	√	Radiologic/Imaging
	Histologic		Physiologic Characteristic
	Other (please describe):		

B. For Molecular Biomarkers, Please Provide a Unique ID

Not applicable.

Rationale for biomarker

Experimental studies of bone biomechanics have established that areal BMD (g/cm^2) is an accurate marker for mechanical strength of bone. Typically these biomechanical studies have gathered data from various animal models of experimentally produced osteopenia including rat, mouse, dog, and ovine skeletal studies, although biomechanical studies of human cadaveric specimens have confirmed many of the preclinical findings [5]. These studies have consistently shown that 60–70% of bone strength is accounted for by DXA BMD. In Figure 2 below, a preclinical animal model study shows the strong relationship between DXA BMD and bone strength as assessed by fracture load:

.

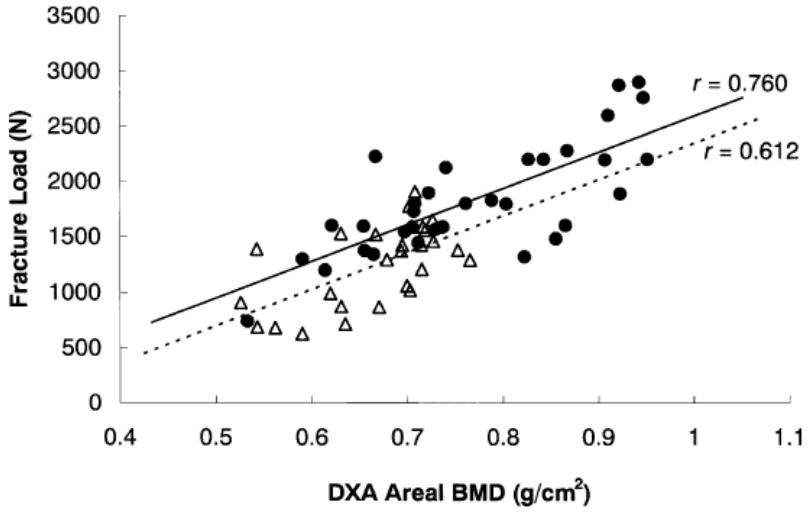


Figure 2 Preclinical and Animal Model Data

Bone mass (areal BMD) measured by DXA is correlated with the fracture load, with r = 0.760 and 0.612 for humerus (–●–) and femur (.....△....), P<0.001 in goat.

Observational data relating BMD at the hip to fracture risk in humans

BMD at the hip has been shown to be highly predictive of various types of fractures (including non-vertebral and hip fractures) in observational studies. One of the first studies of DXA BMD and fracture in human populations [6], the Study of Osteoporotic Fractures (SOF), studied a prospective cohort of 9704 US women from four US clinical sites. The results show a strong relationship between DXA BMD, assessed at the hip, and hip fracture risk (see Fig. 3 below). Note that the incidence of hip fracture in the lowest (age-adjusted) hipBMD quartile is approximately 10 times that of the highest quartile, and there is an almost linear increase in risk with decreasing BMD. For femoral neck (FN) BMD, these results correspond to a relative hazard per standard deviation BMD decrease of 2.6 (2.0, 3.6) with an area under the receiver operating characteristic (ROC) curve of 0.76. Similar results have been shown for total hip BMD. Longer-term studies in this cohort; studies in other cohorts in the US, Europe, and other parts of the world; and meta-analyses [7] have shown similarly strong relationships between DXA BMD and various types of fractures.

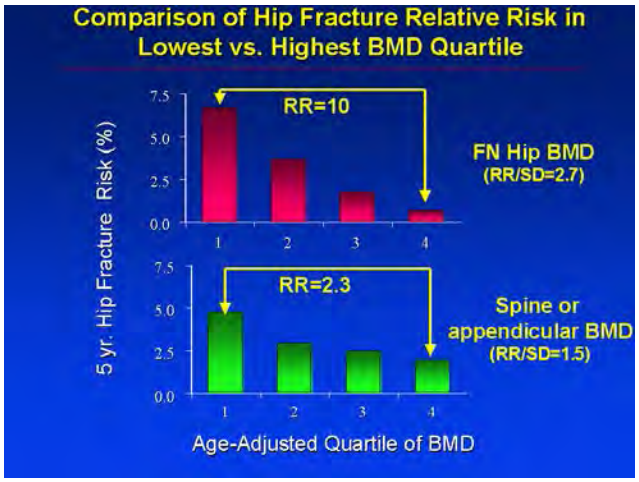


Figure 3 Risk of hip and vertebral fracture in a cohort of US women over age 65 by quartile of DXA BMD at the femoral neck of the hip [6].

Figure 4, below, shows the modeled relationship of 10-year hip fracture risk to hip BMD and shows the role that BMD plays independently of age in predicting fracture risk [8]. A single measure of BMD was recently shown to predict fracture risk for hip and non-vertebral for as long as 25 years [9]. The strong relationship between DXA BMD is recognized by the clinical community: DXA BMD is routinely part of the clinical assessment of osteoporosis risk in older women as well as men.

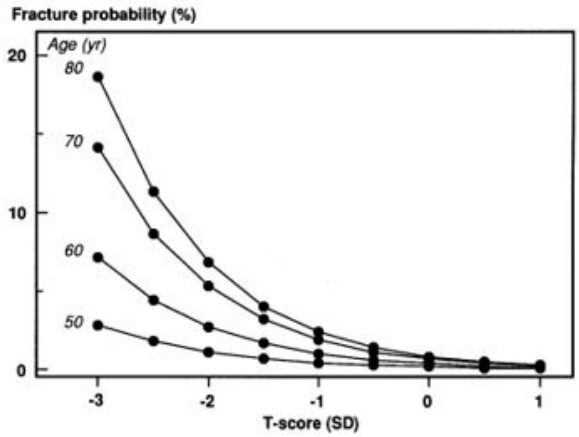


Figure 4 The relationship between BMD at the hip expressed as a T-score and hip fracture probability in women according to age. For any given T-score, the risk is higher with increasing age.

The ability of BMD to predict fracture [10] is much better than the ability of lipid levels to predict cardiovascular disease risk [11] or blood pressure to predict stroke [12]. These strong results are supportive of this surrogate application.

Natural history of the disease and associated risk factors

It has been known since the 1940s that the negative calcium balance and osteoporosis that developed in women after menopause could be reversed by estrogen treatment. Large studies, including the Women's Health Initiative (WHI) [13], have shown that hormonal therapy can decrease risk of hip and other fractures. However, because of the well-documented side effects of long-term estrogen-replacement therapy, such as increased cardiovascular events and breast cancer risks, osteoporosis drug discovery shifted to exploration of the safety and efficacy of agents with other mechanisms of action. These have included bisphosphonates (alendronate, risedronate, ibandronate, zoledronic acid), a physiologic modulator (calcitonin), anabolic agents (teriparatide, abaloparatide), selective estrogen receptor modulators (raloxifene, bazedoxifene), and modulators of fundamental bone cell biology regulation (denosumab) [14]. The largest experience has been with bisphosphonates, which are the most widely used drugs for the prevention of osteoporosis. The opportunistic discovery that they inhibited bone resorption via an anti-osteoclastic effect has resulted in their broad adoption as the first-line treatment after diagnosis of bone loss. Four different approved bisphosphonates are now marketed and are extremely effective in controlling osteoporosis. However, more recently, concerns about the development of rare side effects from bisphosphonate treatments, such as osteonecrosis of the jaw and atypical femoral fractures, have impacted compliance with their use and provided an impetus for discovery and development of anti-osteoporosis agents with novel mechanisms of action. In aggregate, the currently available various osteoporosis treatments markedly reduce vertebral fracture risk (by as much as 86%) [15], and reduce multiple vertebral fractures by as much as 90%. However, the effects of these treatments on hip fracture risk (40–50% reductions) and non-vertebral fracture risk (20–25%) are less pronounced. Therefore it is critically important to develop new interventions that can reduce hip and non-vertebral fracture risk more effectively without the side effects of current medications.

The potential of change in DXA BMD as a surrogate endpoint of interest for fractures in osteoporosis trials has been of some interest for a number of years. Indeed, the potential value of change in BMD to predict fracture reductions during clinical trials was reported over 10 years ago in two meta-regression analyses based on published trials of antiresorptive agents at that time, one for vertebral fractures [16] and the other for non-vertebral fractures [17].

The target population for this biomarker COU is postmenopausal women at high risk of fracture or fracture recurrence. This would potentially include women of all ethnicities over age 50 years (although some younger women may meet this criterion) who meet the definition of osteoporosis used in treatment guidelines in the US, which could include low BMD alone (below a T-score of −2.5), or a more mild reduction in BMD (termed osteopenia) with either a history of fracture, and/or those who meet specific thresholds for risk of hip fracture or major osteoporotic fracture [2]. While a threshold of BMD is part of these definitions, the association of the decline in bone mechanical strength, reflected quantitatively by BMD and the occurrence of fractures, is a continuum.

Is there an established "baseline" for the biomarker in the target patient population compared to healthy controls?

The NHANES III study [18] collected proximal femur BMD by DXA from a nationally representative sample of >7000 men and women age 20 years and older. From these data and others, "normal" BMD values are established and generally referred to as "peak" BMD. In general in women, BMD peaks between ages 20 and 40. Using these same data at "peak," standard deviations are calculated and used in the calculation of T-scores, which is the number of standard deviations that a woman is below "peak" mean. Basing thresholds for treatment on T-scores allows for standardization across machine manufacturers. There are many studies showing that women with lower BMD are at high fracture risk and specifically that women meeting our target population definition of osteoporosis are at higher risk of osteoporotic fractures.

There are systems for assessing fracture risk which include DXA BMD and are in common clinical use. In 1994, WHO defined osteoporosis by comparing DXA BMD values from a young normative database using the T-score concept [19]. Peak bone density is reached in normal individuals by age 30 years and ideally should be maintained at this level throughout life. However, BMD declines with age and as it declines, fracture risk increases. The T-score represents the standard deviation of the patient's BMD from the mean of a young-adult reference population, with a T-score of −2.5 and below defined as osteoporosis, between −1.0 and −2.5 defined as osteopenia, and −1.0 and above defined as normal. BMD data (generally transformed to the T-score metric) were collected and used as enrollment criteria in most of the clinical studies proposed for inclusion in the analysis to validate the BMD by DXA qualification claim for this project. Over the last 10 years, a statistical model, called FRAX (https://www.sheffield.ac.uk/FRAX/) incorporating BMD together with other risk factors has been developed, and is part of the NOF treatment guidelines most commonly used in the US. This model reports the estimated 10-year risk of hip and of major osteoporotic fractures and is included on many reports from DXA scans.

IV. Biomarker Measurement Information

A. General Description of Biomarker Measurement

Enter content here.

B. Test/Assay Information

Indicate whether the biomarker test/assay is one or more of the following:

i. Laboratory Developed Test (LDT) ○ Yes √ No

ii. Research Use Only (RUO) ○ Yes √ No

iii. FDA Cleared/Approved √ Yes ○ No
If yes, provide 510(k)/PMA #: multiple scanner vendors and software providers

iv. If the biomarker is qualified, will the test/assay be performed in a Clinical Laboratory Improvement Amendments (CLIA)–certified laboratory? ○ Yes √ No

v. Is the biomarker test currently under review by the Center for Devices and Radiological Health or the Center for Biologics Evaluation and Research? ○ Yes √ No ○ Don't Know

vi. Is there a standard operating procedure (SOP) for sample collection and storage? ○ Yes √ No

vii. Is there a laboratory SOP for the test/assay methodology? √ Yes ○ No

See Lewiecki et al. [20] Appendix 1 for full publication.

C. Biomarker Measurement

For the BMD measurements described in the published study-level meta-analysis to be included in the qualification evidence package [21], DXA was performed at the proximal femur and/or spine at baseline and follow-up. The duration of the studies and frequency of BMD measurement differed among studies, but most included follow-up DXA BMD measurements performed at 12 and/or 24 months so that changes over these periods could be measured. BMD was measured by different devices marketed by three different companies (Hologic Inc., Bedford, MA; GE Lunar, Madison, WI; Norland Corporation, Fort Atkinson, WI), although currently only Hologic and GE Lunar continue to provide densitometers.

i. Quality Control

<u>Information about quality control material or procedures</u>

DXA is an established methodology that has been in clinical use since 1987. The International Society for Clinical Densitometry (ISCD) was organized in 1993 and has developed standards for routine use of DXA in clinical and research contexts. The most recent update of these standards was published in 2016 as the Best Practices for Scan Acquisition and Analysis [20]. DXA facilities using DXA measurements to evaluate patients and monitor response to osteoporosis therapy should apply these standards to perform high-quality studies and derive clinically useful data.

The most relevant factors are shown in bold:

<u>DXA Best Practices Scan Acquisition and Analysis</u>

1.1. At least one practicing DXA technologist, and preferably all, has a valid certification in bone densitometry.

1.2. Each DXA technologist has access to the manufacturer's manual of technical standards and applies these standards for BMD measurement.

1.3. Each DXA facility has detailed standard operating procedures for DXA performance that are updated when appropriate and available for review by all key personnel.

1.4. The DXA facility must comply with all applicable radiation safety requirements.

1.5. Spine phantom BMD measurement is performed at least once weekly to document stability of DXA performance over time. BMD values must be maintained within a tolerance of ±1.5%, with a defined ongoing monitoring plan that defines a correction approach when the tolerance has been exceeded.

1.6. Each DXA technologist has performed in vivo precision assessment according to standard methods and the facility LSC (least significant change) **has been calculated.**

1.7. The LSC for each DXA technologist should not exceed 5.3% for the lumbar spine, 5.0% for the total proximal femur, and 6.9% for the femoral neck.

Interpretation and Reporting

2.1. At least one practicing DXA interpreter, and preferably all, has a valid certification in bone densitometry.

2.2. The DXA manufacturer and model are noted on the report.

2.3. The DXA report includes a statement regarding scan factors that may adversely affect acquisition/analysis quality and artifacts/confounders, if present.

2.4. The DXA report identifies the skeletal site, region of interest, and body side for each technically valid BMD measurement.

2.5. There is a single diagnosis reported for each patient, not a different diagnosis for each skeletal site measured.

2.6. A fracture risk assessment tool is used appropriately.

2.7. When reporting differences in BMD with serial measurements, only those changes that meet or exceed the LSC are reported as a change. BMD, bone mineral density; DXA, dual-energy x-ray absorptiometry; LSC, least significant change.

Notes on quality control procedures used in past clinical trials

Most studies in osteoporosis have DXA quality control standards. This is particularly true for more recent trials. As DXA has evolved as a technique, quality control guidelines have become more standardized. As a result, recent trials have generally had more rigorous quality control guidelines. An example from one recent trial (MK-0822 Fracture Trial DXA Imaging Charter) is attached as Appendix 2.

Cut point of percentage change in DXA BMD .

Percentage change in DXA BMD has been proposed for this qualification project as the most reliable metric. FDA responded to the proposed use of percentage change in BMD in our original Letter of Intent in 2016, stating that “Percentage change in DXA BMD should be the most appropriate measure as this metric would not be affected by the DXA machine type or normative database used. However, there may be a threshold of percent change in BMD, above which we are comfortable with stating fracture risk reduction has been demonstrated.”

The magnitude and duration of change in the biomarker required to demonstrate a clinically meaningful effect/impact or outcome

Establishing a specific level of BMD “cut point” could be considered as one of the surrogate qualification key objectives. In working with FDA, we believe that a threshold should be high

enough such that the probability of positive clinical endpoint reduction would be close to 100%, had a clinical endpoint study been done. Given the consequences of a "wrong" decision, we would suggest the probability of a positive clinical endpoint result should be much higher than 95%, perhaps as high as 99.9%. This issue will be discussed with FDA. Our preliminary results suggest that a BMD difference between active and control groups for total hip BMD over 24 months > 4–5% is the approximate range where hip and non-vertebral fracture risk reduction would be almost certain.

ii. Quality Assurance

Type of test

DXA utilizes low-dose x-rays at two energies to distinguish between hard and soft tissues.

Interfering substances

Not applicable.

Recognized standards used to support validation of the assay or measuring method

Attention to DXA instrument calibration, data acquisition and analysis, interpretation, and reporting is necessary for correct diagnostic classification and optimal fracture risk assessment by BMD monitoring. A best practices document summarizing the recommendations of the International Society for Clinical Densitometry (ISCD) describes quality standards for BMD testing at DXA facilities and provides guidance for DXA supervisors, technologists, interpreters, and clinicians [20]. These standards are proposed to be part of the surrogate qualification.

Detailed description of the specialized software needed

Examples of software available in different DXA scanners include Hologic Apex and GE-Lunar EnCore software. Current DXA software is highly automated for the placement of the skeletal region of interest (ROI), while the older software versions were completely manual. These software changes include adjustments to the absolute BMD values as well. For instance, the traditional recommendation regarding patient positioning for spine scans involved elevating the legs with a positioning block for pencil-beam systems. Currently, the Hologic fan-beam systems still use the positioning block while GE-Lunar offers the option of not elevating the legs, slightly altering the projection of the spine in the image. Comparability of measurements made using different systems with their associated proprietary software has been the subject of at least one study [22]. In this study, the relationship between Hologic Apex and GE Encore software was defined using linear regression. The BMD values from both systems were converted into sBMD using the Hui et al. formulas for spinal BMD [23]. The standardized bone mineral density (sBMD) values, derived using universal standardized equations, were shown to be equivalent within 1.0% for hip but significantly different for spine between the systems. Spine L1-L4 and L2-L4 sBMD mean differences were 0.042 g/cm^2 (4.1%) and 0.035 g/cm^2 (3.2%), respectively.

In the studies included in the current meta-analysis, where needed, standard equations to convert from Lunar and Norland to Hologic for the lumbar spine, femoral neck, and total hip [24,25] were used to generate standardized BMD (sBMD, mg/cm^2). The lumbar spine region L1-4 was used when available, otherwise L2-L4 was used. Finally, the NHANES data [21] was used to calculate BMD T-score for femoral neck.

iii. Sources and Quantification of Measurement Error

It has been reported that DXA scanner-induced variability may derive from long-term drift (approximately 0.5%/year), short-term drift (approximately 0.2–2.2%/day), inhomogeneity of the x-ray beam intensity over the tabletop (approximately 1%), and changes in internal filtration (approximately 0.5%) [26]. These observed drifts are fortunately negligible with respect to the precision adequate for clinical decision making. Fortunately, measurement drift is not a marked factor in the studies included in the validation meta-regression, where the placebo treatment groups would be subject to the same sources of instrument "drift" and still have shown the validity of BMD as a surrogate for the treatment effect. If approved as a surrogate, BMD difference between treatment groups measured in future trials would similarly not be subject to drift. Despite these recognized quality control procedures, operator certification and high-quality operator performance is still a prerequisite for proper QA in any setting.

D. Additional Considerations for Radiographic Biomarkers

See section IV.C., Biomarker Measurement information, Quality Assurance.

V. Assessment of Benefits and Risks

A. Anticipated Benefits

The potential public health benefits of this project are substantial. This project will address several of the most fundamental obstacles to the development of new treatments for osteoporosis, a disease that presents a large and growing global health burden. The results of these analyses will improve both anti-fracture drug development and testing as well as the clinical use of anti-fracture treatments. For patients with osteoporosis who are considering treatment, this will result in improved treatments, wider choice of treatments, and better guidance for their clinicians regarding how to use these treatments. Overall, this will improve public health by reducing the occurrence of fractures and by allowing the more efficient use of osteoporosis therapies.

Given the potential for accelerated regulatory approval based on a surrogate response, the benefits for the target group (postmenopausal women at high risk of fracture) could include earlier access to alternative novel treatments that are able to reduce their risk of fracture. Perhaps more effective treatments could be developed, compared with what is currently available. Also, since the development costs would not need such large and lengthy trials as are currently needed to prove hip fracture reduction, the development costs would be less, ultimately resulting in lower costs of treatment for patients.

B. Anticipated Risks

There are two primary risks that could result from the qualification of BMD difference as a surrogate measure for fracture risk:

1. If a treatment that is shown to increase BMD results in regulatory approval but is not, in fact, effective in reducing fracture risk, then potentially women could be treated with ineffective medication. This risk seems extremely unlikely since trials of comparable size with BMD differences (active vs. placebo) have shown an extremely high probability of fracture reduction for hip and non-spine fractures, and there is limited risk that the surrogate endpoint measuring increase in bone density will lead to a different clinical outcome. Furthermore, the requirement that the new treatment has been shown effective in increasing bone strength in preclinical

studies and reducing vertebral fractures in humans will minimize this risk. We believe our eventual submission will show that there are no examples of treatments that are effective preclinically, reduce vertebral fractures in a human RCT, and increase total hip BMD by more than 4–5% over 2 years that did not also reduce non-vertebral and hip fracture risk.

2. There may be safety issues with the drug that are not apparent in the small trials that would be sufficient for approval based solely on BMD surrogacy. This risk would be mitigated by performing larger and longer studies, perhaps in a post-marketing context, to assure adequate safety of the drug. The specific size, design, and duration of these trials (or studies) would need to be negotiated between the sponsor and FDA.

C. Risk Mitigation Strategy

If the threshold or percent change in BMD score to define an efficacious anti-osteoporosis treatment effect is insufficient to reliably predict reduction in fracture risk, a higher threshold score may have to be invoked. In addition, as described above in section B, a number of safeguards are built into the COU to further increase the probability that BMD difference above the proposed threshold would represent an agent that would significantly reduce hip and non-vertebral fractures.

D. Conclusions

BMD by DXA is a reliable, well-established, and reproducible measure of bone quality and fracture resistance, is on the causal pathway for development of postmenopausal osteoporosis, and is reasonably likely to predict beneficial response to treatment with a new anti-osteoporosis drug. Preliminary analyses, summarized below, and more extensive analyses that will be included with our eventual qualification package, show a strong quantitative relationship between BMD difference between treatments and fracture risk reduction. There is limited risk of generating inaccurate findings with BMD by DXA that would jeopardize patient safety or limit the treatment with other drugs already proven to be efficacious using conventional fracture reduction endpoints.

VI. Evaluation of Existing Biomarker Information: Summaries

A. Preclinical Information (as appropriate)

Not applicable.

B. Completed Clinical Information (as appropriate)

Overview:

In previous publications, for surrogate endpoint qualification, FDA specifies two types of analyses [27,28]:

1. A **meta-regression at a study level** that compares changes in the proposed surrogate marker (active vs. placebo) to the effect on the clinical endpoint.
2. An **individual patient analysis** that calculates percent of treatment explained **(PTE).**

We are currently working on three papers to address these two types of analyses:

1. A **meta-regression** based on meta-analysis of published trials.
2. A **meta-regression** using individual patient data (IPD) that we have compiled over the last 4 years in the first phase of this project.
3. An analysis of **PTE from the IPD** using a novel time-dependent covariate approach.

We anticipate the all three analyses will be published by the end of 2018 and all have been presented in initial form via abstracts for the past 3 years. Results for each of them are summarized in the background section below.

Background

Compilation of individual patient data: In the course of this project begun in 2013, we have collected data from over 50 randomized trials of osteoporosis medications including individual data from over 150,000 participants, including DXA-BMD and fracture outcomes in most of these participants. These data come primary from trials carried out by a dozen pharmaceutical companies as well as several NIH-sponsored trials. These trials include a total of 14 agents, some of which were eventually approved as well as others that were not submitted for approval after the trials. The agents represent a variety of mechanisms of action and include four bisphosphonates, denosumab, four SERMS, hormone therapy, three anabolic compounds, and odanacatib. In aggregate a total of ~17,800 data points on fracture outcomes were available (vertebral: ~5,200; non-vertebral: ~11,000; hip: ~1,600; see Table 1 below). A full listing of the individual studies is provided in Appendix 3.

Table 1: Bone Quality Database (total N>155,000) as of 8/1/2018

Drug (Approved)	Type	Company	N Trials	N Patients*	Received at UCSF?
Alendronate	Bisphosphonate	Merck	10	20,500	Yes
Risedronate	Bisphosphonate	PG...Actavis	6	16,900	Yes
Ibandronate	Bisphosphonate	Roche/Genentech	5	8,900	Yes
Ibandronate	Bisphosphonate	GlaxoSmithKline	1	93	Yes
Zoledronic Acid	Bisphosphonate (IV)	Novartis	8	~12,000	*PFT & RFT are in* (3 pending)
Denosumab	RL inhibitor	Amgen	1	7,800	Yes
Raloxifene	SERM	Lilly	1	7,700	Yes
PTH (1-34)	Anabolic	Lilly	4	2,400	Yes
Estrogen	HT	WHI/NHLBI	2	27,500	Yes
Lasofoxifene	SERM	Sermonix	4	10,960	Yes
Arzoxifene	SERM	Lilly	1	9,350	Yes
Bazedoxifene	SERM	Pfizer	3	~12,000	Yes
Clodronate	BP	McCloskey	2	5,300	Yes
PTH (1-84)	Anabolic	NPS/Shire	4	~3,000	Yes
Odanacatib	Cat-K inhibitor	Merck	2	~17,000	Yes
Abaloparatide	Anabolic	Radius	—	~3,000	In progress
		Total:	55	155,189	—

*Includes DXA BMD: 103,116.

The data were obtained in a variety of formats and converted into a standard template so that data from all trials are now readily analyzable from a set of six SAS data files including baseline data, DXA-BMD data, clinical fractures, and vertebral fractures. Two other types of data (bone turnover markers and data from hip QCT scans) were collected but are not part of this surrogacy application. These data are fully anonymized per HIPAA and EU requirements and are maintained on secure servers at the University of California, San Francisco (UCSF).

Data analyses: As described in section B.2 above, analyses as required by FDA have been performed. These include two meta-regressions (one from published data and second from our IPD data). Current progress on each of these three analyses is summarized below:

Meta-regressions #1: Study-level meta-regression from published data
Two previous meta-regressions from published data examined the relationship between changes in BMD and fracture risk reduction [16,17] in anti-resorptive trials published up to that time. The two publications included respectively 12 trials with 21,404 women and 18 trials with 26,494 women all conducted prior to 2002. We have updated these analyses and extended them by adding additional medications, BMD sites, and clinical endpoints.

The manuscript describing this analysis has been submitted to the *Journal of Bone and Mineral Research* and has been provisionally accepted pending some in-progress revisions. A current copy of the manuscript is included in the Appendix 4 including detailed methods and results.

To summarize the methods, we first conducted a systematic literature review. For studies meeting inclusion criteria, we performed a regression of the percent increase in BMD (treatment minus placebo) vs. the relative risk reduction for vertebral, hip, and non-vertebral fractures.

In general, the results support a strong and statistically significant relationship between BMD at the hip and each of the three fracture outcomes leading to the conclusion that % change in BMD (active vs. placebo) is strongly associated with corresponding fracture reductions. Results for total hip BMD for three key fracture types are shown below. More BMD/fracture combinations and more details about methods and results are shown in Appendix 4.

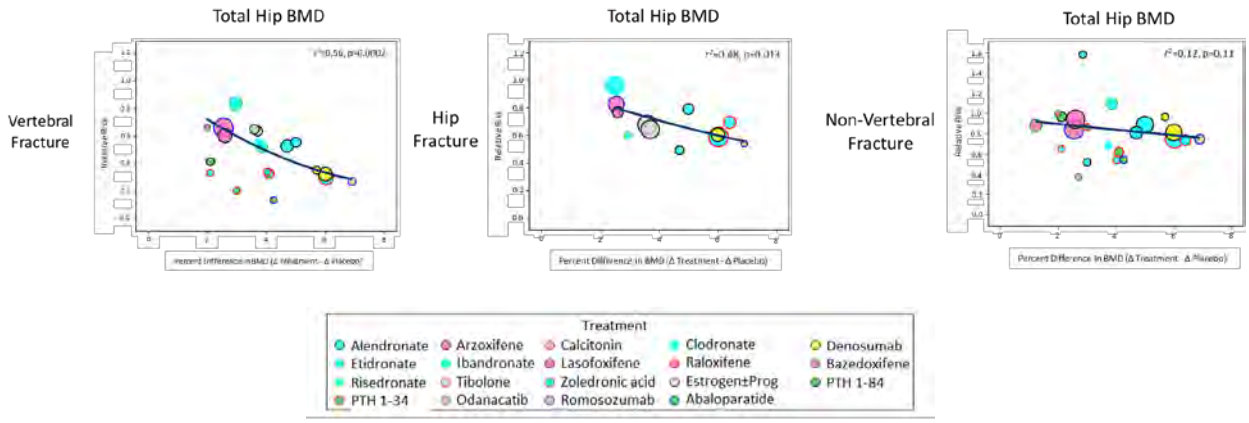


Figure 5 The relationship of change in BMD (% difference in treatment minus placebo) to relative risk for fracture. Each published study is represented by bubble proportion to the precision of the study.

Meta-regressions #2: Study-level meta-regression from individual patient data compiled during the project

Meta-regression #1, based on published data, has some important limitations including that the time for change in BMD varied according to study length and the definitions of fractures varied across studies. Also, looking ahead to qualification of change in BMD as a surrogate for drug approval, we wanted to more closely examine and compare specific times of measurement of change in BMD to determine the shortest that could adequately predict fracture risk. Limitations of published data for these types of meta-regressions have been previously noted [29]. To address these and other limitations, we performed a similar meta-analysis to that described above but used our compiled individual patient data. The current analyses reflect data from 21 trials of osteoporosis treatments (12 bisphosphonate; three SERM; two hormone therapy [HT]; two parathyroid hormone [PTH-like]; and one denosumab and one odanacatib trial) that included BMD measurements for at least one skeletal site: total hip (THBMD), femoral neck (FNBMD), or lumbar spine (LSBMD). The methods used in the data analysis are similar to those in meta-regression #1, although we calculated the change in BMD over fixed periods (12 and over 24 months) and also calculated the fracture definitions from the IPD. This paper was presented in preliminary form to the American Society of Bone and Mineral Research in 2016. A draft of the manuscript (to be submitted by early September 2018) is available on request.

The results below (Fig. 6) show that 24-month % change in total hip BMD (active vs. placebo) can reliably predict reduction in fracture risk. We prefer to focus on 24-month over 12-month BMD change since 24 months more reliably reflects BMD changes in anabolic drugs, the type most likely to be developed in the future.

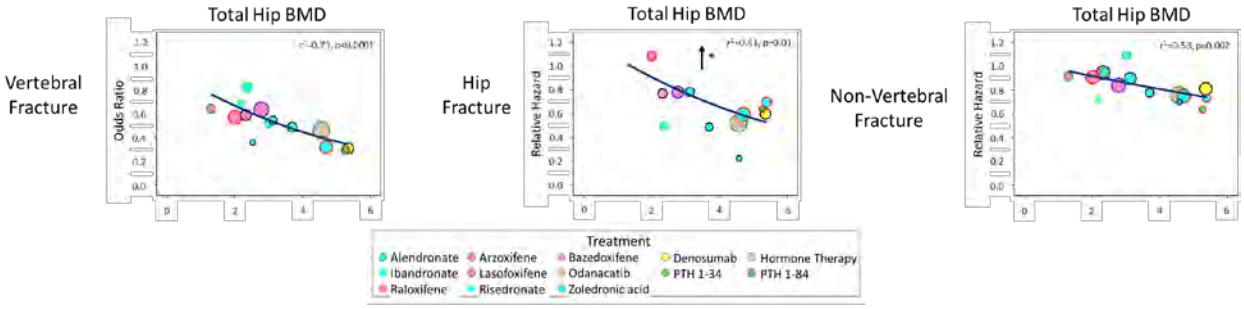


Figure 6 The relationship of change in BMD (% difference in treatment minus placebo) over 24 months to relative risk for fracture calculated from our compiled IPD. Each published study is represented by bubble proportion to the precision of the study.

Analysis of proportion of treatment explained (PTE) from individual patient data: We have made considerable progress on developing methods and performing the analysis calculating PTE from the IPD in our compiled data set. The methods are based on using BMD as a time-dependent covariate for predicting hip and non-vertebral fracture [30]. A slightly different method was used for vertebral fracture where the time to event is unknown. Further details of this analysis are given in an abstract (See Appendix 5) that has been accepted for an oral presentation at the upcoming ASBMR meeting (28 Sept–1 Oct 2018, in Montreal, Quebec).

The primary results are shown in Table 2 below. Overall, the PTEs are quite high: most are over 65% supporting the value of change in BMD as a surrogate measure for fracture.

Table 2: Percent of treatment effect (%, 95% CI) explained by changes in total hip (TH), femoral neck (FN), and lumbar spine (LS) at 24 months for vertebral, hip, and non-vertebral fractures

	ΔTH BMD	ΔFN BMD	ΔLS BMD
Vertebral Fx	62% (52, 73%) p<0.0005	61% (50, 72%) p<0.0005	28% (15, 41%) p<0.0005
Hip Fx	65% (31, 98%) p<0.0005	65% (24, 105%) p=0.002	44% (8, 79%) p=0.016
Non-vertebral Fx	67% (40, 94%) p<0.0005	69% (40, 99%) p<0.0005	54% (23, 84%) P=0.001

We anticipate that this paper will be completed and submitted for publication by November 1, 2018.

Summary of Current Progress

We are pleased at our successful completion and compilation of the individual patient data from over 150,000 people in randomized trials of osteoporosis medications. The analyses using these data (and the published data) will provide a good starting point for our discussions with FDA for analyses that will be

included in our qualification plan that would support approval for % change in BMD as a surrogate for fracture.

C. Summary of Ongoing Information Collection/Analysis Efforts and Current Progress

This qualification effort continues to leverage the active, ongoing, collaborative, and highly successful Bone Quality Project between UCSF and the Foundation for the National Institutes of Health's Biomarkers Consortium (FNIH BC), which was launched in 2013 (https://fnih.org/what-we-do/biomarkers-consortium/programs/bone-quality-project). The ongoing project includes partners from NIH, FDA, industry, and academia and has compiled a unique database from >50 randomized trials of osteoporosis medications, containing data on >150,000 individual patients of whom >80,000 have serial DXA BMD measurements. Access to such a wealth of data presents a unique opportunity to develop change in BMD as a surrogate measure for fracture risk reduction.

No new primary data collection will be required and with the acquisition of the odanacatib data in 2018, we had frozen the data set.

The progress on the two meta-regression analyses is described above and in draft manuscripts aimed for publication in 2018, and available on request. We believe that results from all three analyses are strongly supportive of qualification of change in total hip BMD as a surrogate for hip and non-vertebral fracture consistent with our proposed COU.

We have several types of analyses planned that will be implemented once the primary papers have been completed to expand these results and support this qualification effort. We also anticipate that the analysis plan that we will develop together with FDA will include additional analyses—for example, analyses in which the data will be divided and analyzed as an analysis and validation set, or analyses limited to specific subsets of patients or studies based on factors such as patient age, type of medication tested, etc. We look forward to developing a comprehensive set of analyses with FDA as part of our Qualification Plan.

VII. Knowledge Gaps in Biomarker Development

As part of the Foundation for NIH's Bone Quality Project, we are in the process of collecting patient-level data from most of the randomized controlled trials of osteoporosis drugs. We have already used these data in a recently published paper relating changes in bone turnover markers to fracture risk reduction [31].

A. List and describe any knowledge gaps, including any assumptions, that exist in the application of the biomarker for the proposed COU.

DXA is an established technology that has been used in both research and clinically since it was introduced in the late 1980s. Quality control procedures are well developed and the strong relationship of fracture risk has been established in studies involving many different populations. Furthermore, it is has been used in every major osteoporosis trial and results submitted as part of the approval packages for all osteoporosis treatments.

A number of questions, beyond the overall adequacy of change in BMD as a surrogate, will be explored as part of this biomarker qualification project. Our current COU specifies the population as postmenopausal women with osteoporosis consistent with the NOF treatment guidelines. However, the relationship of our proposed surrogate to fracture reductions might be stronger in more specific subgroups of patients such as women under age 80 or women with osteoporosis defined by BMD T-score below −2.5 alone. The availability of the individual patient data will allow us to explore these subgroups in more detail.

For example, we can limit analyses to those with osteoporosis by hip BMD (e.g., hip BMD T-score < −2.5) or to those patients with a combination of low BMD and history of spine or other fracture risk based on the FRAX model [31]. Another knowledge gap is the optimal fracture endpoint for fracture studies. It would be of interest to consider some additional endpoints commonly used in clinical trials including "all clinical fractures" (non-vertebral fractures plus clinical vertebral fractures) and "major osteoporotic fractures" (clinical vertebral fractures plus a subset of five other non-vertebral fractures). We might find our proposed surrogate is more strongly related to one or another of these endpoints providing further validation for it.

Importantly, we will want to confirm that the surrogacy of BMD would apply to treatments other than anti-resorptives (e.g., anabolics, etc.), supporting the generalizability of our results to new treatments. While most of the studies are of anti-resorptive compounds and studies of treatments with novel mechanisms of action are limited, we do have data for treatments with other mechanisms of action including three anabolic compounds and data from the large odanacatib study, which can be analyzed separately to assess the generalizability of results across treatment types.

In previous publications, FDA has suggested that in order to establish surrogacy, the proportion of treatment effect (PTE) explained by a surrogate marker such as BMD should be calculated. However, there are a variety of methods to perform these calculations. The project will explore ways to calculate PTE plus its confidence interval for % change in BMD for each fracture type. See Letter of Intent Appendix 6-7.

B. List and describe the approach/tools you propose to use to fill in the above-named gaps when evidence is unknown or uncertain (i.e., statistical measures and models, meta-analysis from other clinical trials).

The approach to address the first two evidence gaps is described in section A above. For the PTE calculation, the pooled data from the 21 trials included in the per-patient-level meta-analysis will be used to estimate the PTE at 24 months with a two-stage statistical procedure [32]. First, a linear mixed model is being used to estimate individual BMD trajectory, then a nested pooled logistic regression model is fitted to ascertain the effect of drug treatment on time to vertebral, non-vertebral, and hip fractures. PTE is then calculated as a % change in the coefficient for treatment after adding time-dependent BMD change.

C. Describe the status of other work currently underway and planned for the future toward qualification of this biomarker for the proposed context of use.

It would be of interest to incorporate other variables together with BMD that would support BMD's qualification, such as changes in serum markers of bone turnover or estimated bone strength

derived from finite element modeling of bone strength in QCT scans. We are interested in exploring whether some combination of biomarkers could be more informative than DXA BMD alone using our compiled data.

ATTACHMENTS/APPENDICIES 1-7

Optional* – If you have other supporting information you would like to provide, please submit as attachment(s).

**Optional information will not be posted publicly.*

PUBLICATIONS/REFERENCES

1. Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. N Engl J Med 2016;374:254–62.
2. Cosman F, de Beur SJ, LeBoff MS, Lewiecki EM, Tanner B, Randall S, Lindsay R. Clinician's Guide to Prevention and Treatment of Osteoporosis. Osteoporos Int 2014;25:2359–81.
3. Khosla S, Bilezikian JP, Dempster DW, Lewiecki EM, Miller PD, Neer RM, Recker RR, Shane E, Shoback D, Potts JT. Benefits and risks of bisphosphonate therapy for osteoporosis. J Clin Endocrinol Metab 2012;97:2272–82.
4. Bone HG, Dempster DW, Eisman JA, Greenspan SL, McClung MR, Nakamura T, Papapoulos S, Shih WJ, Rybak-Feiglin A, Santora AC, Verbruggen N, Leung AT, Lombardi A. Odanacatib for the treatment of postmenopausal osteoporosis: development history and design and participant characteristics of LOFT, the Long-Term Odanacatib Fracture Trial. Osteoporos Int 2015;26:699–712.
5. Ammann P, Rizzoli R. Bone strength and its determinants. Osteoporos Int 2003;14 Suppl 3:S13–8.
6. Cummings SR, Black DM, Nevitt MC, Browner W, Cauley J, Ensrud K, Genant HK, Palermo L, Scott J, Vogt TM. Bone density at various sites for prediction of hip fractures. The Study of Osteoporotic Fractures Research Group. Lancet 1993;341:72–5.
7. Marshall DA, Sheldon TA, Jonsson E. Recommendations for the application of bone density measurement. What can you believe? Int J Technol Assess Health Care 1997;13:411–9.
8. Kanis JA, Harvey NC, Cooper C, Johansson H, Oden A, McCloskey EV. A systematic review of intervention thresholds based on FRAX : A report prepared for the National Osteoporosis Guideline Group and the International Osteoporosis Foundation. Arch Osteoporos 2016;11:25.
9. Black DM, Cauley JA, Wagman R, Ensrud K, Fink HA, Hillier TA, Lui LY, Cummings SR, Schousboe JT, Napoli N. The Ability of a Single BMD and Fracture History Assessment to Predict Fracture Over 25 Years in Postmenopausal Women: The Study of Osteoporotic Fractures. J Bone Miner Res 2018;33:389–95.
10. Marshall D, Johnell O, Wedel H. Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. Bmj 1996;312:1254–9.
11. Kannel WB, Castelli WP, Gordon T. Cholesterol in the prediction of atherosclerotic disease. New perspectives based on the Framingham study. Ann Intern Med 1979;90:85–91.
12. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. Stroke 1994;25:40–3.
13. Cauley JA, Robbins J, Chen Z, Cummings SR, Jackson RD, LaCroix AZ, LeBoff M, Lewis CE, McGowan J, Neuner J, Pettinger M, Stefanick ML, Wactawski-Wende J, Watts NB. Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial. Jama 2003;290:1729–38.

14. Khosla S, Hofbauer LC. Osteoporosis treatment: recent developments and ongoing challenges. Lancet Diabetes Endocrinol 2017;5:898–907.
15. Miller PD, Hattersley G, Riis BJ, Williams GC, Lau E, Russo LA, Alexandersen P, Zerbini CA, Hu MY, Harris AG, Fitzpatrick LA, Cosman F, Christiansen C. Effect of Abaloparatide vs Placebo on New Vertebral Fractures in Postmenopausal Women With Osteoporosis: A Randomized Clinical Trial. Jama 2016;316:722–33.
16. Cummings SR, Karpf DB, Harris F, Genant HK, Ensrud K, LaCroix AZ, Black DM. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. Am J Med 2002;112:281–9.
17. Hochberg MC, Greenspan S, Wasnich RD, Miller P, Thompson DE, Ross PD. Changes in bone density and turnover explain the reductions in incidence of nonvertebral fractures that occur during treatment with antiresorptive agents. J Clin Endocrinol Metab 2002;87:1586–92.
18. Looker AC, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP, Johnston CC, Jr., Lindsay RL. Proximal femur bone mineral levels of US adults. Osteoporos Int 1995;5:389–409.
19. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. World Health Organ Tech Rep Ser 1994;843:1–129.
20. Lewiecki EM, Binkley N, Morgan SL, Shuhart CR, Camargos BM, Carey JJ, Gordon CM, Jankowski LG, Lee JK, Leslie WD. Best Practices for Dual-Energy X-ray Absorptiometry Measurement and Reporting: International Society for Clinical Densitometry Guidance. J Clin Densitom 2016;19:127–40.
21. Bouxsein ML, Estell R, Lui LY, Wu L, De Papp AE, Grauer A, Marin F, Cauley JA, Palermo L, Bauer DC, Black DM. Change in bone density and reduction in fracture risk: a meta-analysis of published trials (in press). 2018.
22. Fan B, Lu Y, Genant H, Fuerst T, Shepherd J. Does standardized BMD still remove differences between Hologic and GE-Lunar state-of-the-art DXA systems? Osteoporos Int 2010;21:1227–36.
23. Hui SL, Gao S, Zhou XH, Johnston CC, Jr., Lu Y, Gluer CC, Grampp S, Genant H. Universal standardization of bone density measurements: a method with optimal properties for calibration among several instruments. J Bone Miner Res 1997;12:1463–70.
24. Genant HK, Grampp S, Gluer CC, Faulkner KG, Jergas M, Engelke K, Hagiwara S, Van Kuijk C. Universal standardization for dual x-ray absorptiometry: patient and phantom cross-calibration results. J Bone Miner Res 1994;9:1503–14.
25. Hanson J. Standardization of femur BMD. J Bone Miner Res 1997;12:1316–7.
26. Sievanen H, Oja P, Vuori I. Scanner-induced variability and quality assurance in longitudinal dual-energy x-ray absorptiometry measurements. Med Phys 1994;21:1795–805.
27. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clin Pharmacol Ther 2015;98:34–46.
28. Amur SG, Sanyal S, Chakravarty AG, Noone MH, Kaiser J, McCune S, Buckman-Garner SY. Building a roadmap to biomarker qualification: challenges and opportunities. Biomark Med 2015;9:1095–105.
29. Korn EL, Albert PS, McShane LM. Assessing surrogates as trial endpoints using mixed models. Stat Med 2005;24:163–82.
30. Austin M, Yang YC, Vittinghoff E, Adami S, Boonen S, Bauer DC, Bianchi G, Bolognese MA, Christiansen C, Eastell R, Grauer A, Hawkins F, Kendler DL, Oliveri B, McClung MR, Reid IR, Siris ES, Zanchetta J, Zerbini CA, Libanati C, Cummings SR. Relationship between bone mineral density changes with denosumab treatment and risk reduction for vertebral and nonvertebral fractures. J Bone Miner Res 2012;27:687–93.
31. Kanis JA, Harvey NC, Johansson H, Oden A, Leslie WD, McCloskey EV. FRAX Update. J Clin Densitom 2017;20:360–67.
32. Bauer DC, Black DM, Bouxsein ML, Lui LY, Cauley JA, de Papp AE, Grauer A, Khosla S, McCulloch CE, Eastell R. Treatment-Related Changes in Bone Turnover and Fracture Risk Reduction in Clinical Trials of Anti-Resorptive Drugs: A Meta-Regression. J Bone Miner Res 2018;33:634–42.

1. Administrative structure

This proposal will be formally submitted by the Foundation for NIH Bone Quality Project Team which is led by Dr. Dennis Black, PhD, as PI at UC San Francisco. Co-PI's include Dr. Douglas Bauer, MD at UCSF and Dr. Mary Bouxsein at MGH/Harvard. Other investigators or consultants include Dr. Charles McCulloch, PhD, Head of Biostatistics at UCSF, Dr. Richard Eastell, MD, PhD at U. Sheffield, UK and Dr. Tony Keaveny at O.N Diagnostics. The Project Team is chaired by Dr. Gayle Lester, Program Director at the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). Other members of the Project Team include stakeholders from NIH, FDA, ASBMR as well as funders.

2. Biomarker Qualification Overview

A. Introduction

Over the last 20 years, about 8 new treatments for osteoporosis have been approved by the FDA. The use of these treatments has the potential to greatly decrease the cost and morbidity associated with osteoporosis. The underlying randomized trials supporting these approvals have all had fracture endpoints. Vertebral fracture endpoints are required by FDA but increasingly, non-vertebral and hip fractures have been added as primary and/or secondary endpoints. At the same time, the increasing availability of approved drugs with proven efficacy has led to the ethical requirement of more limited use of placebo or the recruitment of low risk populations. Due to both of these factors, the size of trials has grown considerably so that the most recent trial of odanacatib required a sample size of more than 15,000 patients. Such large trials are no longer feasible to carry out.

At the same time, there is clearly a need for new treatments as evinced by fact that current treatments have rare, but troubling, side effects and while they can reduce risk, the reductions are relatively modest. Thus, the clinical world is caught in a bind of needing new treatments but the size and cost of trials to prove their efficacy and safety is prohibitive. There is a need for innovative solutions which was the subject of a recent FDA workshop. Among these possible solutions, the use of surrogate, biomarker endpoints seems to be the most feasible solution.

B. Proposed context of use

1. Use Statement
   This draft guidance provides qualification recommendations for the use of the treatment-related difference in total hip bone mineral density (BMD), expressed as the mean percent from baseline, compared to placebo, after two or more years as an efficacy response biomarker as a surrogate for hip fracture reduction.

2. Conditions for Qualified Use
   Prediction of reduction in hip fracture risk following treatment:

   a. Patient Population
      Postmenopausal women at high risk for fracture. For example, the subjects should have low bone mineral density and/or a history of prior hip or vertebral fracture. This guidance is not currently intended for use in men or in patients with secondary osteoporosis.

b. Clinical trial context
This biomarker would be used to support a claim of hip fracture reduction in a placebo-controlled clinical trial. The placebo-group should receive adequate calcium and vitamin D. This biomarker would only be valid when the specific agent has been shown to reduce vertebral fractures in a randomized clinical trial.

c. Quantitative imaging biomarker
Total hip BMD should be measured by dual-energy X-ray absorptiometry at baseline and following two or more years of treatment with a potential therapeutic agent. The measurement should be made according to a standard operating procedure and routine quality control procedures should be conducted.

C. High-level data description

1. Overview of evidence base
There is a very large evidence base that could be used to address the relationship of change in hip BMD with treatment and fracture reduction since hip BMD has been included in all randomized, controlled trials that have been conducted for possible anti-fracture drug approvals. These trials have included a variety of fracture endpoints. Most of these trials have been published and we will conduct a meta-analysis of these published data. In addition, as part of our FNIH Bone Quality project, we are collating individual patient-level data from most of these trials and can use it to determine the percent of treatment effect explained. In addition, we plan to use this large database for other analyses to elucidate other aspects of this relationship including subgroup analyses, the impact of treatment type and the impact of study duration on this relationship. A list of such trials, for which our project has obtained or will obtain the patient-level study data, is attached as appendix A. We estimate that there will be a total of 25-30 randomized trials including >100,000 patients with hip BMD and fracture outcomes for which we will have the complete patient-level data.

2. Summary description of proposed methods
We look forward to working closely with the FDA in the detailed planning of these analyses. Specifically, we plan at least two types of analyses which are consistent with FDA guidance (Amur et al 2015):

a. **Study-level meta-analyses** will determine the relationship between BMD change and fracture reduction across studies. There is a very large evidence base from randomized trials that will be used to address the relationship of change in hip BMD with treatment and fracture reduction since hip BMD has been included in all randomized, controlled trials that have been conducted for possible drug approvals. About 15 drugs have been tested in these trials which have included a variety of fracture endpoints. Most of these trials have been published. Our meta-analyses will provide a graphical presentation of the relationship of hip BMD change to fracture reductions across trials. We will also use non-linear regression models to summarize these relationships.

Statistics from these analyses will include a fitted non-linear regression model which will link the change in BMD to a relative risk for fracture incidence. In the preliminary analyses shown, we are using a non-linear model to link the log of the relative risk to the

% change in BMD (active minus placebo). These calculations will also include an R-squared value which will give an idea of the spread of the results around the regression line. In conjunction with the FDA, we will develop a complete statistical analysis plan which may include plans for sensitivity analyses in subsets of studies based on the class of drug (e.g., bisphosphonates only, anti-resorptive only, or anabolic only), study duration (e.g., limited to studies ≥ 2 years) and other factors. We plan to perform these analyses for total hip, femoral neck and lumbar spine BMD using change in BMD over 3 years as well as over 1 year.

We have performed preliminary analyses (currently being finalized), in which we have examined the relationship between the change in total hip BMD (over the full trial length) and hip fracture (Figures 1 and 2). These analyses show a particularly strong relationship between the change in total hip BMD and the reduction in hip fracture (Figure 2): the estimated reduction in hip fracture is 9% for a 2% increase in total hip BMD, whereas the estimated reduction in hip fracture is 42% for a 6% increase in hip BMD. We expect these analyses to be completed by the end of June 2016.

Fig 1: Study level meta-regression of % hip BMD change predicting hip fracture relative risk (preliminary results)*

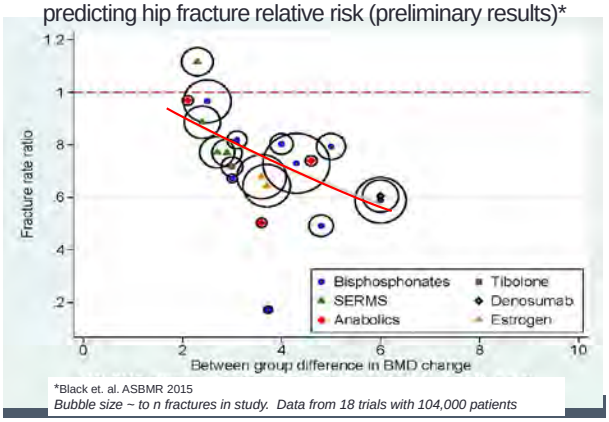
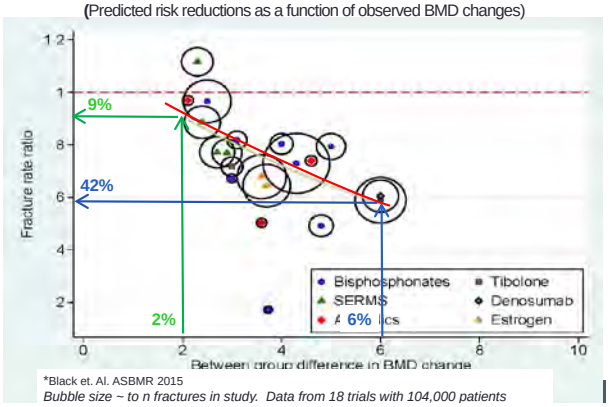


Fig 2: Study level meta-regression of % hip BMD change predicting hip fracture relative risk (preliminary results)*
(Predicted risk reductions as a function of observed BMD changes)



b. **Individual-level meta-analyses from individual patient data:** Using our individual-level data from over 100,000 patients in randomized, controlled trials, we will also provide other detailed meta-analyses pooling these data. These analyses can add to the study-level analyses in a number of ways. Firstly, we can calculate the percent treatment explained (PTE) plus its confidence interval for the % changes in BMD and each fracture type. Secondly, we can provide descriptive analyses based on individual patient characteristics in subgroups of patients. For example, we can limit analyses to those with osteoporosis by hip BMD (e.g., hip BMD T-score < -2.5) or to those patients with a combination of low BMD and history of spine or other fracture or risk based on the FRAX model. We can also examine subgroups defined by age or BMI. Thirdly, we can provide descriptive data using individual characteristics. Importantly, we will want to examine whether our results apply only to specific subgroups of treatments (e.g., anti-resorptives) or can be generalized to treatments which work through new mechanisms. For example, we will categorize individuals by their change in BMD and provide fracture reductions according to % change in BMD. We recognize the limitations and potential biases using post randomization criteria but would hope that these analyses will augment other analyses to give a more complete

description of the change in BMD and its relationship to fracture risk. As with the study-level meta-analyses, we look forward to engaging with the FDA during the planning period to develop a comprehensive analysis plan with a view toward the qualification of DXA hip BMD as biomarker.

c. **Other types of analyses:** We have additional sources of data that may be useful in evaluating other aspects of the relationship between BMD and fracture risk. For example, since UCSF serves as the Coordinating center for the Study of Osteoporotic Fractures (SOF, about 10,000 women over age 65 followed since 1986), we can provide detailed calculations of the relationship between a single assessment of hip BMD and future fracture risk. Other UCSF-Coordinated observational studies include the Osteoporotic Fractures in Men (MrOS) study (about 6,000 men over age 65 followed since 1991) and other observational studies. Many of these analyses have already been performed and published, but additional analyses can be performed which are supportive of this FDA application.

d. Other types of analyses that might be discussed with the FDA include the possibility of combining other types of biomarkers (e.g., bone turnover markers from serum or urine or finite element strength from QCT scans) with hip BMD.

D. Additional resources

In addition to the data sources described in the sections above, we are hopeful to eventually obtain data and include results from drugs currently in development including abaloparatide, odanacatib and romosozumab. These should be helpful in providing validation of our results across new drugs including drugs with different mechanisms of action.

E. Submission to other regulatory agencies

Unknown at this time

3. Process-related questions for FDA

We look forward to working closely with the FDA in developing our plans for analysis. We would be particularly interested in the next steps including meetings, establishing of a timeline for qualification and developing the outline of a briefing book.

We also have some specific questions that we look forward to discussing with FDA as well other members of our consortium. Some examples include the following:

1. We might also want to consider if there is any context under which changes in hip BMD could be used as evidence of anti-fracture efficacy without concurrent or prior evidence of the drug's efficacy in reducing vertebral fracture risk.

2. It would be of interest to incorporate other variables that would support BMD's qualification, such as changes in serum markers of bone turnover or estimated bone strength derived from finite element modeling of bone strength in QCT scans. We are interested in exploring whether some combination of biomarkers could be more informative than DXA BMD alone.

3. In this Letter, for the COU we have assumed that the results would apply only to post-menopausal women. However, there are other groups of interest. For example, could the COU be broader to include men or people with other specific conditions or medication use (e.g., corticosteroids)?

4. Would it be useful to consider only total hip BMD or perhaps changes in other BMD sites including femoral neck BMD or even lumbar spine?

5. If we are going to rely on DXA BMD as surrogate measure, is percentage change in DXA the most reliable metric or should be consider achievement of a specific level of BMD (a "target") a better surrogate?

6. In our analysis plan for the individual data, we plan to include percent of treatment explained (PTE). We would welcome discussion with the FDA about the value of PTE that might be considered adequate for this analysis and whether there are alternative statistical approaches to PTE.

A. Clinical and Biomarker Considerations:

1. *Provide the rationale for focusing primarily on hip fracture as opposed to vertebral fracture (the primary efficacy endpoint for osteoporosis).*

We recognize that the relationship of BMD change to vertebral fracture reduction is particularly strong. In the FDA response to our questions (below), they said that they want to see clinically proven reductions in an RCT in at least one of the 3 major types of fractures. If that is the case, then we believe that in general, it is easiest for sponsor to show vertebral fracture reductions. However, we would look forward to discussing with the FDA the possibility of broadening the COU and saying that reductions for at least one type of fracture would be required for approval and indications for the other two could be shown by change in BMD.

2. *Clarify what is meant by a high risk patient population (for example, list the inclusion criteria used to define a high risk patient in each study).*

This has been clarified in the Legacy Status Update form that the patient population is post-menopausal women who meet osteoporosis consistent with the NOF treatment guidelines which are commonly used throughout the US and base the definition of a combination of BMD, spine/hip fracture history and fracture risk estimated by FRAX. While the inclusion criteria for the trials which provide the evidence base for this proposal have varied, We believe these guidelines are generally consistent with those criteria.

3. *Explain whether you are proposing to establish Bone Mineral Density (BMD) as a validated surrogate endpoint or a reasonably likely surrogate endpoint. Please refer to BEST glossary for definitions at http://www.ncbi.nlm.nih.gov/books/NBK326791/*

We are proposing that the weight of evidence supporting the utility of BMD change, as determined by best practices DXA scan technology, when used to assess post-menopausal osteoporotic women, should be sufficient to establish it as a validated surrogate endpoint to predict fracture risk reduction resulting from intervention with a novel anti-resorptive drug.

B. Statistical and Analytical Considerations:

4. *Provide detailed documentation on the data structure and standardization of the data for the database that will be used to validate total hip BMD as a surrogate for hip fracture reduction. Include details about individual patient data quality, loss/exclusions (if any), and how missing data will be handled, etc.*

Details of the database design and the statistical methods will be developing in conjunction with the FDA in developing the final analysis plan over the next 1-2 years. However, if the FDA would like the description of the SAS database that we have developed to compile the data from the clinical trials, this can be provided at any time.

5. *Provide a detailed statistical analysis plan for both the study level and individual level meta-analyses.*

We appreciate the importance of pre-defining these details and will include them as we develop the analysis plan for qualification in conjunction with the FDA. Details for the two meta-regressions (published level and individual level) are provided in the draft manuscripts that are provided as attachments to the Legacy Status Update form but will developed further in the analysis plan that we hope to develop with the FDA. Details for the IPD analysis of PTE are

outlined in the attached abstract but will be expanded in the draft of the manuscript current being written.

a. *Clearly define the study populations and subgroups of interest. Inclusion and exclusion criteria for analysis and stratified factors should be pre-specified*.

See above

b. *To explore the relationship between total hip BMD and hip fracture, different candidate statistical models can be proposed for our review, e.g. time to fracture analysis or joint modeling approach. The prediction performance of the chosen model should be evaluated. The duration of drug use and other important risk factors should be taken into account. The same analyses can be conducted with placebo treated subjects only to understand the natural relationship between total hip BMD and hip fracture.*

We will include these types of analysis in our discussions with FDA.

c. *We suggest use of forest plots (study level vs. drug class etc.) to present treatment effects. Discussion of heterogeneity among studies or among drug classes should be part of the analysis plan and the clinical report.*

We have performed sensitivity analyses among study duration and drug class and they will be part of our eventual submission to FDA. We would look forward to further discussions with the FDA about how Forrest plots can be used in conjunction with these meta-regression.

d. *As noted in our response to Question D6 below, we can discuss the percent of treatment effect (PTE) explained or other metric you may propose. When preparing your analysis plan, descriptions on how the PTE is calculated, e.g., in the non-linear model that you currently propose, should be included in the plan.*

In conjunction with Drs. Vittinghof and McCulloch at UCSF, we have extensively developed and tested a model for calculating PTE using a time dependent covariate approach using Cox models for non-vertebral and hip fractures and look forward to detailed feedback from FDA about these methods and alternatives.

6. *We would like to understand the variability associated with multiple vendor platforms and any potential variations in the image acquisition process. Provide relevant information including quality controls (QCs), algorithms and reproducibility in your initial briefing package submission*.

For future use of BMD as a surrogate measure, we intend to require rigorous quality control consistent with ISCD recommendations (see further details in Legacy Status Update form). It would be challenging to provide extensive details of QC from previous studies since we do not have this information nor is it published. We appreciate the importance of biomarker QC but have realized that in the context of a surrogate endpoint biomarker, quality control plays a very different role than for a biomarker to be used in an individual patient since this biomarker would be used only with groups and in the context of a controlled trial. We look forward to discussing this very set of issues with FDA.

7. *Provide a discussion of the type and frequency of calibration/QC needed to obtain acceptable performance. How would drift in Dual Energy X-ray Absorptiometry (DXA) device output be accounted for over the duration of a future study?*

See response to question 6.

8. *Provide and justify a complete list of technical specifications for the devices used in the meta-analysis studies, as these will constitute minimum specifications required for future studies. This includes DXA devices and any other devices for additional variables. For example, technical specifications would include DXA precision, Quantitative Computed Tomography (QCT) voxel size (for estimated bone strength derived from Field Emission Microscopy (FEM), etc.*

See response to question 6.

9*. Provide and justify a complete list of practices for DXA measurement and reporting for the sites participating in the meta-analysis, as these will constitute minimum practices required for future studies. For an example, please see EM Lewiecki et al., "Best practices for dual-energy x-ray absorptiometry measurement and reporting: International Society for Clinical Densitometry Guidance," J. Clin. Densitometry: Assessment & Management of Musculoskeletal Health, vol. 19, no. 2, 127-140, 2016*.

We agree that these are excellent recommendations for use of DXA and it is fortunate the procedures for DXA are sufficiently mature that such recommendations are available. We have included extensive details about these recommendations in the Legacy Status Update form.

10. *In section C-2-b, it was mentioned that a detailed meta-analysis would be provided by pooling the data over patients and trials. For different trials, patient population, site, and skill of the technician may impact the precision of the BMD measurements. Discuss the reproducibility of the BMD measurements to justify the appropriateness of the pooling.*

We agree that these differences between studies and over time could influence the reproducibility of DXA measurements for use in individual patients but in the context of this application, where we are focused on mean differences comparing large numbers of patients between treatment groups, we believe the large numbers of patients and comparison between treatments attenuates the impact of these factors for this meta-regression analysis. In addition, the fact that the data in these trials were performed in many, often 100's of centers, it is challenging to retrospectively sort out differences between trials. However we look forward to discussing this further with FDA and determining analyses that might address these issues.

11. *We suggest that you consider splitting the data into training and testing groups upfront so that the testing data can be used to validate the results.*

We would look forward to discussing these analyses with FDA and including them in our analysis plan.

C. *General Considerations:*

12. Provide pdfs of the publication articles mentioned in the LOI.

We will provide PDF's with our analysis plan that we will develop in conjunction with FDA as well as with our final briefing packet.

D. Process-related Questions from the Letter of Intent Submission and FDA Response to Questions:

Question 1. We might also want to consider if there is any context under which changes in hip BMD could be used as evidence of anti-fracture efficacy without concurrent or prior evidence of the drug's efficacy in reducing vertebral fracture risk.

FDA response: It is reasonable to consider whether change in BMD may be acceptable as evidence of anti-fracture efficacy without prior or concomitant direct demonstration of effect on vertebral fracture. In this case, we would expect that change in hip BMD would also be associated with fracture risk reduction at other sites.

Question 2. It would be of interest to incorporate other variables that would support BMD's qualification, such as changes in serum markers of bone turnover or estimated bone strength derived from finite element modeling of bone strength in QCT scans. We are interested in exploring whether some combination of biomarkers could be more informative than DXA BMD alone.

FDA response: Explorations with serum bone turnover markers and finite element modeling of bone strength may be informative. There are limitations in the amount of data available for these analyses however.

Question 3. In this Letter, for the COU we have assumed that the results would apply only to post-menopausal women. However, there are other groups of interest. For example, could the COU be broader to include men or people with other specific conditions or medication use (e.g., corticosteroids)?

FDA response: We agree that these results would apply to postmenopausal women. There is evidence that bone loss due to corticosteroid use is different than that seen in postmenopausal women. Therefore, expanding the context of use to this population may be difficult. For a context of use in men, we have not required fracture efficacy data in this population, and therefore, there are limited data to support a context of use in men.

Question 4. Would it be useful to consider only total hip BMD or perhaps changes in other BMD sites including femoral neck BMD or even lumbar spine?

FDA response: It may be useful to consider other BMD sites. It is recognized that evidence of bone changes may be seen more rapidly in the lumbar spine, so the timing of BMD measurements will be important.

Question 5. If we are going to rely on DXA BMD as surrogate measure, is percentage change in DXA the most reliable metric or should be consider achievement of a specific level of BMD (a "target") a better surrogate?

FDA response: We believe that percent change in BMD would be the appropriate measure, as this metric would not be affected by the DXA machine type or normative database used. However, there may be a threshold of percent change in BMD, above which we are comfortable with stating fracture risk reduction has been demonstrated. If the data show that a threshold BMD is a better correlate, this could be considered.

Question 6. In our analysis plan for the individual data, we plan to include percent of treatment explained (PTE). We would welcome discussion with the FDA about the value of PTE that might

be considered adequate for this analysis and whether there are alternative statistical approaches to PTE.

FDA response: PTE could be considered as an exploratory measure, but details on this analysis approach need to be pre-specified in the planning if you choose to report this measure. We welcome further discussion if you have alternative proposals.

E. References for your consideration:

FDA background document on the benefits and risks of long-term bisphosphonate use for the treatment and prevention of osteoporosis for Meeting of Advisory Committee for Reproductive Health Drugs and Drug Safety and Risk Management Advisory Committee

http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/DrugSafetyandRiskManagementAdvisoryCommittee/UCM270958.pdf

APPENDIX 1: Table 3 (from BAA): Full List of Studies in UCSF-FNIH BQ database

Current data received at UCSF as of August 1, 2018

Study Name (year*)	Drug/sponsor	# of Pts	# with BMD	Fractures			Ctrl^
				Vert	Non-Vert	Hip	
Phase III ALN (1995)	Alendronate/ Merck	994	968	-	68	4	Pbo
FIT (1996/1998)	Alendronate/ Merck	6,459	6,459	344	739	74	Pbo
FOSIT (1999)	Alendronate/ Merck	1,908	1,858	-	52	4	Pbo
MEN'S STUDY (2000)	Alendronate/ Merck	241	241	11	-	-	Pbo
GENERATIONS (2009)	Arzoxifene/ Eli Lilly	9,354	9,339	294	687	46	Pbo
Phase III Baze (2008)	Bazedoxifene/ Pfizer	7,492	5,643	141	258	18	Pbo
Clodronate (2007)	Clodronate/ MRC&Schering	5,579	692	-	-	114	Pbo
FREEDOM (2009)	Denosumab/ Amgen	7,808	7,806	350	556	69	Pbo
WHI-E (2006)	HRT/ NIH	10,739	938	-	1,331	139	Pbo
WHI-E+P (2013)	HRT/ NIH	16,608	1,027	-	2,113	241	Pbo
BONE (2004)	Ibandronate/ Roche	2,929	2,924	167	229	21	Pbo
IV Ibandronate (2004)	Ibandronate/ Roche	2,860	2,855	274	243	26	Pbo
PEARL (2010)	Lasofoxifene/ Sermonix	8,556	8,550	607	760	90	Pbo
IMAGING ODN (2013)	Odanacatib/ Merck	109	95	4	12	2	Pbo
LOFT	Odanacatib/ Merck	16,071	15,720	891	1,084	216	Pbo
TOP (2007)	PTH(1-84)/ Shire	2,532	2,532	59	79	6	Pbo
MORE (1999)	Raloxifene/ Eli Lilly	7,705	7,697	503	677	58	Pbo

Study Name (year*)	Drug/sponsor	# of Pts	# with BMD	Fractures			Ctrl^
				Vert	Non-Vert	Hip	
HIP (2001)	Risedronate/ Actavis	9,331	1,735	497	913	205	Pbo
VERT-MN (2000)	Risedronate/ Actavis	814	799	166	100	17	Pbo
VERT-NA (1999)	Risedronate/ Actavis	1,628	1,615	180	157	15	Pbo
Fx Prevention Trial (2001)	Teriparatide/ Eli Lilly	1,637	1,587	105	119	9	Pbo
HORIZON PFT (2007)	Zoledronic Acid/ Novartis	7,736	7,699	535	679	140	Pbo
HORIZON RFT (2007)	Zoledronic Acid/ Novartis	2,127	2,025	-	186	56	Pbo
EFFECT - Intl (2004)	Alendronate/ Merck	487	487	0	14	1	Act
EFFECT - USA (2004)	Alendronate/ Merck	456	372	-	-	-	Act
Weekly ALN (2000)	Alendronate/ Merck	1,258	1,044	-	81	-	Act
FACT (2005)	ALN & RIS/ Merck	1,053	892	-	66	-	Act
BAZE+CEE (2014)	Baze & HRT/ Pfizer	1,742	-	-	-	-	Act
BAZE + RAL	Baze & RAL/ Pfizer	1,886	513	-	-	-	Act
CORAL	Lasofoxifene/ Sermonix	540	454	-	3	-	Act
Ibandronate Quality (2009)	Ibandronate/ Glaxo Smith Kline	93	93	-	-	-	Act
DIVA (2006)	Ibandronate/ Roche	1,395	1,095	-	43	0	Act
MOBILE	Ibandronate/ Roche	1,609	-	-	-	-	Act
JADE	Lasofoxifene/ Sermonix	497	497	-	-	-	Act
LACE (2009)	Lasofoxifene/ Sermonix	51	44	1	-	2	Act

Study Name (year*)	Drug/sponsor	# of Pts	# with BMD	Fractures			Ctrl^
				Vert	Non-Vert	Hip	
OPAL3	Lasofoxifene/ Sermonix	924	905	-	-	-	Act
OPAL4	Lasofoxifene/ Sermonix	983	977	-	-	-	Act
Phase II Studies	Lasofoxifene/ Sermonix	1,315	1,012	1	-	-	Act
PaTH (2003)	PTH(1-84) & Aln/ NIH (NIAMS)	238	238	-	17	1	Act
CAP	PTH(1-84)/ Shire	375	303	-	2	-	Act
TRES (2007)	PTH(1-84)/ Shire	91	61	6	7	0	Act
FACT (ALN + TPTD) (2005)	TPTD & ALN/ Eli Lilly	203	203	-	-	-	Act
Extension Studies (5)	Various drugs & sponsors	3,707	353	-	-	-	Act
Data Not Yet Received at UCSF							
Phase II (2002)	Zoledronic Acid/ Novartis	351	-	-	5	5	Pbo
ZOL in Men (2012)	Zoledronic Acid/ Novartis	1,199	-	37	-	-	Pbo
IMPACT (2007)	Risedronate/ Actavis	2,302	2,001	27	46	-	Act
EUROFORS (2008)	TPTD/ Eli Lilly	503	467	4	8	1	Act
ZOL/ALN in Men (2010)	Zol & Aln/ Novartis	302	301	10	-	-	Act
ZOL/PTH comb. (2011)	Zol & TPTD/ Novartis	412	-	-	-	-	Act
Total		**155,189**	**103,116**	**5,214**	**11,334**	**1,580**	

*Year of major study publication; ^ Pbo=Placebo Control; Act=Active Control

APPENDIX 2:

Change in bone density and reduction in fracture risk: a meta-regression of published trials

Mary L. Bouxsein[1], Richard Eastell[2], Li-Yung Lui[3], Lucy A. Wu[4], Anne E. de Papp[5], Andreas Grauer[6], Fernando Marin[7], Jane A. Cauley[8], Lisa Palermo[4], Douglas C. Bauer[9], Dennis M. Black[4] for the FNIH Bone Quality Project

[1] Center for Advanced Orthopedic Studies, Department of Orthopedic Surgery, Beth Israel Deaconess Medical Center and Harvard Medical School

[2] Academic Unit of Bone Metabolism, University of Sheffield, Sheffield, UK

[3] California Pacific Medical Center, San Francisco, CA

[4] Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA

[5] Merck & Co., Inc., Kenilworth, NJ USA

[6] Amgen Inc., Thousand Oaks, CA

[7] Eli Lilly and Company, Lilly Research Centre, Windlesham, United Kingdom

[8] Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

[9] Departments of Medicine and Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA, USA

Changes to the manuscript text are highlighted in red text.

Abstract:

Meta-analyses conducted >15 years ago reported that improvements in BMD were associated with reduction in vertebral and non-vertebral fractures in osteoporosis trials. Numerous studies have been conducted since then, incorporating new therapies with different mechanisms of action and enrolling many more subjects. To extend these prior analyses and provide support for the use of BMD as a surrogate endpoint in osteoporosis clinical trials, we conducted a meta-regression of 38 placebo-controlled trials of 19 therapeutic agents to determine the association between improvements in BMD and reductions in fracture risk. We used a linear model to examine the relationship between the mean difference in %BMD change from baseline (at 24 months) between treatment and placebo groups and the logarithm of the relative risk. We found that greater improvements in BMD were strongly associated with greater reductions in vertebral and hip fractures, but not non-vertebral fractures. For example, the change in BMD was strongly associated with reduction in vertebral fracture ($r2 = 0.56$, 0.54 and 0.63 for total hip, femoral neck and lumbar spine BMD, respectively, $p \leq 0.0002$ for all). Accordingly, the analyses would predict that a 2% or 6% difference in total hip BMD would lead to a 28 or 66% reduction, respectively, in vertebral fracture risk. For hip fracture, the r-square values for total hip, femoral neck and lumbar spine BMD difference were 0.48 ($p=0.01$), 0.42 ($p=0.02$) and 0.22 (ns), respectively. For a 2% or 6% improvement in total hip BMD, we might expect a 16 or 40% reduction in hip fracture risk. In conclusion, our results extend prior observations that larger improvements in DXA-based BMD are associated with greater reductions in fracture risk, particularly for vertebral and hip fractures. While these results cannot be directly applied to predict the treatment benefit in an individual patient, they provide compelling evidence that improvements in BMD with osteoporosis therapies may be useful surrogate endpoints for fracture in trials of new therapeutic agents.

Keywords: bone mineral density, clinical trial, meta-regression, osteoporosis, surrogate

Introduction

Currently available osteoporosis treatments markedly reduce vertebral fracture risk, yet the effects of these treatments on hip and non-vertebral fracture risk are less pronounced. It is therefore critically important to develop new interventions that can reduce hip and non-vertebral fracture risk, given the profound personal and societal costs associated with these fractures. However, due to the relatively low rate of hip fractures, randomized trials to assess the efficacy of new treatments on hip fracture risk have become large, long in duration and prohibitively expensive. The enormity of effort and expense required to develop new interventions may inhibit innovation in the treatment of osteoporosis. Accordingly, to facilitate future drug development, it would be valuable to have measures that could be used as surrogate endpoints in place of a fracture outcomes to yield smaller and shorter randomized controlled trials.[1]

A possible surrogate endpoint for fractures in osteoporosis trials is areal bone mineral density (BMD) as measured by dual-energy X-ray absorptiometry (DXA). Indeed, the potential value of BMD to predict fracture reductions during clinical trials was reported over 10 years ago in two meta-regression analyses based on published trials of anti-resorptive agents, one for vertebral fractures[2] and the other for non-vertebral fractures.[3] The study of spine BMD changes and vertebral fracture risk examined 12 trials that enrolled ~21,000 women, whereas the study examining non-vertebral fractures examined 18 trials that enrolled 26,494 women. Notably, neither of these studies addressed the relationship between BMD changes and reduction in hip fracture. Furthermore, since these studies were published, many more randomized trials have been conducted, enrolling an additional 80,000 subjects and investigating additional therapeutic compounds with differing mechanisms of action not included in the prior analyses, including zoledronate[4], ibandronate[5], denosumab[6], lasofoxifene[7], bazedoxifene[8], arzoxifene[9], tibolone[10], odanacatib[11], teriparatide (PTH(1-34))[12-14], PTH(1-84)[15], romozosumab[16,17] and abaloparatide[18], as well as hormone therapy.[19]

We initiated the FNIH Bone Quality Study to investigate possible surrogate endpoints for enhancing drug development in osteoporosis. The FDA has a set of guidelines for how to qualify a biomarker for such purposes, and their recommendations include both a meta-regression of published trials and analyses of patient-level data[20]. The current analysis was undertaken to meet that first recommendation. Thus, we conducted a new meta-analysis of published data from placebo-controlled trials to determine the association between changes in BMD and reductions in vertebral, hip and non-vertebral fractures, incorporating data available from newer osteoporosis drugs with variable mechanisms of action.

Methods

Inclusion of studies

A systematic literature search of the PubMed, Cochrane, and EMBASE databases was initially completed by Cerner Corporation (Research Division, USA) in August 2008 to identify English-language randomized controlled trials (RCTs) that examined osteoporosis prevention and/or treatment. The search terms are provided in Supplementary Materials (**Table S1**). The search was updated in March 2011 and January 2017 in PubMed, by UCSF, using the previously used search terms and restricting to publications after August 29, 2008. We subsequently reviewed this

search to identify any study that met the following criteria: placebo-controlled randomized trial of osteoporosis medications, placebo group of at least 50 participants, BMD at baseline and follow up as assessed by DXA, and data for at least one of the 3 fracture endpoints (vertebral, non-vertebral and/or hip) available. If the number of fractures in both treatment groups combined was fewer than 5 for any of the 3 fracture outcomes, that specific fracture outcome was excluded for that study. Studies targeting specific medical conditions (e.g., rheumatoid arthritis) and treatments (e.g., corticosteroid users) were excluded, as were extension studies. Data from the eligible studies were extracted by one reviewer and verified by a second.

For trials with more than one dose, the active doses of the study medication were combined into one active arm, and BMD changes were averaged and fracture results pooled across treatment arms. For studies that included more than one active treatment arm or an arm with combinations of drugs, only the arm for the primary drug of interest was retained. Thus, in trials that contained another active treatment arm other than the primary drug of interest, we only analyzed the primary active treatment compared to the placebo. For any study in which a dose was discontinued before study end, the results for that dose were excluded.

Meta-regression approach

For the meta-regression, the primary predictor was the BMD difference between treatment and placebo groups at the lumbar spine (LS), femoral neck (FN) and total hip (TH), defined as the difference in percentage DXA BMD change between interventions (% change in active treatment minus % change in placebo) at the end of each study. Since reporting for the published trials varied, the methods used to perform this calculation varied. In some studies, the difference between treatments was available directly. In others, the % change within each treatment group was reported and we calculated the difference between treatment groups. We attempted to abstract data for the lumbar spine, femoral neck and total hip sites. In some studies, all three skeletal sites were available, while for others only one or two were reported. We used the BMD changes from the start to the end of the study, as reported in the published manuscripts.

We performed the analysis for three fracture types: vertebral, non-vertebral and hip fractures. For hip and non-vertebral fractures, we calculated the risk ratio (RR) for treatment using the number of fractures and number of subjects randomized. For vertebral fractures we calculated the RR using the number of fractures and number of subjects with follow-up radiographs. The RR was calculated using a Poisson distribution (Proc Genmod, SAS 9.4).

Fracture definitions varied for all fracture categories, but because our analysis was based on published data we relied on fracture definitions used within each study. For example, most, but not all studies, excluded fractures due to excessive trauma. While the definitions of "excessive trauma" varied across studies, we utilized the study definition. Further, for non-vertebral fractures, some studies included only a limited number of non-vertebral fracture types[21-23] and this was used in the analysis. One study reported hip and pelvis fractures together, which was used for hip fracture analysis[21] and two others reported only clinical vertebral fractures [19,24] which were used in the vertebral fracture analysis.

Statistical analysis

The goal of this analysis was to assess the relationship, across studies, between change in DXA BMD and fracture reduction. We plotted the risk ratio (RR) for the three primary types of fractures (vertebral, non-vertebral, and hip) against the net mean % change in lumbar spine, total hip and femoral neck BMD (treatment minus the placebo group). Each study was plotted as a circle with the area of the circle proportional to the inverse of the standard error of the RR.

Next, we then used a linear model to estimate the relationship between mean percentage difference in BMD (active minus placebo group) and the logarithm of the RR. The studies were weighted by the 1/(standard error of the log RR)^2, a weighting that is roughly proportional to the number of fractures in each study. Using the coefficients from this linear model, we calculated by exponentiation (inducing a non-linear relationship) the % change in BMD vs. the relative risk and plotted this curve. From the linear regressions we also calculated the r^2 with 95% confidence intervals and the statistical significance of the slopes. We used the meta-regression results to estimate the predicted fracture risk reduction across a range of BMD differences. That range was approximately from the lowest to the highest BMD difference among the publications included in the meta-regression: 2% to 6% for the total hip and femoral neck; and 2% to 14% for the lumbar spine.

The main analyses were limited to trials with $\geq$ 100 placebo participants and fracture outcomes with $\geq$ 10 events in the two treatment groups combined. However, we performed sensitivity analyses adding in studies with > 50 participants and number of fractures in both treatment groups combined was fewer than 10 but more than 5 for any of the 3 fracture outcomes. We also performed the following sensitivity analyses: 1) including studies with a duration of 2 or more years, 2) including studies using antiresorptive drugs only, and 3) including trials of bisphosphonates only.

Our approach will be useful only if our model, developed from placebo-controlled studies, will also predict reductions in the increasingly performed active controlled studies (as well as combination and sequential regimens). Thus, to test the robustness of our approach in studies with an active comparator, we applied the meta-regression results to two recently published trials that reported fracture outcomes and BMD, namely the ARCH trial of romosozumab versus alendronate[17] and ACTIVE-Extend trial of abaloparatide followed by alendronate versus placebo followed by alendronate[25]. We used the meta-regression results and published lumbar spine and hip BMD changes to predict the RR for vertebral, non-vertebral and hip fractures, and the standard error from the meta-regression to compute the 95% confidence interval for the RR.

Results

Characteristics of included trials

Thirty-eight randomized, placebo-controlled trials of 19 different therapies met our criteria for inclusion in the meta-analysis (**Table 1**). Our analyses included trials of six bisphosphonates (20 trials), four SERMs (5 trials), calcitonin, estrogen compounds (2 trials), tibolone (1 trial), anti-Rank Ligand antibody (2 trials), PTH (1-84) (1 trial), two PTH-analogs (4 trials), anti-sclerostin antibody

(1 trial), and a cathepsin K inhibitor (1 trial). Trial size ranged from 246 to over 16,000 participants and trial duration ranges from 1 to 8 years of follow-up. Most of the trials enrolled postmenopausal women only (n=32), 5 trials enrolled women and men, and 1 trial only enrolled men.

The change in femoral neck, total hip and lumbar spine BMD within active and placebo groups are shown in **Table 2**. Differences in BMD change between groups (active – placebo) ranged from 1.21% to 6.9% for the total hip, 2.1% to 5.9% for the femoral neck, and 0.73% to 13.3% for the lumbar spine (**Table 3**). The number of fractures in each trial included in the study ranged from 14 to 607, 20 to 232 and 12 to 934 for vertebral, hip and non-vertebral fractures, respectively. Risk ratios ranged from 0.14 to 1.03 for vertebral fractures, 0.47 to 1.12 for hip fracture and 0.37 to 1.58 for non-vertebral fractures.

Meta-regression results

Combining all eligible trials, the analyses of BMD change and vertebral fracture included up to 111,183 subjects and 4,557 fractures (**Table 4**). Results from the meta-analysis indicated that greater improvements in total hip, femoral neck BMD and lumbar spine BMD were all strongly associated with a greater reduction in vertebral fracture (r^2=0.56, r^2=0.54, and r^2=0.63, respectively, p≤0.0002 for all, **Table 4, Figure 1**). For example, our results revealed that a 2% improvement in total hip BMD was associated with a 28% reduction in vertebral fracture, whereas a 6% improvement in total hip BMD was associated with a 66% reduction in vertebral fracture (**Table 5**). In comparison, a 2% improvement in lumbar spine BMD was associated with a 28% reduction in vertebral fracture, whereas 8% and 14% improvements in lumbar spine BMD were associated 62% and 79% reductions in vertebral fracture, respectively.

The analyses of BMD change and hip fracture included trials with up to 94,469 subjects and 882 hip fractures. The results revealed that greater improvements in both total hip and femoral neck BMD were strongly associated with larger reductions in hip fracture risk (r^2 = 0.48, p=0.01 and r^2 = 0.42, p=0.02, respectively, **Table 4, Figure 2**). For example, a 2% improvement in total hip BMD was associated with a 16% reduction in hip fracture, whereas a 6% improvement in total hip BMD was associated with a 40% reduction in hip fracture, with similar results for femoral neck BMD (**Table 5**). In contrast, the association between improvements in lumbar spine BMD and hip fracture was weaker and did not reach statistical significance.

Finally, the analyses of BMD change and non-vertebral fractures included trials with up to 92,556 subjects and 6,383 fractures and showed that improvements in BMD at all three skeletal sites were weakly associated with the reduction in non-vertebral fracture (**Table 4, Figure 3**).

We conducted several sensitivity analyses. The associations between BMD change and fracture risk reduction were largely were similar when we restricted the analyses to trials with a duration of ≥2 years, with slightly stronger associations between hip BMD changes and reduction in vertebral fracture than when all trials were included **(Supplementary Table 2)**. In addition, the r^2 values were of similar magnitude when we restricted the analyses to anti-resorptive therapies

only **(Supplementary Table 3)** or to studies of bisphosphonates only **(Supplementary Table 4),** except for slightly higher associations between hip BMD and vertebral fractures.

Robustness of model predictions in active comparator trials
As the main goal of this project is to qualify hip BMD as a surrogate endpoint for hip and non-vertebral fractures for future osteoporosis trials, we focused on those two fracture types. Our analysis of two active comparator studies revealed that the predicted point estimate for relative risks (RR) for hip and non-vertebral fractures derived from the meta-regression were within the 95% confidence interval for the RR reported in these trials (**Supplementary Table 5**).

Discussion

Our overall goal in these analyses was to determine the extent to which change in DXA-based BMD might function as a surrogate for fracture risk reduction in future clinical osteoporosis trials. We found that change in BMD across all published randomized trials is strongly predictive of hip and vertebral fracture reduction. In particular, BMD changes at the total hip and femoral neck are similarly predictive of both hip and vertebral fractures. In contrast, lumbar spine BMD changes were predictive only of vertebral fracture risk. BMD changes at all skeletal sites were weakly associated with reductions in non-vertebral fracture in these analyses. Provided that preclinical studies demonstrate normal or improved bone quality and biomechanical properties following treatment, these results suggest that new drugs that can increase hip BMD substantially are almost certain to decrease risk of hip and vertebral fractures.

The association between BMD changes and vertebral fracture was generally stronger than that between BMD changes and hip and non-vertebral fracture. While falls may play a role in up to half of vertebral fractures[26,27], the weaker associations between BMD change and hip and no-vertebral fracture may be attributable, in part, to the prominent contribution of fall-related factors to these fractures.

We found that the association of changes in BMD with reductions in non-vertebral fractures did not achieve statistical significance for the two hip BMD sites (although p=0.05 for lumbar spine BMD). We can think of at least two reasons that these results were not statistically significant. First, the effect of treatment on the reduction in non-vertebral fractures, even for the most effective agents, is generally much weaker (20 to 25% reduction) compared to observed reductions in vertebral (40 to 70%) or hip fracture (40% to 50%) for effective agents.[28] Since the range of reductions (only 0 to about 25%) is so much smaller, the slope of the line relating BMD change to fracture relative risk is therefore not very steep compared to the other fracture types. Since the r^2 and statistical significance of the relationship is a function of steepness of the line as well as the scatter around the line, contributes to the lower significance of non-vertebral compared to vertebral and hip fractures. Secondly, definitions of non-vertebral fracture have varied greatly across studies. For example, in some studies[21-23], non-vertebral fractures included only 6 skeletal sites, while other studies had much broader definitions or definitions of non-vertebral fractures were not explicitly stated.[29,30] The criteria for exclusion of fractures due to trauma also varied greatly across studies. Taken together, this heterogeneity in non-vertebral

fracture definitions added scatter to the points around the regression line, attenuating the magnitude of relationship, and could help explain the lack of significance.

Two previous meta-analyses have examined the relationship between changes in BMD and fracture risk reduction. In 2002, Cummings and colleagues analyzed the relationship of spine BMD changes to vertebral fracture reduction.[2] They included 12 trials that included 21,404 women, and similar to our approach, examined changes in BMD over the entire study. Though we now include 30 trials in the analysis of lumbar spine BMD improvements and vertebral fracture, our results are remarkably consistent with what they reported. Hochberg et al[3], also in 2002, analyzed change in hip and spine BMD over 1 year and its relationship to reduction in non-vertebral fracture over the duration of each study in 18 studies with 26,494 women. They found a much stronger relationship between BMD improvement and non-vertebral fracture reduction than we did, and in contrast to our study, reported a significant association between improvements in hip BMD at the end of the study and reduction in non-vertebral fracture risk. These discrepant results may be due to the larger number of subjects included in the current analyses, which had nearly 4-fold more subjects than the prior study. Our approaches also differed in that we restricted our analyses to studies with at least 10 fracture events and combined results for studies with more than one active treatment arm.

We acknowledge that the percent change in BMD, as we used, depends on the baseline value of BMD, and thus, it would be ideal to also perform these analyses using the absolute change in BMD and/or adjust for the baseline BMD value. However, this was not possible for a variety of reasons. Most importantly, not all studies reported baseline BMD values (see **Table 1**). Furthermore, among the baseline values that were reported, some are in g/cm^2 while others are in t-score units. An additional problem, and the main reason that BMD changes are almost always reported as % changes, is that there are a variety of manufacturers and models of densitometers, and the absolute values across machines are not necessarily comparable.

This analysis has a number of significant limitations that must be considered when interpreting the results. None of the studies limited the patients to the "oldest old" and it is possible that the association between changes in BMD and reductions in fracture may differ in this group. Further, there are few studies in men, as well as few studies of anabolic agents and those that are available are of limited duration. Importantly, we pooled data from trials with study durations varying from one[16,31] to eight[24] years. The implicit assumption is that improvements in BMD over varying time frames have similar relationships to fracture risk. However, this assumption may not be true. For example, bisphosphonates strongly increase BMD over the first year of treatment with smaller decreases over the next 2 to 3 years and then a plateau.[32] Fracture reductions are generally not evident until after 18 or 24 months of treatment[33], inferring that a study that is only a single year in duration may have underestimated the eventual treatment effect. Alternately, teriparatide has little impact on hip BMD over the first 18 months, but still reduces non-vertebral fractures.[12] Another concern is that we correlated BMD change until the end of the study with fracture reductions. However, fractures occurring during a study may influence BMD changes due to inactivity, adherence or loss to follow-up and therefore could induce bias. However, earlier BMD changes were not available in most studies.

Another issue, as discussed above, is that fracture definitions vary across studies, particularly for non-vertebral fractures and with respect to exclusion of traumatic fractures. Furthermore, some trials recorded fractures as adverse events whereas in others, fractures were primary endpoints and were therefore confirmed via rigorous adjudication processes. Another limitation is that not all three BMD sites were included in all studies. Thus, a comparison of the relationship between BMD improvement and fracture reduction across all studies and fracture types is not possible and each analysis included a slightly different subset of the studies. As part of the Foundation for NIH's Bone Quality Project, we are in the process of collecting individual patient-level data from many of the randomized controlled trials of osteoporosis drugs[34]. Future analyses using these individual patient level data will allow us to address several of the limitations described above and inherent in working only with published data. For example, we will be able to use a consistent definition of non-vertebral fracture and a consistent follow-up duration across trials.

Nevertheless, our study also has a number of strengths. Notably, the current analyses include 38 trials with up to 111,000 enrolled subjects, more than double the number of trials and 4-fold the number of study subjects in the prior meta-analyses of BMD change and fracture risk reduction. Importantly, this is the first study to show that greater improvements in BMD are associated with larger reductions in hip fracture. This finding is particularly important, as placebo-controlled trials designed to show hip fracture efficacy with drug treatments are already large, and the required size of trials will expand even further as active-comparator trials are implemented. Finally, we were able to include therapies with diverse mechanisms of action, and although the small number of trials precluded any formal analyses, the association between BMD improvements and fracture reduction appeared consistent across most trials. This lends support to the idea that improvements in BMD may be useful in trials of new therapeutic agents with varied mechanisms of action.

Importantly, this type of analysis, developed from placebo-controlled studies, will be useful for future drug development only if it also predicts anti-fracture efficacy in actively controlled studies, as well as combination and sequential treatment regimens. We are only aware of two actively-controlled studies with BMD and fracture outcomes that we could use to test the robustness of the meta-regression model, namely the ARCH[17] and ACTIVE-Extend[25] trials. We used the meta-regression results to predict the anti-fracture efficacy based on the published BMD changes in these trials. This exercise showed that the predicted relative risks for non-vertebral and hip fracture from our models were similar to the observed relative risks, providing support for the robustness of this approach in the setting of active-controlled trials and sequential treatment regimens.

In summary, our results confirm and extend prior observations that larger improvements in DXA-based BMD are associated with greater reductions in fracture risk, particularly for vertebral and hip fractures. Notably our findings were valid across therapeutic agents with varied mechanisms of action. While these results cannot be directly applied to predict the treatment benefit in an individual patient, they provide compelling evidence that improvements in BMD with

osteoporosis therapies may be useful as surrogate endpoints for fracture in trials of new therapeutic agents.

Acknowledgements
The Bone Quality Project would like to thank the following groups for their contribution and support:

Scientific and financial support for the FNIH Bone Quality Project are made possible through direct contributions by: AgNovos Healthcare, American Society for Bone and Mineral Research, Amgen Inc., Daiichi Sankyo, Inc., Dairy Research Institute, Eli Lilly and Company, Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, and Roche Diagnostics Corporation. We thank Jean Hietpas, Charles McCulloch, Lisa Palermo, Lucy Wu, Gayle Lester, Sanya Fanous-Whitaker, and Steve Hoffmann for their leadership and expertise on the Project Team.

In-kind data to support the project was provided by Actavis, Amgen Inc., Bayer Schering Pharma Oy, Eli Lilly and Company, GlaxoSmithKline, Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA, National Institute of Arthritis and Musculoskeletal and Skin Diseases & National Heart, Lung, and Blood Institute, Novartis, Pfizer, Inc., Roche Diagnostics Corporation, and Sermonix.

Figure legends

Figure 1. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in vertebral fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.

Figure 2. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in hip fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.

Figure 3. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in non-vertebral fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.

References

1. Bouxsein ML, Delmas PD. Considerations for development of surrogate endpoints for antifracture efficacy of new treatments in osteoporosis: a perspective. J Bone Miner Res. Aug 2008;23(8):1155-67.
2. Cummings SR, Karpf DB, Harris F, Genant HK, Ensrud K, LaCroix AZ, et al. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. Am J Med. Mar 2002;112(4):281-9.
3. Hochberg MC, Greenspan S, Wasnich RD, Miller P, Thompson DE, Ross PD. Changes in bone density and turnover explain the reductions in incidence of nonvertebral fractures that occur during treatment with antiresorptive agents. J Clin Endocrinol Metab. Apr 2002;87(4):1586-92.
4. Black DM, Delmas PD, Eastell R, Reid IR, Boonen S, Cauley JA, et al. Once-yearly zoledronic acid for treatment of postmenopausal osteoporosis. N Engl J Med. May 3 2007;356(18):1809-22.
5. Chestnut III C, Skag A, Christiansen C, Recker R, Stakkestad J, Hoiseth A, et al. Effects of oral ibandronate administered daily or intermittently on fracture risk in postmenopausal osteoporosis. J Bone Min Res. 2004;19(8):1241-49.
6. Cummings SR, San Martin J, McClung MR, Siris ES, Eastell R, Reid IR, et al. Denosumab for prevention of fractures in postmenopausal women with osteoporosis. N Engl J Med. Aug 20 2009;361(8):756-65. Epub 2009/08/13.
7. Cummings SR, Ensrud K, Delmas PD, LaCroix AZ, Vukicevic S, Reid DM, et al. Lasofoxifene in postmenopausal women with osteoporosis. N Engl J Med. Feb 25 2010;362(8):686-96.
8. Silverman SL, Christiansen C, Genant HK, Vukicevic S, Zanchetta JR, de Villiers TJ, et al. Efficacy of bazedoxifene in reducing new vertebral fracture risk in postmenopausal women with osteoporosis: results from a 3-year, randomized, placebo-, and active-controlled clinical trial. J Bone Miner Res. Dec 2008;23(12):1923-34.
9. Cummings SR, McClung M, Reginster JY, Cox D, Mitlak B, Stock J, et al. Arzoxifene for prevention of fractures and invasive breast cancer in postmenopausal women. J Bone Miner Res. Feb 2011;26(2):397-404.
10. Cummings SR, Ettinger B, Delmas PD, Kenemans P, Stathopoulos V, Verweij P, et al. The effects of tibolone in older postmenopausal women. N Engl J Med. Aug 14 2008;359(7):697-708.
11. Bonnick S, De Villiers T, Odio A, Palacios S, Chapurlat R, DaSilva C, et al. Effects of odanacatib on BMD and safety in the treatment of osteoporosis in postmenopausal women previously treated with alendronate: a randomized placebo-controlled trial. J Clin Endocrinol Metab. Dec 2013;98(12):4727-35.
12. Neer RM, Arnaud CD, Zanchetta JR, Prince R, Gaich GA, Reginster JY, et al. Effect of parathyroid hormone (1-34) on fractures and bone mineral density in postmenopausal women with osteoporosis. N Engl J Med. May 10 2001;344(19):1434-41.
13. Nakamura T, Sugimoto T, Nakano T, Kishimoto H, Ito M, Fukunaga M, et al. Randomized Teriparatide [human parathyroid hormone (PTH) 1-34] Once-Weekly

Efficacy Research (TOWER) trial for examining the reduction in new vertebral fractures in subjects with primary osteoporosis and high fracture risk. The Journal of clinical endocrinology and metabolism. Sep 2012;97(9):3097-106. Epub 2012/06/23.
14. Fujita T, Fukunaga M, Itabashi A, Tsutani K, Nakamura T. Once-Weekly Injection of Low-Dose Teriparatide (28.2 mug) Reduced the Risk of Vertebral Fracture in Patients with Primary Osteoporosis. Calcified tissue international. Feb 2014;94(2):170-5. Epub 2013/08/22.
15. Greenspan SL, Bone HG, Ettinger MP, Hanley DA, Lindsay R, Zanchetta JR, et al. Effect of recombinant human parathyroid hormone (1-84) on vertebral fracture and bone mineral density in postmenopausal women with osteoporosis: a randomized trial. Ann Intern Med. Mar 6 2007;146(5):326-39.
16. Cosman F, Crittenden DB, Adachi JD, Binkley N, Czerwinski E, Ferrari S, et al. Romosozumab Treatment in Postmenopausal Women with Osteoporosis. N Engl J Med. Oct 20 2016;375(16):1532-43.
17. Saag KG, Petersen J, Brandi ML, Karaplis AC, Lorentzon M, Thomas T, et al. Romosozumab or Alendronate for Fracture Prevention in Women with Osteoporosis. N Engl J Med. Oct 12 2017;377(15):1417-27.
18. Miller PD, Hattersley G, Riis BJ, Williams GC, Lau E, Russo LA, et al. Effect of Abaloparatide vs Placebo on New Vertebral Fractures in Postmenopausal Women With Osteoporosis: A Randomized Clinical Trial. JAMA. Aug 16 2016;316(7):722-33.
19. Cauley JA, Robbins J, Chen Z, Cummings SR, Jackson RD, LaCroix AZ, et al. Effects of estrogen plus progestin on risk of fracture and bone mineral density: the Women's Health Initiative randomized trial. JAMA. Oct 01 2003;290(13):1729-38.
20. Amur S, LaVange L, Zineh I, Buckman-Garner S, Woodcock J. Biomarker Qualification: Toward a Multiple Stakeholder Framework for Biomarker Development, Regulatory Acceptance, and Utilization. Clin Pharmacol Ther. Jul 2015;98(1):34-46.
21. Harris ST, Watts NB, Genant HK, McKeever CD, Hangartner T, Keller M, et al. Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. Vertebral Efficacy With Risedronate Therapy (VERT) Study Group. Jama. Oct 13 1999;282(14):1344-52.
22. Reginster J, Minne HW, Sorensen OH, Hooper M, Roux C, Brandi ML, et al. Randomized trial of the effects of risedronate on vertebral fractures in women with established postmenopausal osteoporosis. Vertebral Efficacy with Risedronate Therapy (VERT) Study Group. Osteoporos Int. 2000;11(1):83-91.
23. McClung MR, Geusens P, Miller PD, Zippel H, Bensen WG, Roux C, et al. Effect of risedronate on the risk of hip fracture in elderly women. Hip Intervention Program Study Group. N Engl J Med. Feb 1 2001;344(5):333-40.
24. Jackson RD, Wactawski-Wende J, LaCroix AZ, Pettinger M, Yood RA, Watts NB, et al. Effects of conjugated equine estrogen on risk of fractures and BMD in postmenopausal women with hysterectomy: results from the women's health initiative randomized trial. J Bone Miner Res. Jun 2006;21(6):817-28.

25. Bone HG, Cosman F, Miller PD, Williams GC, Hattersley G, Hu MY, et al. ACTIVExtend: 24 Months of Alendronate after 18 Months of Abaloparatide or Placebo for Postmenopausal Osteoporosis. J Clin Endocrinol Metab. May 24 2018.
26. Freitas SS, Barrett-Connor E, Ensrud KE, Fink HA, Bauer DC, Cawthon PM, et al. Rate and circumstances of clinical vertebral fractures in older men. Osteoporos Int. May 2008;19(5):615-23. Epub 2007/11/27.
27. Cooper C, Atkinson EJ, Kotowicz M, O'Fallon WM, Melton LJ, 3rd. Secular trends in the incidence of postmenopausal vertebral fractures. Calcif Tissue Int. Aug 1992;51(2):100-4.
28. Black DM, Rosen CJ. Clinical Practice. Postmenopausal Osteoporosis. N Engl J Med. Jan 21 2016;374(3):254-62.
29. Hooper MJ, Ebeling PR, Roberts AP, Graham JJ, Nicholson GC, D'Emden M, et al. Risedronate prevents bone loss in early postmenopausal women: a prospective randomized, placebo-controlled trial. Climacteric. Sep 2005;8(3):251-62.
30. Fogelman I, Ribot C, Smith R, Ethgen D, Sod E, Reginster JY. Risedronate reverses bone loss in postmenopausal women with low bone mass: results from a multinational, double-blind, placebo-controlled trial. BMD-MN Study Group. J Clin Endocrinol Metab. May 2000;85(5):1895-900.
31. Pols HA, Felsenberg D, Hanley DA, Stepan J, Munoz-Torres M, Wilkin TJ, et al. Multinational, placebo-controlled, randomized trial of the effects of alendronate on bone density and fracture risk in postmenopausal women with low bone mass: results of the FOSIT study. Foxamax International Trial Study Group. Osteoporos Int. 1999;9(5):461-8.
32. Khosla S, Bilezikian JP, Dempster DW, Lewiecki EM, Miller PD, Neer RM, et al. Benefits and risks of bisphosphonate therapy for osteoporosis. J Clin Endocrinol Metab. Jul 2012;97(7):2272-82. Epub 2012/04/24.
33. Black DM, Thompson DE, Bauer DC, Ensrud K, Musliner T, Hochberg MC, et al. Fracture risk reduction with alendronate in women with osteoporosis: the Fracture Intervention Trial. FIT Research Group. J Clin Endocrinol Metab. 2000;85(11):4118-24.
34. Bauer DC, Black DM, Bouxsein ML, Lui LY, Cauley JA, de Papp AE, et al. Treatment-Related Changes in Bone Turnover and Fracture Risk Reduction in Clinical Trials of Anti-Resorptive Drugs: A Meta-Regression. J Bone Miner Res. Apr 2018;33(4):634-42.
35. Liberman UA, Weiss SR, Broll J, Minne HW, Quan H, Bell NH, et al. Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. The Alendronate Phase III Osteoporosis Treatment Study Group. N Engl J Med. Nov 30 1995;333(22):1437-43.
36. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, et al. Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. Lancet (London, England). Dec 7 1996;348(9041):1535-41. Epub 1996/12/07.
37. Cummings SR, Black DM, Thompson DE, Applegate WB, Barrett-Connor E, Musliner TA, et al. Effect of alendronate on risk of fracture in women with low bone density but without vertebral fractures: results from the Fracture Intervention Trial. Jama. Dec 23-30 1998;280(24):2077-82.

38. Hosking D, Chilvers CE, Christiansen C, Ravn P, Wasnich R, Ross P, et al. Prevention of bone loss with alendronate in postmenopausal women under 60 years of age. Early Postmenopausal Intervention Cohort Study Group. The New England journal of medicine. Feb 19 1998;338(8):485-92. Epub 1998/01/28.
39. Pols HA, Felsenberg D, Hanley DA, Stepan J, Munoz-Torres M, Wilkin TJ, et al. Multinational, placebo-controlled, randomized trial of the effects of alendronate on bone density and fracture risk in postmenopausal women with low bone mass: results of the FOSIT study. Fosamax International Trial Study Group. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 1999;9(5):461-8. Epub 1999/11/07.
40. Greenspan SL, Schneider DL, McClung MR, Miller PD, Schnitzer TJ, Bonin R, et al. Alendronate improves bone mineral density in elderly women with osteoporosis residing in long-term care facilities. A randomized, double-blind, placebo-controlled trial. Ann Intern Med. May 21 2002;136(10):742-6.
41. McCloskey E, Selby P, Davies M, Robinson J, Francis RM, Adams J, et al. Clodronate reduces vertebral fracture risk in women with postmenopausal or secondary osteoporosis: results of a double-blind, placebo-controlled 3-year study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. May 2004;19(5):728-36. Epub 2004/04/08.
42. McCloskey EV, Beneton M, Charlesworth D, Kayan K, deTakats D, Dey A, et al. Clodronate reduces the incidence of fractures in community-dwelling elderly women unselected for osteoporosis: results of a double-blind, placebo-controlled randomized study. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. Jan 2007;22(1):135-41. Epub 2006/10/18.
43. Watts NB, Harris ST, Genant HK, Wasnich RD, Miller PD, Jackson RD, et al. Intermittent cyclical etidronate treatment of postmenopausal osteoporosis. N Engl J Med. Jul 12 1990;323(2):73-9.
44. Chesnut CH, 3rd, Skag A, Christiansen C, Recker R, Stakkestad JA, Hoiseth A, et al. Effects of oral ibandronate administered daily or intermittently on fracture risk in postmenopausal osteoporosis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. Aug 2004;19(8):1241-9. Epub 2004/07/03.
45. Recker R, Stakkestad JA, Chesnut CH, 3rd, Christiansen C, Skag A, Hoiseth A, et al. Insufficiently dosed intravenous ibandronate injections are associated with suboptimal antifracture efficacy in postmenopausal osteoporosis. Bone. May 2004;34(5):890-9. Epub 2004/05/04.
46. Lyles KW, Colon-Emeric CS, Magaziner JS, Adachi JD, Pieper CF, Mautalen C, et al. Zoledronic acid and clinical fractures and mortality after hip fracture. N Engl J Med. Nov 1 2007;357(18):1799-809.
47. Boonen S, Reginster JY, Kaufman JM, Lippuner K, Zanchetta J, Langdahl B, et al. Fracture risk and zoledronic acid therapy in men with osteoporosis. N Engl J Med. Nov 2012;367(18):1714-23. Epub 2012/11/02.
48. Nakamura T, Fukunaga M, Nakano T, Kishimoto H, Ito M, Hagino H, et al. Efficacy and safety of once-yearly zoledronic acid in Japanese patients with primary

osteoporosis: two-year results from a randomized placebo-controlled double-blind study (ZOledroNate treatment in Efficacy to osteoporosis; ZONE study). Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. Jan 2017;28(1):389-98. Epub 2016/09/16.
49. Chesnut CH, 3rd, Silverman S, Andriano K, Genant H, Gimona A, Harris S, et al. A randomized trial of nasal spray salmon calcitonin in postmenopausal women with established osteoporosis: the prevent recurrence of osteoporotic fractures study. PROOF Study Group. Am J Med. Sep 2000;109(4):267-76.
50. Nakamura T, Matsumoto T, Sugimoto T, Hosoi T, Miki T, Gorai I, et al. Clinical Trials Express: fracture risk reduction with denosumab in Japanese postmenopausal women and men with osteoporosis: denosumab fracture intervention randomized placebo controlled trial (DIRECT). The Journal of clinical endocrinology and metabolism. Jul 2014;99(7):2599-607. Epub 2014/03/22.
51. Itabashi A, Yoh K, Chines AA, Miki T, Takada M, Sato H, et al. Effects of bazedoxifene on bone mineral density, bone turnover, and safety in postmenopausal Japanese women with osteoporosis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research. Mar 2011;26(3):519-29. Epub 2010/09/15.
52. Ettinger B, Black DM, Mitlak BH, Knickerbocker RK, Nickelsen T, Genant HK, et al. Reduction of vertebral fracture risk in postmenopausal women with osteoporosis treated with raloxifene: results from a 3-year randomized clinical trial. Multiple Outcomes of Raloxifene Evaluation (MORE) Investigators. Jama. Aug 18 1999;282(7):637-45. Epub 1999/10/12.
53. Bone HG, Downs RW, Jr., Tucci JR, Harris ST, Weinstein RS, Licata AA, et al. Dose-response relationships for alendronate treatment in osteoporotic elderly women. Alendronate Elderly Osteoporosis Study Centers. The Journal of clinical endocrinology and metabolism. Jan 1997;82(1):265-74. Epub 1997/01/01.
54. Bone HG, Greenspan SL, McKeever C, Bell N, Davidson M, Downs RW, et al. Alendronate and estrogen effects in postmenopausal women with low bone mineral density. Alendronate/Estrogen Study Group. The Journal of clinical endocrinology and metabolism. Feb 2000;85(2):720-6. Epub 2000/02/26.
55. Orwoll E, Ettinger M, Weiss S, Miller P, Kendler D, Graham J, et al. Alendronate for the treatment of osteoporosis in men. N Engl J Med. Aug 31 2000;343(9):604-10.
56. Yan Y, Wang W, Zhu H, Li M, Liu J, Luo B, et al. The efficacy and tolerability of once-weekly alendronate 70 mg on bone mineral density and bone turnover markers in postmenopausal Chinese women with osteoporosis. Journal of bone and mineral metabolism. 2009;27(4):471-8. Epub 2009/04/04.
57. Reid IR, Brown JP, Burckhardt P, Horowitz Z, Richardson P, Trechsel U, et al. Intravenous zoledronic acid in postmenopausal women with low bone mineral density. N Engl J Med. Feb 28 2002;346(9):653-61.
58. Greenspan SL, Perera S, Ferchak MA, Nace DA, Resnick NM. Efficacy and safety of single-dose zoledronic acid for osteoporosis in frail elderly women: a randomized clinical trial. JAMA internal medicine. Jun 2015;175(6):913-21. Epub 2015/04/14.

59. Bone HG, Bolognese MA, Yuen CK, Kendler DL, Wang H, Liu Y, et al. Effects of denosumab on bone mineral density and bone turnover in postmenopausal women. J Clin Endocrinol Metab. Jun 2008;93(6):2149-57. Epub 2008/04/03.
60. Morii H, Ohashi Y, Taketani Y, Fukunaga M, Nakamura T, Itabashi A, et al. Effect of raloxifene on bone mineral density and biochemical markers of bone turnover in Japanese postmenopausal women with osteoporosis: results from a randomized placebo-controlled trial. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. Oct 2003;14(10):793-800. Epub 2003/09/05.
61. Reid IR, Eastell R, Fogelman I, Adachi JD, Rosen A, Netelenbos C, et al. A comparison of the effects of raloxifene and conjugated equine estrogen on bone and lipids in healthy postmenopausal women. Archives of internal medicine. Apr 26 2004;164(8):871-9. Epub 2004/04/28.

Table 1. Characteristics of placebo-controlled studies included in the meta-regression analyses

Author (Year)	Drug	Dose	Unit	Study sample size	Age (yrs) mean (range)	Gender (% women)	Placebo Baseline BMD T-score		
							TH	FN	LS
Anti-sclerostin Antibody									
Cosman (2016) [16]	Romosozumab	210	mg/mo	7180	70.9 (55-90)	100	-2.46	-2.74	-2.71
Bisphosphonate									
Liberman (1995) [35]	Alendronate	5, 10, 20 then 5*	mg/d	994	64.0 (45-80)	100	**	**	**
Black (1996) [36]	Alendronate	5 then 10	mg/d	2027	70.8 (55-81)	100	--	-2.48^	-2.63^
Cummings (1998) [37]	Alendronate	5 then 10	mg/d	4432	67.7 (54-81)	100	--	-2.21^	-2.15^
Hosking (1998) [38]	Alendronate	2.5, 5*,#	mg/d	1499	53.3 (45-59)	100	-0.75^	--	-0.97^
Pols (1999) [39]	Alendronate	10	mg/d	1908	62.8 (≤85)	100	-1.74^	**	**
Greenspan (2002) [40]	Alendronate	10	mg/d	327	78.5 (65-91)	100	--	--	--
McCloskey (2004) [41]	Clodronate	800	mg/d	593	67.5 (--)	100	-2.46^	--	-3.30^
McCloskey (2007) [42]	Clodronate	800	mg/d	5592	79.6 (≥75)	100	-1.57^	--	--
Watts (1990) [43]	Etidronate	400	mg/d	423	65.1 (≤75)	100	--	**	**
Chesnut (2004) [44]	Ibandronate	2.5, 20itmt*	mg/d	2946	69.0 (55-80)	100	-1.70	-2.00	-2.80
Recker (2004) [45]	Ibandronate	0.5, 1*	mg/3 mo	2862	67.0 (55-76)	100	-1.83	--	-2.79
Harris (1999) [21]	Risedronate	5#	mg/d	1641	68.5 (≤85)	100	--	-2.60	-2.40

Fogelman (2000) [30]	Risedronate	5#	mg/d	357	64.5 (≤80)	100	--	-1.85^	-2.91
Reginster (2000) [22]	Risedronate	5#	mg/d	816	71.0 (≤85)	100	--	**	-2.77
McClung (2001) [23]	Risedronate	2.5, 5*	mg/d	9331	77.7 (70-100)	100	--	-3.70 (S)	--
Hooper (2005) [29]	Risedronate	2.5, 5*	mg/d	383	52.7 (42-63)	100	--	-0.65^	-0.43
Black (2007) [4]	Zoledronic acid	5	mg/y	7765	73.1 (65-89)	100	-2.39^	-2.73^	-2.34^ (S)
Lyles (2007) [46]	Zoledronic acid	5	mg/y	2127	74.5 (≥50)	76.1	-1.98^	-1.73^	--
Boonen (2012) [47]	Zoledronic acid	5	mg/y	1199	-- (50-85)	0	-1.72	-2.24	--
Nakamura (2017) [48]	Zoledronic acid	5	mg/y	665	74.2 (65-89)	93.9	-2.20 (S)	-2.94 (S)	-2.97 (S)
Calcitonin									
Chesnut (2000) [49]	Calcitonin	100, 200, 400*	IU/d	1255	68.3 (--)	100	--	--	-1.79^
Cathepsin K Inhibitor									
Bonnick (2013) [11]	Odanacatib	50	mg/wk	246	71.3 (≥60)	100	-2.05	-2.38	-2.49
Estrogen Compounds									
Cauley (2003) [19]	Estrogen	0.625	mg/d	16608	63.2 (50-79)	100	-0.91 (S)	--	-1.26 (S)
Jackson (2006) [24]	Estrogen+Progestin	0.625	mg/d	10739	63.6 (50-79)	100	-0.81 (S)	--	-1.13 (S)
Cummings (2008) [10]	Tibolone	1.25	mg/d	4538	68.3 (60-85)	100	-1.80	--	-2.90
PTH(1-84) and PTH Analogs									

Miller (2016) [18]	Abaloparatide	80#	*u*g/d	1645	68.8 (49-86)	100	-1.90	-2.20	-2.90
Neer (2001) [12]	PTH 1-34	20, 40*	*u*g/d	1637	69.5 (--)	100	-1.90^	-1.82^	-2.60
Nakamura (2012) [13]	PTH 1-34	56.5	*u*g/wk	578	75.3 (65-95)	96.0	-2.10 (S)	-2.40 (S)	-2.60 (S)
Fujita (2013) [14]	PTH 1-34	28.2	*u*g/wk	316	71.5 (--)	94.9	--	--	-3.02
Greenspan (2007) [15]	PTH 1-84	100	*u*g/d	2532	64.4 (≥45)	100	-1.89	-2.21	-2.96
RANKL Inhibitor									
Cummings (2009) [6]	Denosumab	60	mg/6 mo	7808	72.3 (60-90)	100	-1.91	-2.17	-2.84
Nakamura (2014) [50]	Denosumab	60#	mg/6 mo	956	69.4 (≥50)	95.1	-1.95	-2.29	-2.73
SERM									
Cummings (2011) [9]	Arzoxifene	20	mg/d	9354	67.5 (60-85)	100	-1.37	-1.86	-2.22
Silverman (2008) [8]	Bazedoxifene	20, 40*,#	mg/d	5643	66.4 (55-85)	100	--	-1.80	-2.40
Itabashi (2011) [51]	Bazedoxifene	20, 40*	mg/d	429	63.4 (≤85)	100	-2.31^	-2.48^	-3.79^
Cummings (2010) [7]	Lasofoxifene	0.25, 0.5*	mg/d	8556	67.4 (59-80)	100	--	-2.20	-3.00
Ettinger (1999) [52]	Raloxifene	60, 120*	mg/d	7705	66.1 (31-80)	100	**	**	**
Studies not included in the main analysis&									
Bone (1997) [53]	Alendronate	1, 2.5, 5*	mg/d	359	70.7 (60-85)	100	--	--	**
Bone (2000) [54]	Alendronate	10#	mg/d	142	61.4 (42-82)	100	--	--	-2.50
Orwoll (2000) [55]	Alendronate	10	mg/d	241	63.0 (31-87)	0	-2.10	-2.30	-2.10

Yan (2009) [56]	Alendronate	70	mg/wk	560	64.9 (≤85)	100	**	**	**
Reid (2002) [57]	Zoledronic acid	0.25, 0.5, 1, 2, 4*	Various	351	64.1 (45-80)	100	-1.90^	--	-2.79^
Greenspan (2015) [58]	Zoledronic acid	5	mg/y	181	85.5 (≥65)	100	-2.00	-2.10	-0.60
Bone (2008) [59]	Denosumab	60	mg/6 mo	332	59.4 (--)	100	--	--	-1.66
Morii (2003) [60]	Raloxifene	60, 120*	mg/d	284	64.7 (≤80)	100	--	--	-3.99^
Reid (2004) [61]	Raloxifene	60, 150*,#	mg/d	461	52.9 (40-60)	100	-0.51^	--	-0.70^

Dashes indicate that no data were available

PTH = Parathyroid hormone; SERM = Selective estrogen receptor modulator

TH = Total hip; FN = Femoral neck; LS = Lumbar spine

*Active dosage groups were combined for analysis

** DXA machines other than Hologic were used. No standardized calculation to Hologic was available

Additional group of dosage/drug type not included in the analysis

^ Calculated from reported mean BMD using Looker 98 NHANES non-Hispanic White women reference

(S): BMD measured in subset only

& Studies with >= 5 but <10 fractures or with >=50 but <100 ppts in placebo group

Table 2. BMD changes in studies included in the meta-regression

Author (Year)	Drug	Study duration (yrs)	Δ TH BMD (%)		Δ FN BMD (%)		Δ LS BMD (%)	
			Placebo	Active	Placebo	Active	Placebo	Active
Anti-sclerostin Antibody								
Cosman (2016) [16]	Romosozumab	1	0.00	6.80	-0.70	5.20	0.00	13.30
Bisphosphonate								
Liberman (1995) [35]	Alendronate	3	NA	NA	-1.30	3.53	-0.80	6.83
Black (1996) [36]	Alendronate	3	-1.50	3.20	-0.40	3.70	1.80	8.00
Cummings (1998) [37]	Alendronate	4	-1.60	3.40	-0.80	3.80	1.50	8.30
Hosking (1998) [38]	Alendronate	2	-1.40	1.45	NA	NA	-1.80	2.90
Pols (1999) [39]	Alendronate	1	0.10	3.10	-0.20	2.30	0.10	5.00
Greenspan (2002) [40]	Alendronate	2	NA	NA	-0.10	3.30	2.00	6.40
McCloskey (2004) [41]	Clodronate	3	-3.03	0.70	NA	NA	0.64	4.35
McCloskey (2007) [42]	Clodronate	3	-4.10	-1.60	NA	NA	NA	NA
Watts (1990) [43]	Etidronate	2	NA	NA	-0.25	2.65	1.30	4.70
Chestnut (2004) [44]	Ibandronate	3	-0.70	3.15	-0.60	2.60	1.30	6.10
Recker (2004) [45]	Ibandronate	3	-1.30	1.70	-1.10	1.00	0.95	4.40
Harris (1999) [21]	Risedronate	3	NA	NA	-1.20	1.60	1.10	5.40
Fogelman (2000) [30]	Risedronate	2	NA	NA	-1.00	1.30	0.00	4.10
Reginster (2000) [22]	Risedronate	3	NA	NA	-1.10	2.00	1.10	7.00

McClung (2001) [23]	Risedronate	3	NA	NA	NR	NR	NA	NA
Hooper (2005) [29]	Risedronate	2	NA	NA	-2.50	0.28	-2.50	1.13
Black (2007) [4]	Zoledronic acid	3	-2.00	4.02	-1.00	4.06	0.20	6.90
Lyles (2007) [46]	Zoledronic acid	5	-0.90	5.50	-0.70	3.60	NA	NA
Boonen (2012) [47]	Zoledronic acid	2	0.20	2.30	0.10	3.40	1.60	7.70
Nakamura (2017) [48]	Zoledronic acid	2	-0.70	3.30	-0.50	3.50	0.30	7.90
Calcitonin								
Chestnut (2000) [49]	Calcitonin	5	NA	NA	NA	NA	0.50	1.23
Cathepsin K Inhibitor								
Bonnick (2013) [11]	Odanacatib	2	-1.90	0.80	-0.90	1.70	-0.30	2.30
Estrogen Compounds								
Cauley (2003) [19]	Estrogen	6	0.14	3.70	NA	NA	1.50	6.00
Jackson (2006) [24]	Estrogen+Progestin	8	-1.90	1.80	NA	NA	1.90	7.10
Cummings (2008) [10]	Tibolone	3	NA	NA	-1.35	1.75	1.40	6.20
PTH(1-84) and PTH Analogs								
Miller (2016) [18]	Abaloparatide	1.5	-0.10	4.18	-0.43	3.60	0.63	11.20
Neer (2001) [12]	PTH 1-34	1.8	-1.00	3.10	-0.70	3.95	1.10	11.70
Nakamura (2012) [13]	PTH 1-34	1.4	0.10	3.10	-0.50	1.80	0.30	6.70
Fujita (2013) [14]	PTH 1-34	1.5	NA	NA	NA	NA	0.50	4.40
Greenspan (2007) [15]	PTH 1-84	1.4	-1.09	1.02	-0.69	1.78	-0.32	6.53
RANKL Inhibitor								
Cummings (2009) [6]	Denosumab	3	-1.00	5.00	NA	NA	0.05	9.25

Nakamura (2014) (50)	Denosumab	2	-1.10	4.60	-1.10	4.00	0.10	9.10
SERM								
Cummings (2011) (9)	Arzoxifene	3	-1.20	1.40	-1.20	1.70	0.90	3.70
Silverman (2008) (8)	Bazedoxifene	3	-0.83	0.39	NA	NA	0.88	2.30
Itabashi (2011) (51)	Bazedoxifene	2	-0.97	1.02	-1.01	1.45	-0.65	2.59
Cummings (2010) (7)	Lasofoxifene	5	NR	NR	NR	NR	NR	NR
Ettinger (1999) (52)	Raloxifene	3	NA	NA	-1.20	1.05	0.50	3.15

PTH = Parathyroid hormone; SERM = Selective estrogen receptor modulator
TH = Total hip; FN = Femoral neck; LS = Lumbar spine
NA = Not available; NR = Not report

Table 3. BMD differences and fracture risk reduction values used in the meta-regression

Author	Drug	Difference in BMD, active-placebo (%)			Fracture reduction RR (N fractures)		
		TH	FN	LS	VFx	Hip Fx	NVFx
Anti-sclerostin Antibody							
Cosman (2016) [16]	Romosozumab	6.90	5.90	13.30	0.27 (75)	0.54 (20)	0.75 (131)
Bisphosphonate							
Liberman (1995) [35]	Alendronate	--	4.83	7.63	0.52 (39)	--	0.79 (83)
Black (1996) [36]	Alendronate	4.70	4.10	6.20	0.53 (223)	0.49 (33)	0.81 (270)
Cummings (1998) [37]	Alendronate	5.00	4.60	6.80	0.56 (121)	0.79 (43)	0.89 (555)
Hosking (1998) [38]	Alendronate	2.85	--	4.70	--	--	1.58 (58)
Pols (1999) [39]	Alendronate	3.00	2.50	4.90	--	**	0.52 (56)
Greenspan (2002) [40]	Alendronate	--	3.40	4.40	--	**	0.73 (31)
McCloskey (2004) [41]	Clodronate	3.73	--	3.71	0.54 (96)	**	0.69 (35)
McCloskey (2007) [42]	Clodronate	2.50	--	--	--	0.96 (114)	--
Watts (1990) [43]	Etidronate	--	2.90	3.40	0.44 (25)	**	1.21 (62)
Chestnut (2004) [44]	Ibandronate	3.85	3.20	4.80	0.52 (149)	--	1.10 (256)
Recker (2004) [45]	Ibandronate	2.95	2.10	3.45	0.84 (251)	0.60 (24)	--
Harris (1999) [21]	Risedronate	--	2.80	4.30	0.64 (154)	0.80 (27)	0.64 (85)
Fogelman (2000) [30]	Risedronate	--	2.30	4.10	0.53 (25)	--	0.55 (20)
Reginster (2000) [22]	Risedronate	--	3.10	5.90	0.60 (142)	0.82 (20)	0.71 (87)
McClung (2001) [23]	Risedronate	--	2.75	--	--	0.73 (232)	0.84 (934)

Hooper (2005) [29]	Risedronate	--	2.78	3.63	1.03 (31)	--	0.65 (14)
Black (2007) [4]	Zoledronic acid	6.02	5.06	6.71	0.30 (402)	0.59 (140)	0.75 (680)
Lyles (2007) [46]	Zoledronic acid	6.40	4.30	--	--	0.70 (56)	0.74 (186)
Boonen (2012) [47]	Zoledronic acid	2.10	3.30	6.10	0.33 (37)	--	0.65 (13)
Nakamura (2016) [48]	Zoledronic acid	4.03	4.07	7.61	0.34 (39)	**	0.54 (57)
Calcitonin							
Chestnut (2000) [49]	Calcitonin	--	--	0.73	0.79 (241)	0.47 (22)	0.80 (167)
Cathepsin K Inhibitor							
Bonnick (2013) [11]	Odanacatib	2.70	2.70	2.60	--	--	0.37 (22)
Estrogen Compounds							
Cauley (2003) [19]	Estrogen	3.60	--	4.50	0.65 (101)	0.68 (125)	--
Jackson (2006) [24]	Estrogen+Progestin	3.70	--	5.20	0.64 (112)	0.64 (119)	--
Cummings (2008) [10]	Tibolone	--	3.10	4.80	0.56 (196)	0.72 (24)	0.74 (288)
PTH(1-84) and PTH Analogs							
Miller (2016) [18]	Abaloparatide	4.25	4.01	10.37	0.14 (34)	**	0.54 (51)
Neer (2001) [12]	PTH 1-34	4.10	4.65	10.60	0.33 (105)	**	0.62 (119)
Nakamura (2012) [13]	PTH 1-34	3.00	2.30	6.40	0.20 (44)	--	0.87 (28)
Fujita (2013) [14]	PTH 1-34	--	--	3.90	0.22 (27)	--	0.60 (13)
Greenspan (2007) [15]	PTH 1-84	2.11	2.47	6.91	0.42 (60)	**	0.97 (144)
RANKL Inhibitor							
Cummings (2009) [6]	Denosumab	6.00	--	9.20	0.32 (350)	0.61 (69)	0.81 (531)
Nakamura (2014) [50]	Denosumab	5.70	5.10	9.00	0.35 (66)	--	0.97 (39)
SERM							

Cummings (2011) [9]	Arzoxifene	2.60	2.80	2.90	0.60 (288)	0.77 (46)	0.94 (688)
Silverman (2008) [8]	Bazedoxifene	1.21	--	1.42	--	--	0.88 (273)
Itabashi (2011) [51]	Bazedoxifene	1.99	2.46	3.24	0.66 (14)	--	1.00 (12)
Cummings (2010) [7]	Lasofoxifene	2.55	2.95	3.05	0.66 (607)	0.83 (93)	0.84 (795)
Ettinger (1999) [52]	Raloxifene	--	2.25	2.70	0.59 (503)	1.12 (58)	0.91 (677)

Dashes indicate that no data was available
PTH = Parathyroid hormone; SERM = Selective estrogen receptor modulator
TH = Total hip; FN = Femoral neck; LS = Lumbar spine
** Not included in the meta-analysis due to <10 fracture outcomes

Table 4. Results of meta-regression relating BMD change to fracture reduction

	Δ Total Hip BMD	Δ Femoral Neck BMD	Δ Lumbar Spine BMD
Vertebral Fracture			
# of studies included	20	24	30
# of subjects	91340	73904	111183
# of fractures	3174	3630	4557
r^2 (95% CI)	0.56 (0.26,0.70)	0.54 (0.27,0.68)	0.63 (0.41,0.73)
p-value	0.0002	<0.0001	<0.0001
Hip Fracture			
# of studies included	12	13	15
# of subjects	85010	69557	94469
# of fractures	882	816	863
r^2 (95% CI)	0.48 (0.07,0.67)	0.42 (0.05,0.63)	0.22 (0.00,0.46)
p-value	0.013	0.017	0.08
Non-Vertebral Fracture			
# of studies included	22	28	32
# of subjects	74513	84981	92556
# of fractures	4999	6383	6340
r^2 (95% CI)	0.12 (0.00,0.34)	0.12 (0.00,0.31)	0.12 (0.00,0.30)
p-value	0.11	0.07	0.05

Table 5. Estimated fracture risk reduction associated with BMD improvement

	Vertebral Fracture	Hip Fracture	Non-Vertebral Fracture
Δ Total Hip BMD			
2%	28%	16%	10%
4%	51%	29%	16%
6%	66%	40%	21%
Δ Femoral Neck BMD			
2%	28%	15%	11%
4%	55%	32%	19%
6%	72%	46%	27%
Δ Lumbar Spine BMD			
2%	28%	22%	11%
8%	62%	38%	21%
14%	79%	51%	30%

Table S1. Search Terms

	fracture OR fractures OR bone fracture OR bone fractures OR BMD OR bone density OR bone mineral density OR bone density OR density, bone density, bone mineral
AND	RCT OR trial OR clinical trial OR clinical trial OR randomized controlled trial OR clinical trials, randomized OR controlled clinical trials, randomized
AND	Placebo
AND	Humans NOT animals
AND	anti-resorptive OR bone density conservation agents OR osteoporosis therapy OR osteoporosis treatment OR osteoporosis drug OR antiresorptive agents OR antiresorptive drugs OR antiresorptive OR bisphosphonate OR bisphosphonates OR diphosphonate OR diphosphonates OR alendronate OR alendronate OR fosamax OR etidronate OR didrone OR didtronel OR risedronate OR actonel OR boniva OR ibandronate OR zoledronate OR zoledronic acid OR zometa OR reclast OR tiludronate OR Skelid OR pamidronate OR Aredia OR clodronate OR Bonefos OR neridronate OR olpadronate
OR	selective estrogen receptor modulators OR Selective Estrogen Receptor Modulators OR SERM OR raloxifene OR raloxifene OR evista OR evista OR lasofoxifene OR Oporia OR Fablyn OR bazedoxifene OR arzoxifene OR clomifene OR Clomid OR Clomifert OR afimoxifene OR TamoGel OR ormeloxifene OR tamoxifen OR tamoxifen OR nolvadex OR Nolvadex OR toremifene
OR	rank ligand OR RANK ligand OR RANKL OR receptor activator nuclear factor ligand OR denosumab OR xgeva
OR	parathyroid hormone OR human parathyroid hormone 1 34 OR parathyroid hormone 1 34 OR parathyroid hormone 1 84 OR parathyroid hormone OR teriparatide OR teriparatide OR forteo OR Forsteo OR pth 1 34 OR pth 1 84 OR PTH OR recombinant parathyroid hormone OR recombinant PTH OR preos
OR	hormone replacement therapy OR hormone replacement therapy, post menopausal OR postmenopausal hormone replacement therapy OR hormone replacement therapy OR hormone therapy OR estrogen replacement therapy OR conjugated estrogens OR estrogen OR progesterone
OR	tibolone OR Livial
OR	strontium OR strontium
OR	calcitonin OR salmon calcitonin OR Cibacalcin OR Miacalcin OR Calcimar

Table S2. Results of meta-regression relating BMD change to fracture reduction, study duration ≥ 2 years subset

	Δ Total Hip BMD	Δ Femoral Neck BMD	Δ Lumbar Spine BMD
Vertebral Fracture			
# of studies included	15	19	24
# of subjects	76950	59514	96477
# of fractures	2856	3312	4212
r^2 (95% CI)	0.73 (0.43,0.82)	0.64 (0.35,0.76)	0.63 (0.37,0.74)
p-value	<0.0001	<0.0001	<0.0001
Hip Fracture			
# of studies included	11	12	14
# of subjects	77830	62377	87289
# of fractures	862	796	843
r^2 (95% CI)	0.45 (0.04,0.66)	0.38 (0.02,0.61)	0.19 (0.00,0.45)
p-value	0.023	0.033	0.12
Non-Vertebral Fracture			
# of studies included	16	22	25
# of subjects	58215	68683	75942
# of fractures	4470	5854	5798
r^2 (95% CI)	0.16 (0.00,0.41)	0.10 (0.00,0.31)	0.08 (0.00,0.27)
p-value	0.13	0.15	0.18

Table S3. Results of meta-regression relating BMD change to fracture reduction, anti-resorptive subset

	Δ Total Hip BMD	Δ Femoral Neck BMD	Δ Lumbar Spine BMD
Vertebral Fracture			
# of studies included	15	19	24
# of subjects	76950	59514	96477
# of fractures	2856	3312	4212
r^2 (95% CI)	0.73 (0.43,0.82)	0.64 (0.35,0.76)	0.63 (0.37, 0.74)
p-value	<0.0001	<0.0001	<0.0001
Hip Fracture			
# of studies included	11	12	14
# of subjects	77830	62377	87289
# of fractures	862	796	843
r^2 (95% CI)	0.45 (0.04, 0.66)	0.38 (0.02, 0.61)	0.19 (0.00, 0.45)
p-value	0.023	0.033	0.12
Non-Vertebral Fracture			
# of studies included	17	23	26
# of subjects	60123	70591	77850
# of fractures	4526	5910	5854
r^2 (95% CI)	0.12 (0.00,0.36)	0.07 (0.00,0.27)	0.07 (0.00,0.26)
p-value	0.18	0.22	0.19

Table S4. Results of meta-regression relating BMD change to fracture reduction, bisphosphonate subset

	Δ Total Hip BMD	Δ Femoral Neck BMD	Δ Lumbar Spine BMD
Vertebral Fracture			
# of studies included	8	13	14
# of subjects	22485	27925	28518
# of fractures	1318	1638	1734
r^2 (95% CI)	0.63 (0.09,0.78)	0.70 (0.33,0.81)	0.52 (0.14,0.69)
p-value	0.018	0.0004	0.004
Hip Fracture			
# of studies included	6	8	6
# of subjects	24805	32228	20770
# of fractures	410	575	287
r^2 (95% CI)	0.51 (0.00,0.72)	0.27 (0.00,0.56)	0.03 (0.00,0.37)
p-value	0.11	0.19	0.73
Non-Vertebral Fracture			
# of studies included	10	16	16
# of subjects	25267	38756	29500
# of fractures	2166	3389	2362
r^2 (95% CI)	0.14 (0.00,0.44)	0.02 (0.00,0.22)	0.06 (0.00,0.30)
p-value	0.28	0.61	0.34

Table S5. Relative risk predicted from meta-regression results from published placebo-controlled studies and observed relative risk or hazard ratio for active-comparator trials.

ARCH Trial[17]– Head to head active comparator: Romosozumab (12 mo) vs. ALN (12 mo)				
BMD Site		Relative Risk (95% CI) Predicted by ΔBMD at 12 mo		
	ΔBMD Difference at 12 mo (Romo vs ALN)	Vertebral Fracture	Non-Vertebral Fracture	Hip Fracture
LS BMD	8.7%	0.36 (0.35-0.37)	0.78 (0.76-0.79)	0.60 (0.57-0.62)
FN BMD	3.2%	0.55 (0.50-0.60)	0.84 (0.80-0.89)	0.74 (0.68-0.81)
TH BMD	3.4%	0.55 (0.51-0.60)	0.86 (0.82-0.89)	0.75 (0.70-0.79)
Observed RR or HR (95% CI)		0.63 (0.47-0.85)	0.74 (0.54-1.01)	0.64 (0.33-1.26)
ARCH Trial[17]–Sequential therapy: Romosozumab (12 mo)-> ALN (12 mo) vs. ALN (24 mo)				
BMD Site		Relative Risk (95% CI) Predicted by ΔBMD at 24 mo		
	ΔBMD Difference at 24 mo (Romo->ALN vs. Aln->Aln)	Vertebral Fracture	Non-Vertebral Fracture	Hip Fracture
LS BMD	8.1%	0.38 (0.37-0.39)	0.79 (0.77-0.81)	0.61 (0.59-0.64)
FN BMD	3.7%	0.49 (0.44-0.53)	0.82 (0.78-0.87)	0.70 (0.64-0.77)
TH BMD	3.7%	0.52 (0.48-0.57)	0.85 (0.81-0.89)	0.73 (0.68-0.77)
Observed RR (95% CI)		0.52 (0.40-0.66)	0.81 (0.66-0.99)	0.62 (0.42-0.92)
ACTIVE-Extend Trial[25] – Sequential therapy: ABL (18 mo)->ALN (24 mo) vs. PBO (18 mo)->ALN (24 mo)				
BMD Site		Relative Risk (95% CI) Predicted by ΔBMD at 43 mo		
	ΔBMD Difference at 43 mo (ABL->ALN vs PBO->ALN)	Vertebral Fracture	Non-Vertebral Fracture*	Hip Fracture*
LS BMD	8.1%	0.38 (0.37-0.39)	0.79 (0.77-0.80)	0.61 (0.59-0.64)
FN BMD	3.8%	0.47 (0.43-0.52)	0.82 (0.77-0.86)	0.69 (0.64-0.76)
TH BMD	3.9%	0.50 (0.46-0.55)	0.84 (0.81-0.88)	0.71 (0.67-0.76)
Observed RR (95% CI)		0.16 (no CI reported)	0.63 (0.41 – 0.98)	Not reported

*The non-vertebral fracture and hip fracture are reported for the "primary analysis" in the manuscript with a median follow-up of 2.7 yrs

Figure 1. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in vertebral fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.

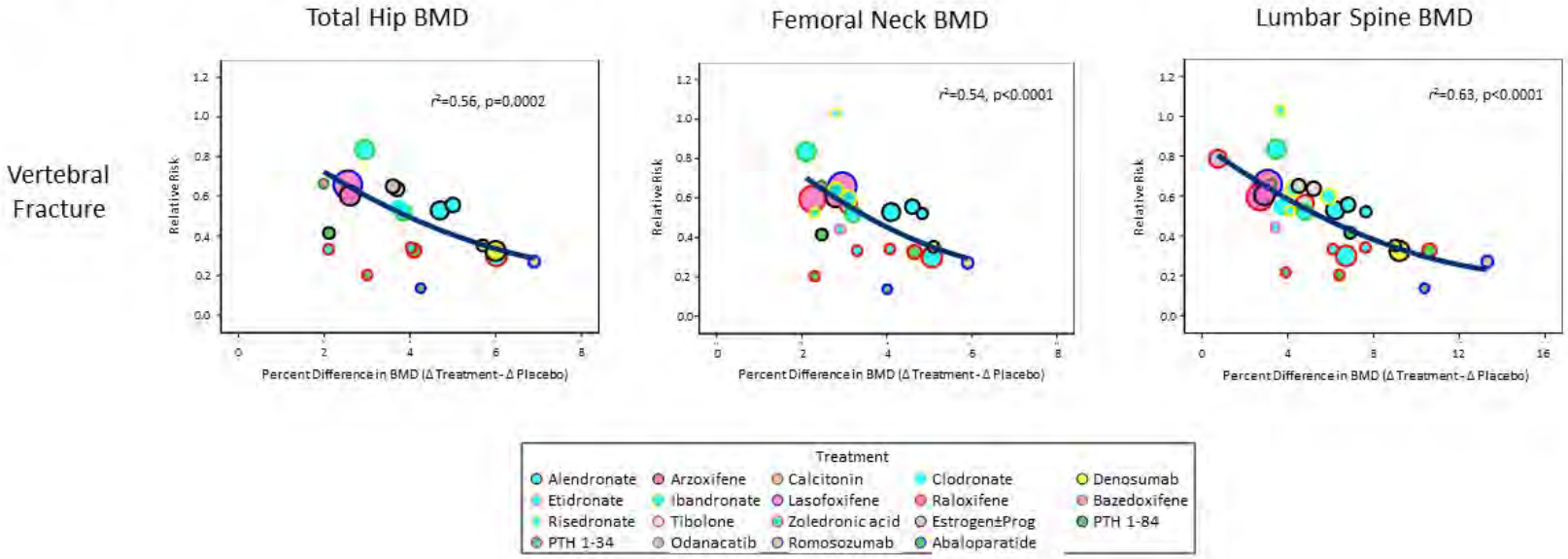


Figure 2. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in hip fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.

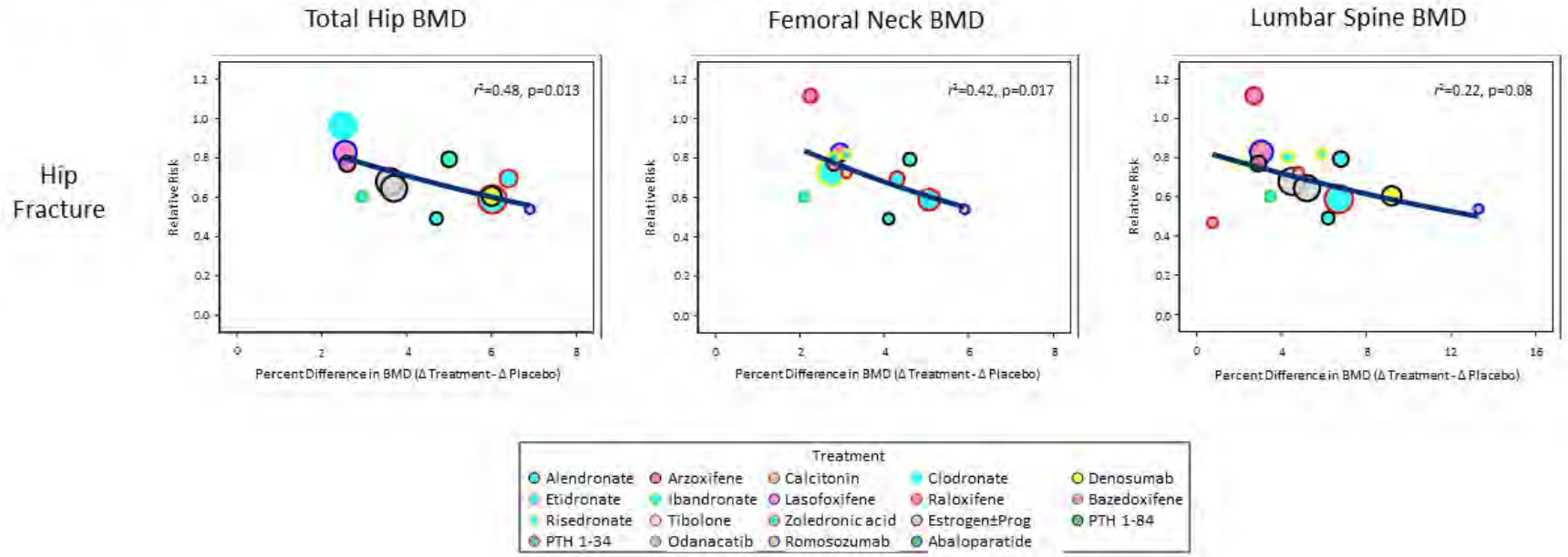
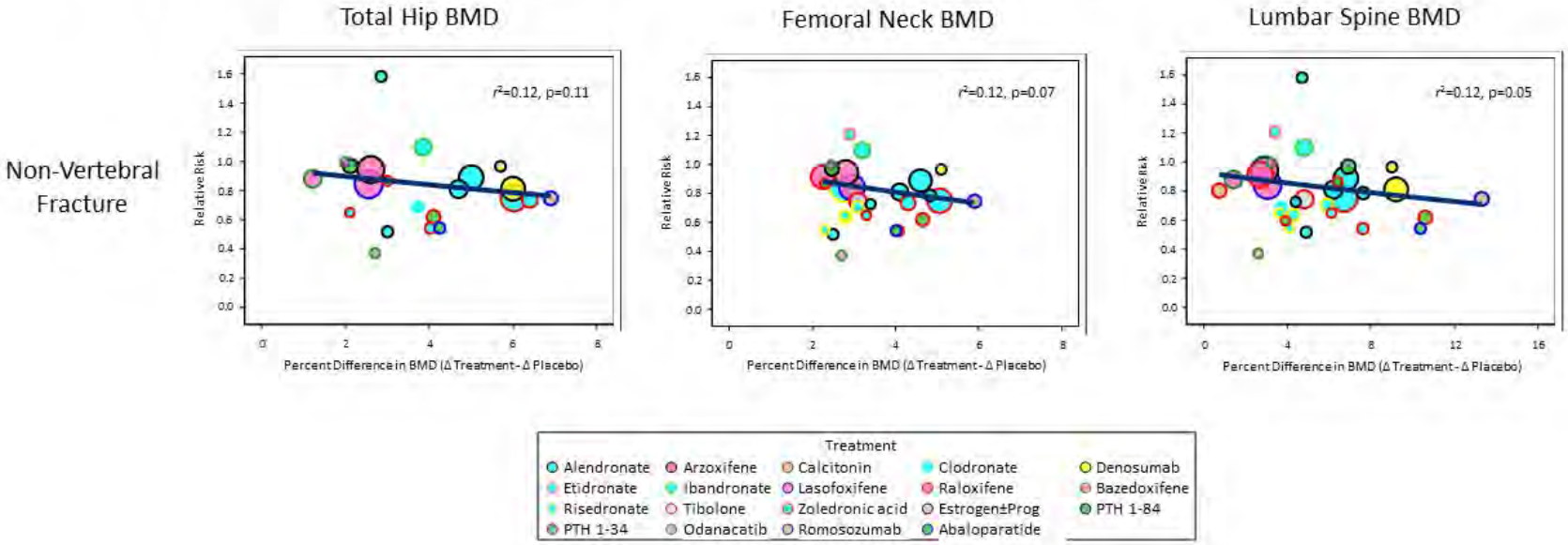


Figure 3. Association between treatment-related differences in BMD (Active-Placebo, in %) and reduction in non-vertebral fracture risk. Individual trials are represented by circles with areas that are approximately proportional to the number of fractures in the trial. Drugs of the same class are represented by the symbols of the same color.



APPENDIX 3: -

Journal of Clinical Densitometry: Assessment & Management of Musculoskeletal Health, vol. 19, no. 2, 127–140, 2016

1094-6950/19:127–140/$36.00
http://dx.doi.org/10.1016/j.jocd.2016.03.003

Best Practices

Best Practices for Dual-Energy X-ray Absorptiometry Measurement and Reporting: International Society for Clinical Densitometry Guidance

E. Michael Lewiecki,*[*,1] *Neil Binkley,*[2] *Sarah L. Morgan,*[3] *Christopher R. Shuhart,*[4] *Bruno Muzzi Camargos,*[5] *John J. Carey,*[6] *Catherine M. Gordon,*[7] *Lawrence G. Jankowski,*[8] *Joon-Kiong Lee,*[9] *and William D. Leslie*[10] *on behalf of the International Society for Clinical Densitometry

[1]New Mexico Clinical Research & Osteoporosis Center, Albuquerque, NM, USA; [2]Osteoporosis Clinical Center and Research Program, University of Wisconsin, Madison, WI, USA; [3]Division of Clinical Immunology and Rheumatology, Department of Medicine, UAB Osteoporosis Prevention and Treatment Clinic, University of Alabama at Birmingham, Birmingham, AL, USA; [4]Swedish Medical Group, Seattle, WA, USA; [5]Rede Mater Dei de Saúde - Densimater, Belo Horizonte, Brazil; [6]Galway University Hospitals, National University of Ireland, Galway, Ireland; [7]Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH, USA; [8]Illinois Bone and Joint Institute, LLC., Morton Grove, IL, USA; [9]JK Lee Orthopaedics & Traumatology, Petaling Jaya, Malaysia; and [10]University of Manitoba, Winnipeg, Manitoba, Canada

Abstract

Dual-energy X-ray absorptiometry (DXA) is a technology that is widely used to diagnose osteoporosis, assess fracture risk, and monitor changes in bone mineral density (BMD). The clinical utility of DXA is highly dependent on the quality of the scan acquisition, analysis, and interpretation. Clinicians are best equipped to manage patients when BMD measurements are correct and interpretation follows well-established standards. Poor-quality acquisition, analysis, or interpretation of DXA data may mislead referring clinicians, resulting in unnecessary diagnostic evaluations, failure to evaluate when needed, inappropriate treatment, or failure to provide medical treatment, with potentially ineffective, harmful, or costly consequences. Misallocation of limited healthcare resources and poor treatment decisions can be minimized, and patient care optimized, through meticulous attention to DXA instrument calibration, data acquisition and analysis, interpretation, and reporting. This document from the International Society for Clinical Densitometry describes quality standards for BMD testing at DXA facilities worldwide to provide guidance for DXA supervisors, technologists, interpreters, and clinicians. High-quality DXA testing is necessary for correct diagnostic classification and optimal fracture risk assessment, and is essential for BMD monitoring.

Received 03/7/16; Accepted 03/8/16.

Disclosure: EML has served on scientific advisory boards for Amgen, Merck, Eli Lilly, Radius Health, AgNovos Healthcare, Alexion, Shire, and AbbVie. NB has served on scientific advisory boards for Amgen, Merck, and Eli Lilly. CRS serves on scientific and commercial advisory boards for Amgen. WDL has received speaker fees from Amgen, Eli Lily, and Novartis. BMC has received speaker fees from GE, Hologic, Aché, EMS, Hypermarcas and Medimaps. EML has received institutional grant/research support from Amgen, Merck, and Eli Lilly. NB has received institutional grant/research support from Amgen, Merck, Eli Lilly, and GE Healthcare.WDL received research grants from Amgen (all paid to the institution). JJC was served on scientific advisory boards for MSD, Abbvie, Amgen, Prcoter & Gamble, Hospira,Servier, Pfizer, Bristol Myers Squibb, and Roche, and received speaker fees from MSD, Roche, Abbvie, Eli-Lilly, Bristol Myers Squibb, Hospira, and Pfizer. Research Grants: Abbvie, MSD, P&G, and Centocor.

*Address correspondence to: E. Michael Lewiecki, MD, New Mexico Clinical Research & Osteoporosis Center, 300 Oak St. NE, Albuquerque, NM 87106. E-mail: mlewiecki@gmail.com

Key Words: Accreditation; certification; DXA; osteoporosis; quality.

Introduction

Dual-energy X-ray absorptiometry (DXA) is a quantitative radiological procedure for measuring bone mineral density (BMD), a major determinant of bone strength *(1)*. DXA measurements are used to diagnose osteoporosis *(2)*, monitor changes in BMD over time *(3)*, and estimate fracture risk, *(4)* and, as such, are often integral to therapeutic intervention recommendations. Indeed, BMD by DXA is a component of osteoporosis treatment guidelines in the United States *(5,6)*, Canada *(7)*, Europe *(8)*, United Kingdom *(9)*, and elsewhere *(10)*. Femoral neck BMD by DXA is an important risk factor input for the World Health Organization (WHO) fracture risk assessment algorithm (FRAX) *(11)*. DXA also has applications beyond BMD testing, including vertebral fracture assessment *(12)*, analysis of body composition *(13)*, hip structural analysis *(14)*, and trabecular bone score determination *(15)*. Physicians rely on DXA measurements to manage patients with skeletal disorders. Poor-quality DXA acquisition/analysis and/or incorrect reporting of the results may result in the ordering of unnecessary diagnostic tests, failing to order needed tests, or inappropriately starting, stopping, or changing treatment. Such errors in clinical practice are unfortunately common, sometimes costly, and potentially harmful to patients *(16–21)*. DXA scans in growing children and adolescents are particularly challenging and errors are common with respect to both data acquisition and interpretation *(22)*. These errors can lead to the inappropriate initiation of skeletal agents, many of which have unknown side effects in pediatric patients, and other inappropriate management decisions.

A central DXA system is composed of a padded table for the patient, an X-ray source, a radiation detector, computer hardware and software, and usually a printer for generating a hard copy of data, graphs, and images *(23)*. These sophisticated scientific instruments are manufactured with rigorous technical standards. Upon completion of the manufacturing process, the DXA system is transported to the end-user facility and assembled by a technician who checks system calibration to assure the accuracy (more correctly referred to as "trueness") of the measurements and makes adjustments as needed. The DXA technologist(s) may receive basic training from the manufacturer (e.g., by an applications specialist) in quality assessment, instrument maintenance, patient positioning, data acquisition, and analysis. Following densitometer installation, there may be local regulatory requirements that apply to the system (e.g., radiation safety standards and inspection) or for the technologist (e.g., training as a radiological technologist, licensure, certification). The physician who is responsible for supervising a DXA facility, interpreting the DXA results, and signing off on the report must have sufficient training to assure that the data are correct and that interpretation and reporting conform to current standards in the field *(24)*. Typically, US state and local regulations do not require any specific qualifications for DXA interpretation *(25)*, despite the important technical aspects of the test discussed here. US Medicare regulations only require some qualifications of supervising physicians in independent diagnostic testing facilities *(26)*, but not in hospital facilities or private clinical practices. In Canada, 3 provinces currently have a requirement for International Society for Clinical Densitometry (ISCD) certification for physicians who are reporting or supervising a DXA facility. In Brazil, certification by the Brazilian Radiology Society (Colégio Brasileiro de Radiologia) is required for any physician to perform DXA acquisition, analysis, and reporting. Technical certification, issued by the Brazilian Society of Radiologic Technologists (Conselho de Técnicos em Radiologia), is required for other allied healthcare professionals to perform DXA acquisitions. Globally, requirements for training, performing, and interpreting DXA scans by healthcare professionals are variable.

The generation of high-quality DXA reports requires an understanding of potential sources of errors, including changes in instrument calibration, improper patient positioning or analysis, recognition of confounding artifacts, and correct selection of reference databases for T- and Z-score calculation, thus requiring skilled technologists and interpreting physicians to assure production of a high-quality report. Over time, densitometer calibration may change due to degradation of the components (e.g., X-ray tube and detector), moving the instrument to a different location, or a variety of other factors. The skills of a DXA technologist may improve with experience or worsen over time, or a highly proficient technologist may leave and be replaced by one who is less skilled. Similarly, a physician involved may be dedicated to very high DXA quality or may view DXA as a sideline to other responsibilities. For all of these reasons, the reliability of DXA measurements and reports is sometimes in doubt, thereby having potential adverse effects on the management of patients *(16–19)*.

The ISCD is an international organization with global membership dedicated to advancing excellence in the assessment of skeletal health by promoting education and understanding of the clinical applications of bone mass measurement and other skeletal health assessment technologies. The ISCD strives to assure proficiency and quality in the assessment of skeletal health through education, certification, and accreditation in bone densitometry. To highlight the essential components of quality DXA testing, the ISCD herein identifies DXA Best Practices (Box). The DXA Best Practices are not meant to be a comprehensive list of all features that characterize a high-quality DXA facility, but rather these practices identify a basic set of essential markers that are consistent with high quality. For the purposes of this document, quality is defined as the degree to which DXA measurements and interpretation are consistent with current professional standards to facilitate desired

Box. DXA Best Practices

Scan Acquisition and Analysis

1.1. At least one practicing DXA technologist, and preferably all, has a valid certification in bone densitometry.
1.2. Each DXA technologist has access to the manufacturer's manual of technical standards and applies these standards for BMD measurement.
1.3. Each DXA facility has detailed standard operating procedures for DXA performance that are updated when appropriate and available for review by all key personnel.
1.4. The DXA facility must comply with all applicable radiation safety requirements.
1.5. Spine phantom BMD measurement is performed at least once weekly to document stability of DXA performance over time. BMD values must be maintained within a tolerance of ±1.5%, with a defined ongoing monitoring plan that defines a correction approach when the tolerance has been exceeded.
1.6. Each DXA technologist has performed in vivo precision assessment according to standard methods and the facility LSC has been calculated.
1.7. The LSC for each DXA technologist should not exceed 5.3% for the lumbar spine, 5.0% for the total proximal femur, and 6.9% for the femoral neck.

Interpretation and Reporting

2.1. At least 1 practicing DXA interpreter, and preferably all, has a valid certification in bone densitometry.
2.2. The DXA manufacturer and model are noted on the report.
2.3. The DXA report includes a statement regarding scan factors that may adversely affect acquisition/analysis quality and artifacts/confounders, if present.
2.4. The DXA report identifies the skeletal site, region of interest, and body side for each technically valid BMD measurement.
2.5. There is a single diagnosis reported for each patient, not a different diagnosis for each skeletal site measured.
2.6. A fracture risk assessment tool is used appropriately.
2.7. When reporting differences in BMD with serial measurements, only those changes that meet or exceed the LSC are reported as a change.

BMD, bone mineral density; DXA, dual-energy X-ray absorptiometry; LSC, least significant change.

health outcomes. These DXA Best Practices are intended to serve as a guide and expectation for DXA supervisors, technologists, interpreters, and clinicians. Following these DXA Best Practices aids patients, referring healthcare providers, and payers by facilitating recognition of high-quality DXA services. DXA Best Practices are applicable worldwide for adult and pediatric DXA testing, recognizing that adaptations may be required according to local circumstances and country-specific standards.

Overview of High-Quality DXA Performance

Quality DXA studies require instrument calibration within an acceptable range of tolerance, rigorous attention to detail in assuring correct scan acquisition and analysis, understanding serial BMD "test–retest" precision, and appropriate application of guidelines for interpretation and reporting. This can be achieved through bone densitometry training and validated by certification for the DXA technologist and interpreting physician; the implementation of what is learned from training can be confirmed through facility accreditation.

Implementation of DXA Best Practices

The ISCD recommends that DXA facilities establish standard operating procedures (SOPs) as a guide for adherence to DXA Best Practices. For others (e.g., patients, referring physicians, and payers) interested in assessing competency of those responsible for bone densitometry, technologist and interpreter certification provides a measure of attaining basic DXA knowledge; DXA facility accreditation provides additional assurance that high-quality DXA is being performed.

Methodology

These DXA Best Practices are derived from the ISCD Official Positions *(13,24,27–34)* that are developed and periodically updated through Position Development Conferences held regularly since 2001. The ISCD is the only organization exclusively dedicated to advancing excellence in the assessment of skeletal health, doing so through education, certification, accreditation, and development of evidence-based quality standards. The ISCD Official Positions have been established through a process of

rigorous review of the best medical evidence by internationally recognized experts in skeletal health assessment, often in collaboration with other stakeholder organizations. Evaluation of the evidence when developing Official Positions is conducted using a modification of the RAND Corporation and University of California at Los Angeles method (RAM) *(35)*. This method has been used worldwide to determine whether medical procedures are expected to provide a specific health benefit that exceeds the potential negative consequences by such a wide margin that the procedure or indication is worth doing. The rationale for use of the RAM in the development of the ISCD Official Positions is based on its ability to combine the best available scientific evidence with the collective judgment of the expert panel consisting of a broad range of professionals within and outside of the ISCD.

Scan Acquisition and Analysis

At Least One Practicing DXA Technologist, and Preferably All, Has a Valid Certification in Bone Densitometry

Rationale. Measurement of BMD by DXA is technically demanding, with reliability of the output (BMD, T-score, and Z-score) dependent on technologist training and skill. By receiving training in DXA acquisition and analysis, passing an examination and receiving certification in bone densitometry, a technologist provides assurance that a basic skill set has been acquired. Keeping the certification current through continuing medical education and/or subsequent examinations demonstrates that these skills have been maintained and evolved with new developments in the field. Ideally all DXA technologists should be fully trained and certified in bone densitometry; however, a single certified technologist at each DXA facility may be capable of educating, supervising, and monitoring the quality of DXA studies by other technologists at the same facility. If children are being scanned at a DXA facility, at least 1 technologist should ideally have undergone additional instruction in pediatric densitometry (ISCD pediatric bone density course or similar training), as the adjustment of Z-score for height and other clinical variables is critically important *(36)*.

Comments. As part of the training and certification process, technologists come to recognize that densitometer maintenance, scan acquisition, and scan analysis must be rigorously conducted according to standard procedures *(24)*. This approach provides the interpreter with valid data needed to generate a correct and clinically useful DXA report, thereby giving the referring healthcare provider appropriate information to make wise patient care decisions. With updates in DXA software, changes in DXA systems, and evolution of quality standards (e.g., reference database standardization for T-score calculation), it is necessary that DXA technologists stay current in the field.

Failure to follow standard procedures may result in invalid data, which can be misleading and potentially harmful for patient care *(16,17,19,37,38)*. Examples of DXA errors abound. These include incorrect patient positioning and/or analysis, failure to consider confounding artifacts that affect BMD values, and inappropriate reference database use for T-score derivation. Additional errors include failure to recognize densitometer drift or shift that could lead to reporting an inappropriate BMD change, thus leading to alteration of therapy, failure to change therapy, and/or unnecessary diagnostic studies. Another common error is failure to perform precision assessment, resulting in inability to distinguish between an apparent BMD difference that is simply within the range of error of the test vs one that is statistically significant.

DXA certification provides evidence that a basic body of knowledge has been acquired. A "valid" certification is one that is currently active (i.e., not expired). Certification should be maintained through proof of continuing education in the DXA field and/or reexamination because of evolving technologies and standards in bone densitometry. Accreditation of a DXA facility by a neutral third party is a formal declaration that the facility meets international standards for development, implementation, and maintenance of the certification program. Examples of accrediting agencies include the National Commission for Certifying Agencies *(39)* and the American National Standards Institute *(40)*. These agencies were developed to ensure the health, welfare, and safety of the public.

Each DXA Technologist Has Access to the Manufacturer's Manual of Technical Standards and Applies These Standards for BMD Measurement

Rationale. There are important manufacturer-specific differences in DXA hardware, software, instrument operation, and requirements for patient positioning *(18)*. DXA systems use complex digital technologies that generate numerical data, the validity of which is highly dependent on the application of appropriate manufacturer-specific standard methods of operation. The manufacturer's manual of instructions, in print or electronic format, is the primary resource for quality control standards, instrument maintenance, patient scanning, and data analysis.

Comments. Each DXA system is delivered with a manual of instructions that may be in printed form, embedded in computer software, on external electronic media, or online. This manual is an important resource to understand proper instrument use. As time passes, some of the information in the manual may be revised or updated. However, accessibility, understanding, and application of the manual's contents by facility staff is likely to vary widely depending on the initial level of interest, changes in staffing, and procedures for assuring continuity of quality standards. Deviations from recommended procedures that may adversely affect the validity of BMD measurements include the use

of a nonstandard phantom (41), failure to recognize and correct changes in instrument calibration (17), and nonstandard patient positioning (42).

Each DXA Facility Has Detailed SOPs for DXA Performance That Are Updated When Appropriate and Available for Review by All Key Personnel

Rationale. Measurement of BMD by DXA is a process that requires integration of procedures that can be placed into 3 categories: pretesting (e.g., patient scheduling, preparation, and education, as well as instrument calibration and maintenance), testing (e.g., selection of skeletal sites to measure, scan mode, patient positioning), and post-testing (e.g., analysis, interpretation, reporting). SOPs that are carefully conceived, drafted, executed, and maintained provide a systematic method for assuring that all components that contribute to quality DXA are recognized and instituted.

Comments. Establishing effective procedures for implementing and maintaining quality standards is an important element of reliability in radiological procedures. Standardization of radiological processes can reduce errors and improve patient safety (43). Individuals involved in all aspects of bone densitometry should participate in the development of SOPs (44). Examples of elements in effective SOPs include a statement of the SOP purpose, scope of the SOPs, related documentation, definitions of terms, responsible staff, exact steps of the procedure, error analysis (i.e., a systematic method to analyze errors for the purpose of improving performance, with correction steps when errors are found), required quality control methods for the procedure, and guidelines for reporting DXA results. Examples of SOPs for some DXA procedures are available online (45).

The DXA Facility Must Comply With All Applicable Radiation Safety Requirements

Rationale. DXA scanning uses ionizing radiation in the form of X-rays, which can theoretically cause harm despite the extremely low radiation dose. For both patient and technologist safety, all applicable radiation safety guidelines and requirements must be followed to minimize the risk from diagnostic radiation.

Comments. Radiation safety issues with DXA have been identified and described (46). While it is not possible to precisely quantitate random effects from the low doses of ionizing radiation associated with DXA, for purposes of radiation protection, there is assumed to be a linear relationship between dose and adverse effects, with no threshold below which adverse effects are not possible (47). The typical level of background radiation to which the general population is exposed, not including radiation due to medical procedures, has been estimated to be about 2.5 mSv/yr (48). A DXA scan is associated with radiation exposure (effective dose) of about 5 μSv or 0.005 mSv. At facilities where young children and adolescents are scanned, these concepts are considered very carefully by radiation safety committees; the scrutiny of clinical and research protocols is often stricter than that for adults.

Three concepts related to DXA scanning should be considered in protecting the public and technologists from radiation harm (46): justification—a DXA scan should not be performed unless there is net benefit to the patient; optimization—radiation exposure should be as low as reasonably achievable by limiting the time of exposure, maximizing the distance from the source of radiation, and using shielding when appropriate; and regulation—adherence to all applicable regulations (e.g., by city, state/province, country) to minimize excessive radiation exposure from diagnostic procedures.

Spine Phantom BMD Measurement Is Performed at Least Once Weekly to Document the Stability of DXA Performance Over Time; BMD Values Must Be Maintained Within a Tolerance of ± 1.5%, with an Ongoing Monitoring Plan That Defines a Correction Approach When the Tolerance Is Exceeded

Rationale. The accuracy and precision of BMD measurements by DXA can be adversely affected by changes in instrument performance that may occur suddenly (calibration "shift") or slowly (calibration "drift"). To detect these changes and know that BMD measurements are stable over time, a phantom (standardized object with known BMD) should be scanned at regular intervals. This provides assurance that the X-ray source, radiation detectors, and software algorithms are operating correctly. The scanning of a phantom verifies densitometer performance and assures that DXA results are stable over time (49).

Comments. Phantom scanning can determine when a DXA system is out of calibration and requires service. Phantom scanning does not calibrate the system but is an independent test object that can be scanned as a patient proxy. This allows monitoring of the system to identify problems within the calibration process itself (49). A suitable quality control program requires periodic scanning of a phantom of known BMD, bone mineral content, and area. The phantom is semianthropomorphic and made of either aluminum or hydroxyapatite. Longitudinal scanning of a phantom over time assures that instrument performance parameters of the entire imaging and processing chain are stable over time.

When a manufacturer recommends phantom scanning at specified intervals, this should be done as advised. BMD, bone mineral content, and areas of the phantom should be plotted on a graph based on Shewhart plots (23,50,51). To construct a Shewhart plot, the anthropometric phantom is scanned 10 times and the mean phantom BMD is established as the baseline. The phantom is then scanned on a regular basis according to manufacturer's directions and/or the DXA facility's SOPs, with the results recorded and monitored. On the Shewhart plot, a band ± 1.5% (±3 standard deviations [SDs]) around the phantom mean BMD

delineates the upper and lower limits *(47,49)*. If the phantom value falls outside the upper or lower control limit, the phantom should be rescanned. If the rescan value also falls outside of acceptable ranges, then patient scanning should be postponed until machine service occurs. The Shewhart plots should be reviewed regularly to assure that there is no short-term shift or long-term drift in BMD values. Following routine preventive or other scanner maintenance, the phantom should be scanned 10 times without repositioning between scans. If the mean BMD of these 10 scans differs from the mean of prior daily phantom scans by more than the established limits, then the machine should be recalibrated and a new mean of 10 further scans is established *(47,49)*. Depending on the DXA manufacturer, the Shewhart plot may be automatically generated or may need to be created manually. Facilities may wish to invoke more rigorous phantom scanning protocols (i.e., daily phantom scanning and tighter phantom limits), as many facilities have long-term CVs <0.5%.

Each DXA Technologist Has Performed In Vivo Precision Assessment According to Standard Methods and the Facility Least Significant Change (LSC) Has Been Calculated

Rationale. All quantitative tests in medicine have inherent uncertainty. With DXA BMD measurement, the main sources of variability are patient factors, the technologist, and the instrument *(52)*. Knowledge of the magnitude of this random uncertainty is essential to determine when a BMD "change" is real *(46)*. BMD precision (i.e., reproducibility of the measurement) is the ability of the same densitometer and technologist to obtain the same result when measuring a patient multiple times over a short period *(46)*. When a follow-up BMD measurement differs by the LSC or more, the clinician can conclude that a real loss or gain in BMD has occurred.

Comments. Determination of LSC requires precision assessment. This involves repeat BMD measurements in individuals representative of the clinic's patient population according to a well-established methodology *(53)*. Generally, this consists of measuring 30 patients twice, or 15 patients 3 times, with repositioning between scans. Precision assessment is not a research study and should not require institutional review board approval *(46)*. However, as precision assessment exposes the patient to additional radiation beyond that of a single DXA, the patient should be informed of the reason for precision assessment and agreement (verbal or written) obtained prior to performing the second scan. Precision error is subsequently calculated as the root mean square SD. The LSC with 95% confidence is the precision error × 2.77; this value is easily determined using online calculators *(54)*. Variation in patient position during scan acquisition and variability in subsequent analysis are important factors that influence BMD precision. When multiple technologists are performing BMD measurements at a facility, it is recommended that the average LSC of all technologists be used *(24)*. If a DXA facility has not performed precision assessment, then quantitative comparison of serial BMD measurements is not possible.

The LSC for Each DXA Technologist Should Not Exceed 5.3% for the Lumbar Spine, 5.0% for the Total Hip, and 6.9% for the Femoral Neck

Rationale. BMD precision error values acceptable for clinical practice were determined by a meta-analysis of published BMD precision studies *(55)*. In the studies comprising this meta-analysis, precision values were reported as percent coefficient of variation (%CV) rather than absolute SD values in gram per square centimeter, the latter of which is recommended in clinical practice *(56)*.

Comments: Technologist precision and quantitative BMD comparisons in clinical practice should use the LSC expressed as an absolute value in gram per square centimeter *(53)*. This is preferable to using %CV as it is less affected by the baseline BMD value; as an example, the same absolute change in BMD with a very low baseline BMD would represent a greater percentage change compared with a higher baseline BMD. DXA precision calculators that are available online *(54)* can be set to express precision as either gram per square centimeter or %CV. As such, it is possible to determine whether the technologists are meeting the precision standards. If a technologist has exceeded these acceptable values, retraining is necessary. If the LSC is very large, then expected changes in BMD over time with disease or treatment cannot be detected within a clinically useful time interval.

Interpretation and Reporting

At Least One Practicing DXA Interpreter, and Preferably All, Has a Valid Certification in Bone Densitometry

Rationale. DXA interpretation requires awareness and understanding of issues that include patient positioning, data analysis, precision assessment and LSC, reference databases, diagnostic criteria, and treatment guidelines. DXA reports must provide information that is correct and meaningful for the referring healthcare provider. By passing an examination and receiving a certification in bone densitometry, an interpreter provides evidence that a basic skill set has been acquired; keeping the certification current through continuing medical education relevant to DXA and/or subsequent examinations shows that these skills have been maintained as the field has evolved. Ideally, all DXA interpreters should be well trained and certified in bone densitometry; however, a single certified interpreter at each DXA facility may be capable of educating, supervising, and monitoring the quality of other interpreters at the same facility.

Comments. Standards for measuring BMD, diagnosing osteoporosis, assessing fracture risk, and treatment

recommendations are continually evolving. Examples of common mistakes (16,17,19,37,38) that could result in an incorrect interpretation of DXA include the following: failure to recognize the presence of an artifact that invalidates BMD measurement, use of an invalid skeletal site for diagnostic classification, reporting a different diagnosis and fracture risk for each skeletal site and region of interest (ROI) measured, reporting T-scores when Z-scores should be used, using an incorrect reference database for generating T-scores or Z-scores, comparing T-scores when interpreting serial DXA studies rather than BMD in gram per square centimeter, entering incorrect information into the FRAX algorithm, and giving inappropriate recommendations for evaluation and treatment due to inadequate understanding of applicable guidelines. In interpreting the scans of children and adolescents with chronic disease (as DXA-derived measures of areal BMD can be confounded by bone size), the Z-score may need adjustment for height, and in some clinical settings, bone age, to ensure that the Z-score is not confounded by delayed skeletal growth and/or maturation (36).

DXA certification provides evidence that a basic body of knowledge has been acquired. A "valid" certification is one that is currently active (i.e., not expired). As standards and guidelines for DXA and osteoporosis management evolve, it is necessary that DXA interpreters stay current in the field. Certification should be maintained through proof of continuing education and/or reexamination because of evolving technologies and standards in bone densitometry.

The DXA Manufacturer and Model Are Noted on the Report

Rationale. There are important differences in hardware, software, reference databases, and operational protocols among DXA manufacturers. A patient with BMD measured on 1 manufacturer's densitometer may have a different BMD and/or T-score when measured on another, even when there is no real difference in BMD. Quantitative comparison with a previous DXA study requires that BMD be measured on the same instrument at the same facility, with knowledge of LSC, unless a cross-calibration study has been done between the different instruments.

Comments. Differences in manufacturer's recommendations for patient positioning, bone edge detection algorithms, calibration methods, ROIs, and reference databases are largely responsible for discrepancies in BMD values measured with DXA systems of different manufacturers (49,57). Comparing results of measurements on different machines requires cross-calibration procedures (29,55), but there is a statistical penalty (i.e., greater LSC with reduced sensitivity for detecting change) paid for these comparisons (58). Identification of the DXA manufacturer is helpful for referring physicians to validate that a quantitative comparison is possible.

The DXA Report Includes a Statement Regarding Scan Factors That May Adversely Affect Acquisition/Analysis Quality and Artifacts/ Confounders, if Present

Rationale. DXA results depend greatly on the skills of the technologist to properly position the patient and subsequently analyze the data for interpretation and reporting. Collectively, these functions are referred to as acquisition and analysis. Manufacturer's training, thorough knowledge of technical manuals, and adherence to SOPs are prerequisites for quality acquisition and analysis. The consequences of faulty acquisition and analysis are well documented (16–18), and at times alter or invalidate DXA interpretation. The interpreter must alert the referring provider of these possibilities and their consequences through a clear statement of scan technical quality. Artifacts that may confound BMD measurements are commonly classified as internal (intrinsic to the patient when disrobed) or external (able to be removed).

Comments. Acquisition and analysis errors may require repeat analysis, repeat scanning, or having the patient return for scan of an additional skeletal site. Important clinical consequences can ensue from these errors, including missed opportunities for treatment, unnecessary treatment, inappropriate laboratory testing, failure to perform appropriate laboratory tests, return visits, and additional healthcare costs (16,17). Lack of awareness of anatomic variation in vertebral segmentation can create confusion with DXA analysis and can have meaningful adverse effects on the interpretation of the results (59). In a 2008 survey, referring physicians thought it important that the DXA interpreter provide information about the technical quality and limitations of the report (60). Internal artifacts can represent common consequences of aging (e.g., degenerative spine changes and aortic calcification) or medical interventions (e.g., hip prosthesis and inferior vena cava filter). External artifacts related to clothing, jewelry, or other man-made objects should be removed, when possible, before proper scan acquisition. Careful preprocedure questioning and astute observation by technologists can mitigate or eliminate impacts of artifacts. Sometimes, serious disease states (e.g., Paget's disease of bone, osteolytic or osteoblastic malignancies) are suggested on the DXA images; these should be noted on the report so that appropriate evaluation can be initiated.

The DXA Report Identifies the Skeletal Site, ROI, and Body Side for Each Technically Valid BMD Measurement

Rationale. The identification of the skeletal site, ROI, and body side (when applicable) documents the exact area scanned; this allows the technologist to scan the same ROI in follow-up studies, provides interpreters with essential information when generating results, and allows referring

healthcare providers to document that the same skeletal sites were used to monitor BMD change over time.

Comment. An important component of DXA interpretation involves scrutinizing the skeletal images to assess patient positioning, correctness of edge detection, potentially confounding artifacts, and placement of margins to delineate ROIs *(49)*. If scanning of any skeletal site is not technically valid, the values for that site should not be reported. Failure to properly identify skeletal sites and use of improper ROIs, particularly on follow-up scanning, can potentially provide incorrect data for use in clinical care. Technical standards exist regarding skeletal sites and ROIs for scanning and reporting *(24)*. For lumbar spine BMD, L1–L4 should be measured, only excluding vertebrae that are affected by local structural change or artifact, using at least 2 vertebrae for diagnostic classification. Anatomically abnormal vertebrae may be excluded from analysis if they are clearly abnormal and nonassessable within the resolution of the system, supported by more than a 1.0 T-score difference between the vertebra in question and adjacent vertebrae *(24)*. Lateral spine BMD measurement should not be used for diagnosis. For hip BMD, only the femoral neck and total proximal femur ROIs should be used for diagnostic classification in adults. The mean hip BMD can be used for monitoring in adults and older adolescents (age >15 yr), with total proximal femur being preferred. However, in children and young adolescents, the hip should generally be excluded as a skeletal assessment site, as positioning in this age group is challenging and skeletal landmarks that guide consistent positioning are not well developed. For forearm DXA measurements, use of the 33% radius (one-third radius) of the nondominant forearm is recommended for diagnosis; other forearm ROIs are not recommended *(24)*. In children and adolescents, total body less head is the recommended assessment site for baseline and ongoing monitoring of bone health. The whole body scan also provides a measurement of body composition, which may be helpful in the ongoing evaluation of youth with chronic diseases.

There Is a Single Diagnosis Reported for Each Patient, Not a Different Diagnosis for Each Skeletal Site Measured

Rationale. The densitometric diagnosis of osteoporosis in clinical practice is made by applying the WHO criteria *(2)* to each appropriate patient using a limited number of skeletal sites *(24)*. This allows for a consistent diagnostic classification for application to treatment guidelines and fracture risk assessment. The WHO criteria are not applicable to premenopausal women, men under age 50 yr, children, and adolescents.

Comment. The ISCD Official Positions state that osteoporosis may be diagnosed in postmenopausal women and in men aged 50 yr and older if the T-score of the lumbar spine, total proximal femur, femoral neck, or 33% radius is ≤–2.5, using a uniform Caucasian (nonrace adjusted) female normative database to derive T-scores for women and men of all ethnic groups *(24)*. This convention should be used in reporting DXA scans; however, application of this recommendation may vary according to local requirements *(24)*. Manufacturers are advised to use National Health and Nutrition Examination Survey III young-adult Caucasian female BMD data as the reference standard for femoral neck and total proximal femur T-score calculation and to continue to use their own reference databases for lumbar spine T-score calculation *(24)*. However, country-specific guidelines related to the use of T-scores may differ from international guidelines *(61)*. As an example, in Japan, T-scores are not used for diagnostic classification *(61)*; therefore, statements regarding T-scores for diagnosis are not applicable in Japan. If local reference data are available, they should be used to calculate Z-scores but not T-scores. Guidelines have been developed for BMD measurement, interpretation, and reporting in children and in adolescents *(34)*, as well as in premenopausal women and in men <50 yr of age *(24)*; interpreters should be aware of, and follow, these guidelines.

A Fracture Risk Assessment Tool Is Used Appropriately

Rationale. In some locations, the therapeutic intervention threshold (i.e., the cut-point at which pharmacologic therapy is recommended) historically was based on the BMD T-score alone. However, the majority of "osteoporosis-related" fractures occur in individuals with low bone mass (osteopenia) or normal BMD *(62,63)*. To improve targeting of interventions to those most likely to sustain fractures, various fracture risk assessment tools have been developed for adult patients. The FRAX tool developed by the WHO is most widely used. It is well studied and has many country-specific versions. FRAX utilizes clinical risk factors with or without femoral neck BMD to estimate the 10-yr risk for major osteoporosis-related fractures (clinical spine, forearm, hip, or shoulder) and for hip fracture alone. Other calculators exist; for example, the Garvan calculator allows inclusion of the number of prior fractures and falls *(64)*. In some regions of the world, therapeutic intervention thresholds are linked to fracture risk estimates. Like all tools, it is important to use these calculators as intended; for example, FRAX is intended to assess fracture risk and to assist in treatment decisions in individuals between the ages of 40 and 90 yr. Additionally, it is important to recognize when to check "yes" in the FRAX calculator for a given clinical risk factor. For example, to consider alcohol consumption as a risk factor, it needs to be 3 or more units per day with 1 unit defined as a 285-mL glass of beer, a 30-mL serving of liquor, or 120 mL of wine. These definitions are listed on the FRAX website and include a useful frequently asked question page that all users should refer to.

Comment. These calculators are not meant to replace clinical judgment and it is not necessary to rigidly follow treatment guidelines based upon such results. While

Table 1
Examples of Resources for DXA Training and Certification/Accreditation

Organization	Description	Weblink
American Bone Health	Limited permit X-ray technician	https://americanbonehealth.org/limited-permit-x-ray-technician-school-bone-densitometry
American College of Radiology	Practice parameter for the performance of DXA	http://www.acr.org/~/media/eb34da2f786d4f8e96a70b75ee035992.pdf
American Registry of Radiologic Technologists	Training and certification for technologists	https://www.arrt.org/pdfs/Disciplines/Handbooks/BD-Handbook.pdf https://www.arrt.org/pdfs/disciplines/clinical-experience/bd-clinical-experience.pdf
American Society of Radiologic Technologists	Training and certification for technologists	http://www.asrt.org/students/study-guides/bone-densitometry https://www.asrt.org/docs/default-source/educators/bonedensitometrycurriculum.pdf http://www.asrt.org/events-and-conferences/event-calendar
Auntminnie.com	Bone densitometry course for technologists	http://www.auntminnie.com/(F(AiaAhFYYF2NIpZ-LQYAK9zBSaE53uNbrdw8TMEotZJ4C_auBzpJsKf51OZTxmuNjXb903IJaUqAs9rhc5QxVyVpLxTkY0MGovcJoYpYoY40DAE80cW6r0WGxQOr8qjHkOA557w2))/index.aspx?sec=lin&sub=def&erd=83
CAR	CAR Bone Mineral Densitometry Accreditation Program	http://www.car.ca/en/accreditation/bmd.aspx
DEXA Solutions	Link to training and certification	http://www.dexasolutions.com/Resources/Certification.aspx
GE Healthcare (Lunar)	DXA training	http://www3.gehealthcare.com/en/education/product_education_-_technical/lunar_bone_densitometry
Hologic	DXA training	http://www.hologic.com/training/dxa-101-basics-bone-densitometry
Swissray (Norland)	DXA training	http://www.swissray.com/product.php?action=view&cid=16
International Society for Clinical Densitometry	Training courses for DXA certification for clinicians and technologists, facility accreditation	http://www.iscd.org/education/cmece-live-courses/osteoporosis-essentials/ http://www.iscd.org/certification/ http://www.iscd.org/accreditation/
Medical Technology Management Institute	Bone densitometry training course	http://www.mtmi.net/courses/reg_BD.php
OAR	Accredited Densitometry Technologist CME	https://cme.oarinfo.ca/cme/uploaded/2015-CBMD-Tech-ADT-2016-brochure.pdf
OAR	OAR Canadian Bone Mineral Densitometry Facility Accreditation	http://cbmd.ca/
Study.com	Bone density technician training and degree program options	http://study.com/articles/Bone_Density_Technician_Training_and_Degree_Program_Options.html

Note: This is not an all-inclusive list. Other organizations in other countries may have excellent resources as well. Inclusion of programs in this table does not represent an endorsement of the ISCD; the quality of training in preparation for certification and/or accreditation may vary.

Abbr: DXA, dual-energy X-ray absorptiometry; CAR, Canadian Association of Radiologists; CME, continuing medical education; OAR, Ontario Association of Radiologists.

Table 2
Examples of Helpful Books on Bone Densitometry

Bonnick SL, Lewis LA. Bone Densitometry for Technologists, Springer, New York, NY. 2012.
Genant KH. Bone Densitometry and Osteoporosis, Springer, New York, NY. 2011.
Guglielmi G (ed.). Osteoporosis and Bone Densitometry Measurements (Medical Radiology), Springer–Verlag Berlin Heidelberg. 2013.
Hamdy RC, Lewiecki EM. Osteoporosis, Oxford University Press, New York, NY. 2013.
Licata AA, Williams SE. A DXA Primer for the Practicing Clinician: a Case-Base Manual for Understanding and Interpreting Bone Densitometry, Springer, New York, NY. 2013.
Sawyer AJ, Bachrach LK, Fung E. Bone Densitometry in Growing Patients: Guidelines for Clinical Practice, Humana Press, Totowa, NJ. 2007.
Saag KG, Morgan SL, Clines GA. Diagnosis and Management of Osteoporosis, Professional Communications Inc, West Islip, NY. 2013.
The American Society for Bone and Mineral Research, Primer on the Metabolic Bone Diseases and Disorders of Mineral Metabolism, 8th ed, John Wiley & Sons, Ames, IA. 2013.
Dual energy X ray absorptiometry for bone mineral density and body composition assessment. IAEA Human Health Series. No. 15. Vienna: International Atomic Energy Agency. 2010.
Body Composition assessment from birth to two years of age. IAEA Human Health Series. No, 22. Vienna: International Atomic Energy Agency. 2013.

Note: This listing of examples is not exhaustive and is only representative; this does not indicate an endorsement of the ISCD.

fracture risk calculators are a substantial step forward, they are not without limitations. For example, the FRAX calculator requires dichotomous (i.e., yes or no) answers for risk factors, which are actually associated with a range of risks depending on modifying factors such as dose, length of exposure, or severity. Additionally, as the number of prior osteoporosis-related fractures increases or the dose of glucocorticoids rises, the risk of future fractures increases, yet these considerations are not included in the FRAX algorithm. In children, the correlation between BMD and fracture risk is not well established; a FRAX algorithm for the pediatric population does not yet exist.

When Reporting Differences in BMD With Serial Measurements, Only Those Changes That Meet or Exceed the LSC Are Reported as a Change

Rationale. To determine when a difference in DXA measured BMD reflects a true biological change vs a simple measurement variability, each facility needs to calculate its individual LSC. Briefly, this is accomplished by measuring a patient twice on the same day using the same instrument with the scans being performed by the same technologist. When 30 patients (60 scans) have been obtained, the LSC can be calculated using the root mean square standard deviation approach. The LSC can also be calculated using 3 scans obtained on 15 patients. The ISCD and others have developed online calculators to facilitate this process *(54)*. Although calculation of LSC by this method may underestimate long-term measurement variability *(65,66)*, it is a widely used pragmatic approach to patient care.

Comment. Once a center has determined LSC values for the clinically relevant sites (usually L1–L4 spine, total proximal femur, and femoral neck), the LSC values should be applied to serial scans. The LSC should be calculated for other ROIs (e.g., L2–L4, L3–L4, 33% radius, and total radius) if serial comparison for any of these is desired. The ISCD Official Positions include operational details on LSC calculation and reporting *(24)*). For comparison, the current BMD measurement is subtracted from the prior scan and the absolute difference is assessed. If the difference is less than the LSC, this is simply measurement variance and should not be identified as a change. Simply put, a "change" that is not statistically significant is no change and should be reported as such. When the difference between scans is greater than the facility LSC, this change should be reported as an increase or decrease in BMD.

Resources to Support DXA Quality

Resources for education in bone densitometry and the conditions evaluated with DXA technology include scientific journals (e.g., *Journal of Clinical Densitometry*, *Journal of Bone and Mineral Research*, *Osteoporosis International*, *Bone*, *Calcified Tissue International*, and *Journal of Clinical Endocrinology and Metabolism*), instructional courses (Table 1), and books (Table 2). A glossary of DXA terminology and common acronyms is provided in Table 3. The ISCD has a selection of instructional courses devoted to various uses of DXA (e.g., vertebral fracture recognition, pediatric DXA, and body composition testing) and collaborates with the International Osteoporosis Foundation to regularly update a course (Osteoporosis Essentials) in bone densitometry and osteoporosis treatment.

Certification is a procedure by which a third party gives written assurance that a product, process, or service

Table 3
Glossary

Terminology
Acquisition. The process of positioning and scanning the patient on the DXA table.
Accreditation of a certification program. Declaration by a neutral third party (e.g., ANCI, NCCA) that the program meets national and/or international standards for development, implementation, and maintenance of the certification program.
Accreditation of a DXA facility. A process through which a DXA facility is validated as providing quality bone density tests.
Analysis. Assessing and correcting, if necessary, computer default selections for bone edges, regions of interest, and intervertebral space markers; selecting reference databases; and generating data for interpretation.
Artifact. Internal or external factors that can alter the DXA measurements.
Certification. Validation that an individual has acquired a basic level of knowledge on bone densitometry.
Calibration. The process of correcting differences between known reference values and actual measured DXA values.
Fracture risk assessment tool. A validated system for estimating fracture risk in populations.
Interpretation. The process of reviewing the images and data of a DXA scan to provide a diagnosis, assessment of fracture risk, and comparison with any previous studies, while recognizing limitations, if any, in the quality of the test.
Least significant change. The smallest change in BMD that is statistically significant.
Phantom. A standardized object with known BMD that is measured regularly to assess the stability of DXA measurements.
Precision assessment. The methodology of scanning multiple patients more than once that provides the data for calculating the LSC.
Reference database. Data for mean BMD and standard deviation of a defined population that is used to calculate T-scores and Z-scores.
Region of interest. A standardized portion of bone(s) for measuring BMD.
Reporting. The translation of data from acquisition and analysis into a clinically useful report.
Shewhart plot. A graph for recording serial phantom measurements to determine the stability of the DXA system.
Sievert. A derived unit of ionizing radiation dose; 1 Sv = 100 rem (Roentgen equivalent man).
Standard operating procedures. A document that provides necessary information for DXA usage for each DXA facility.
T-score. The standard deviation difference between a patient's BMD and that of a young-adult reference population.
Z-score. The standard deviation difference between a patient's BMD and that of an age-, sex-, and ethnicity-matched reference population.

Acronyms
ANSI. American National Standards Institute
ARRT. American Registry of Radiologic Technologists
ASRT. American Society of Radiologic Technologists
BMD. Bone mineral density
DXA. Dual-energy X-ray absorptiometry
FRAX. WHO fracture risk assessment tool
ISCD. International Society for Clinical Densitometry
ISO. International Organization for Standardization
LSC. Least significant change
NCCA. National Commission for Certifying Agencies
NHANES. National Health and Nutrition Examination Survey
ROI. Region of interest
SOPs. Standard operating procedures
Sv. Sievert
WHO. World Health Organization

conforms to specified requirements. Certification in bone densitometry is provided by organizations such as the American Registry of Radiologic Technologists (for radiological technologists) and the ISCD (for technologists and DXA interpreters).

Accreditation of a professional or personnel certification program provides impartial, third-party validation that the program has met recognized national and international credentialing industry standards for development, implementation, and maintenance of the programs. Agencies that accredit certification programs include the National Commission for Certifying Agencies *(39)*, the American National Standards Institute *(40)*, and others that adhere to principles established by the International Organization for Standardization. The International Organization for Standardization is an independent, nongovernmental international organization with a membership of 162 national standards bodies *(67)*. The ISCD programs for Certified Clinical Densitometrist and Certified Bone Densitometry Technologist are accredited by the National Commission for Certifying Agencies.

Facility accreditation is offered by organizations that include the ISCD *(68)*, Ontario Association of Radiologists *(69)*, Canadian Association of Radiologists *(70)*, the Brazilian College of Radiology, and the Brazilian Association of Bone Health Assessment and Metabolism *(71)*. Programs such as these provide the highest level of assurance that essential elements for quality bone density testing have been implemented at a DXA facility.

Acknowledgments

The ISCD thanks all participants in the many Position Development Conferences resulting in ISCD Official Positions that provided the foundation for DXA Best Practices.

References

1. Bouxsein ML, Seeman E. 2009 Quantifying the material and structural determinants of bone strength. Best Pract Res Clin Rheumatol 23:741–753.
2. Kanis JA. 1994 Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. WHO Study Group. Osteoporos Int 4:368–381.
3. Miller PD, Bonnick SL, Rosen CJ. 1996 Consensus of an international panel on the clinical utility of bone mass measurements in the detection of low bone mass in the adult population. Calcif Tissue Int 58:207–214.
4. Marshall D, Johnell O, Wedel H. 1996 Meta-analysis of how well measures of bone mineral density predict occurrence of osteoporotic fractures. BMJ 312:1254–1259.
5. Cosman F, de Beur SJ, LeBoff MS, et al. 2014 Clinician's guide to prevention and treatment of osteoporosis. Osteoporos Int 25:2359–2381.
6. Grossman JM, Gordon R, Ranganath VK, et al. 2010 American College of Rheumatology 2010 recommendations for the prevention and treatment of glucocorticoid-induced osteoporosis. Arthritis Care Res (Hoboken) 62:1515–1526.
7. Papaioannou A, Morin S, Cheung AM, et al. 2010 2010 Clinical practice guidelines for the diagnosis and management of osteoporosis in Canada: summary. CMAJ 182:1864–1873.
8. Kanis JA, McCloskey EV, Johansson H, et al. 2013 European guidance for the diagnosis and management of osteoporosis in postmenopausal women. Osteoporos Int 24:23–57.
9. Compston J, Bowring C, Cooper A, et al. 2013 Diagnosis and management of osteoporosis in postmenopausal women and older men in the UK: National Osteoporosis Guideline Group (NOGG) update 2013. Maturitas 75:392–396.
10. Wang M, Bolland M, Grey A. 2015 Management recommendations for osteoporosis in clinical guidelines. Clin Endocrinol doi:10.1111/cen.13000; (Oxf) Epub.
11. Kanis JA, Oden A, Johansson H, et al. 2009 FRAX(R) and its applications to clinical practice. Bone 44:734–743.
12. Lewiecki EM, Laster AJ. 2006 Clinical applications of vertebral fracture assessment by dual-energy X-ray absorptiometry. J Clin Endocrinol Metab 91:4215–4222.
13. Shepherd JA, Baim S, Bilezikian JP, Schousboe JT. 2013 Executive summary of the 2013 International Society for Clinical Densitometry Position Development Conference on Body Composition. J Clin Densitom 16:489–495.
14. Broy SB, Cauley JA, Lewiecki ME, et al. 2015 Fracture risk prediction by non-BMD DXA measures: the 2015 ISCD Official Positions part 1: hip geometry. J Clin Densitom 18: 287–308.
15. Silva BC, Broy SB, Boutroy S, et al. 2015 Fracture risk prediction by non-BMD DXA measures: the 2015 ISCD Official Positions part 2: trabecular bone score. J Clin Densitom 18:309–330.
16. Lewiecki EM, Lane NE. 2008 Common mistakes in the clinical use of bone mineral density testing. Nat Clin Pract Rheumatol 4:667–674.
17. Lewiecki EM, Binkley N, Petak SM. 2006 DXA quality matters. J Clin Densitom 9:388–392.
18. Watts NB. 2004 Fundamentals and pitfalls of bone densitometry using dual-energy X-ray absorptiometry (DXA). Osteoporos Int 15:847–854.
19. Messina C, Bandirali M, Sconfienza LM, et al. 2015 Prevalence and type of errors in dual-energy X-ray absorptiometry. Eur Radiol 25:1504–1511.
20. Kim TY, Schafer AL. 2016 Variability in DXA Reporting and other challenges in osteoporosis evaluation. JAMA Intern Med 176:393–395.
21. Fenton JJ, Robbins JA, Amarnath AL, Franks P. 2016 Osteoporosis overtreatment in a regional health care system. JAMA Intern Med 176:391–393.
22. Gafni RI, Baron J. 2004 Overdiagnosis of osteoporosis in children due to misinterpretation of dual-energy X-ray absorptiometry (DEXA). J Pediatr 144:253–257.
23. Bonnick SL, Lewis LA. 2013 Bone Densitometry for Technologists. New York, London: Springer.
24. Shepherd JA, Schousboe JT, Broy SB, et al. 2015 Executive summary of the 2015 ISCD position development conference on advanced measures from DXA and QCT: fracture prediction beyond BMD. J Clin Densitom 18:274–286.
25. MultiState Associates Incorporated. Semi-weekly reports to the International Society for Clinical Densitometry on state legislative and regulatory activity relating to DXA or bone density studies. 2006–2015; Data on file.
26. Centers for Medicare & Medicare Services. 2011 Code of Federal Register. Billing and Enrollment Guideline for PHYS-078 Independent Diagnostic Testing Facilities (IDTF).

Available at: https://downloads.cms.gov/medicare-coverage-database/lcd_attachments/23448_6/CBGPHYSMED078April09.pdf. Accessed March 15, 2016.
27. Lenchik L, Leib ES, Hamdy RC, et al. 2002 Executive summary: International Society for Clinical Densitometry Position Development Conference, Denver, Colorado, July 20–22, 2001. J Clin Densitom 5(Suppl):S1–S3.
28. The Writing Group for the International Society for Clinical Densitometry Position Development Conference. 2004 2004 Executive summary: International Society for Clinical Densitometry Position Development Conference. J Clin Densitom 7:7–12.
29. Binkley N, Bilezikian JP, Kendler DL, et al. 2006 Official positions of the International Society for Clinical Densitometry and executive summary of the 2005 Position Development Conference. J Clin Densitom 9:4–14.
30. Baim S, Binkley N, Bilezikian JP, et al. 2008 Official positions of the International Society for Clinical Densitometry and executive summary of the 2007 ISCD Position Development Conference. J Clin Densitom 11:75–91.
31. Gordon CM, Bachrach LK, Carpenter TO, et al. 2008 Dual energy X-ray absorptiometry interpretation and reporting in children and adolescents: the 2007 ISCD pediatric official positions. J Clin Densitom 11:43–58.
32. Hans DB, Kanis JA, Baim S, et al. 2011 Joint official positions of the International Society for Clinical Densitometry and International Osteoporosis Foundation on FRAX((R)) executive summary of the 2010 Position Development Conference on Interpretation and Use of FRAX((R)) in Clinical Practice. J Clin Densitom 14:171–180.
33. Schousboe JT, Shepherd JA, Bilezikian JP, Baim S. 2013 Executive summary of the 2013 International Society for Clinical Densitometry Position Development Conference on Bone Densitometry. J Clin Densitom 16:455–466.
34. Crabtree NJ, Arabi A, Bachrach LK, et al. 2014 Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: the revised 2013 ISCD pediatric official positions. J Clin Densitom 17:225–242.
35. Fitch K, Bernstein SJ, Aguilar B, et al. 2001 The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, California: The RAND Corporation.
36. Gordon CM, Leonard MB, Zemel BS, International Society for Clinical Densitometry. 2014 2013 Pediatric position development conference: executive summary and reflections. J Clin Densitom 17:219–224.
37. Morgan SL, Lopez-Ben R, Nunnally N, et al. 2008 The effect of common artifacts lateral to the spine on bone mineral density in the lumbar spine. J Clin Densitom 11:243–249.
38. El Maghraoui A, Roux C. 2008 DXA scanning in clinical practice. QJM 101:605–617.
39. Institute for Credentialling Excellence. 2014 NCCA Accreditation. Available at: http://www.credentialingexcellence.org/ncca. Accessed March 15, 2016.
40. American National Standards Institute. 2016 Accreditation Services. Available at: https://www.ansica.org/wwwversion2/outside/Default.asp. Accessed March 15, 2016.
41. Frimeth J, Galiano E, Webster D. 2010 Some physical and clinical factors influencing the measurement of precision error, least significant change, and bone mineral density in dual-energy X-ray absorptiometry. J Clin Densitom 13:29–35.
42. Cheng XG, Nicholson PH, Boonen S, et al. 1997 Effects of anteversion on femoral bone mineral density and geometry measured by dual energy X-ray absorptiometry: a cadaver study. Bone 21:113–117.
43. Donnelly LF, Dickerson JM, Goodfriend MA, Muething SE. 2009 Improving patient safety: effects of a safety program on performance and culture in a department of radiology. AJR Am J Roentgenol 193:165–171.
44. Larson DB, Kruskal JB, Krecke KN, Donnelly LF. 2015 Key concepts of patient safety in radiology. Radiographics 140277.
45. International Society for Clinical Densitometry. 2015 Test Study Material. Available at: http://www.iscd.org/certification/test-study-material. Accessed March 15, 2016.
46. Baim S, Wilson CR, Lewiecki EM, et al. 2005 Precision assessment and radiation safety for dual-energy X-ray absorptiometry: position paper of the International Society for Clinical Densitometry. J Clin Densitom 8:371–378.
47. Khan AA, Colquhoun A, Hanley DA, et al. 2007 Standards and guidelines for technologists performing central dual-energy X-ray absorptiometry. J Clin Densitom 10:189–195.
48. Ramalingaswami V. Iodine and thyroid cancer. 1969;12:111.
49. The Writing Group for the International Society for Clinical Densitometry Position Development Conference. 2004 2004 International Society for Clinical Densitometry Position Development Conference. Technical standardization for dual-energy X-ray absorptiometry. J Clin Densitom 7:27–36.
50. Pearson D, Cawte SA. 1997 Long-term quality control of DXA: a comparison of Shewhart rules and Cusum charts. Osteoporos Int 7:338–343.
51. Bonnick SL. 2010 Bone Densitometry in Clinical Practice: Application and Interpretation. New York, NY: Humana.
52. Bennett HS, Dienstfrey A, Hudson LT, et al. 2006 Standards and measurements for assessing bone health-workshop report co-sponsored by the International Society for Clinical Densitometry (ISCD) and the National Institute of Standards and Technology (NIST). J Clin Densitom 9:399–405.
53. Lenchik L, Kiebzak GM, Blunt BA, International Society for Clinical Densitometry, Position Development Panel, Scientific Advisory Committee. 2002 What is the role of serial bone mineral density measurements in patient management? J Clin Densitom 5(Suppl):S29–S38.
54. International Society for Clinical Densitometry. 2015 ISCD Calculators. Available at: http://www.iscd.org/resources/calculators/. Accessed March 15, 2016.
55. Shepherd JA, Lu Y, Wilson K, et al. 2006 Cross-calibration and minimum precision standards for dual-energy X-ray absorptiometry: the 2005 ISCD official positions. J Clin Densitom 9:31–36.
56. Leslie WD, Manitoba Bone Density Program. 2006 The importance of spectrum bias on bone density monitoring in clinical practice. Bone 39:361–368.
57. Boonen S, Kaufman JM, Reginster JY, Devogelaer JP. 2003 Patient assessment using standardized bone mineral density values and a national reference database: implementing uniform thresholds for the reimbursement of osteoporosis treatments in Belgium. Osteoporos Int 14:110–115.
58. Shepherd JA, Morgan SL, Lu Y. 2008 Comparing BMD results between two similar DXA systems using the generalized least significant change. J Clin Densitom 11:237–242.
59. Peel NF, Johnson A, Barrington NA, et al. 1993 Impact of anomalous vertebral segmentation on measurements of bone mineral density. J Bone Miner Res 8:719–723.
60. Binkley N, Krueger D. 2009 What should DXA reports contain? Preferences of ordering health care providers. J Clin Densitom 12:5–10.
61. Soen S, Fukunaga M, Sugimoto T, et al. 2013 Diagnostic criteria for primary osteoporosis: year 2012 revision. J Bone Miner Metab 31:247–257.

62. Schuit SC, van der Klift M, Weel AE, et al. 2004 Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. Bone 34:195–202.
63. Wainwright SA, Marshall LM, Ensrud KE, et al. 2005 Hip fracture in women without osteoporosis. J Clin Endocrinol Metab 90:2787–2793.
64. Garvan Institute. 2016 Fracture Risk Calculator. Available at: http://www.garvan.org.au/bone-fracture-risk/. Accessed March 15, 2016.
65. Hangartner TN. 2007 A study of the long-term precision of dual-energy X-ray absorptiometry bone densitometers and implications for the validity of the least-significant-change calculation. Osteoporos Int 18:513–523.
66. Leslie WD. 2008 Factors affecting short-term bone density precision assessment and the effect on patient monitoring. J Bone Miner Res 23:199–204.
67. International Organization for Standardization. 2016 Standards. Available at: http://www.iso.org/iso/home/standards.htm. Accessed March 15, 2016.
68. International Society for Clinical Densitometry. 2014 Facility Accreditation. Available at: http://www.iscd.org/accreditation/. Accessed March 15, 2016.
69. Ontario Association of Radiologists. 2016 Canadian BMD Facility Accreditation Program. Available at: https://oarinfo.ca/cbmd/resources. Accessed March 15, 2016.
70. Canadian Association of Radiologists. 2016 Accreditation Programs: Bone Mineral Density (BMD). Available at: http://www.car.ca/en/accreditation/bmd.aspx. Accessed March 15, 2016.
71. Associação Brasileira de Avaliação Óssea e Osteometabolismo. 2005 Programma de Controle de Qualidade em Densitometria Clinica. Available at: http://www.proquad.org.br/index_proquad.php. Accessed March 15, 2016.

APPENDIX 4:

2018 ASBMR Abstract

Change in BMD as a Surrogate for Fracture Risk Reduction in Osteoporosis Trials: Results from Pooled, Individual-level Patient Data from the FNIH Bone Quality Project

Dennis M. Black, Eric Vittinghoff, Richard Eastell, Douglas Bauer, Li-Yung Lui, Lisa Palermo, Charles McCulloch, Jane Cauley, Sundeep Khosla, Fernando Marin, Anne DePapp, Andreas Grauer, Mary Bouxsein

To facilitate future drug development in osteoporosis, the FNIH Bone Quality Project aims to identify a surrogate that can be used in place of fracture in future clinical trials. To do so, we compiled and pooled individual patient data from 25 randomized, placebo-controlled osteoporosis trials that enrolled >125,000 subjects. Changes in BMD are an obvious candidate surrogate. One step in the validation of a surrogate biomarker is to estimate the percent of treatment explained (PTE) by the surrogate. Previous analyses of individual trials reported PTEs for BMD change ranging from 4% to ~80%, but analysis methods differed, and single studies were too small to give precise estimates.

We used our pooled data to estimate PTE for BMD change, using a two-stage procedure. In the first stage, we used linear mixed models (LMM) to estimate participant-specific BMD trajectories. In the second, we fit nested pooled logistic regression (PLR) models for the effect of treatment on time to vertebral, non-vertebral and hip fractures. PTE was calculated as the percentage change in the coefficient for treatment after adding time-dependent BMD change, based on the first-stage LMM, to the base PLR model. Confidence intervals for PTE were obtained using the method of seemingly unrelated regressions. We analyzed PTE for femoral neck, total hip and lumbar spine BMD for vertebral, hip and non-vertebral fractures including people with at least one BMD measurement. We examined the change in BMD at 24 months, and included all fractures to the end of the trials.

The current analyses included 21 trials (including bisphosphonates, SERMs, estrogen, odanacatib, denosumab, PTH1-34, and PTH1-84) with 83,395 subjects with DXA who suffered 4515 vertebral, 6608 non-vertebral and 873 hip fractures. We found that PTE for 24 month change in femoral neck BMD was 65% (95% CI: 53-77%) for vertebral, 74% (95% CI: 43-105%) for non-vertebral and 65% (24-106%) for hip fractures (see Table). The results were similar for total hip, but PTE was lower for lumbar spine BMD.

Our results indicate that changes in total hip and femoral neck BMD explain a large proportion of the treatment-related reduction in fracture risk for all 3 fracture types. Importantly, this analysis included therapies with differing mechanisms of action. These data, along with others generated from this project, support the potential value of DXA BMD change as a surrogate for fractures in future trials.

Table. Percent of treatment effect (%, 95% CI) explained by changes in total hip (TH), femoral neck (FN) and lumbar spine (LS) at 24 months for vertebral, hip and non-vertebral fractures.

	ΔTH BMD	ΔFN BMD	ΔLS BMD
Vertebral Fx	66% (55, 77%) p<0.0005	65% (53, 77%) p<0.0005	27% (14, 40%) p<0.0005
Hip Fx	60% (27, 93%) p<0.0005	65% (24, 106%) p=0.002	42% (7, 76%) p=0.019
Non-vertebral Fx	69% (41, 96%) p<0.0005	74% (43, 105%) p<0.0005	56% (25, 87%) p<0.0005

BIOCLINICA®

CHARTER FOR INDEPENDENT IMAGING ASSESSMENT

DXA, X-ray, QCT, HR-pQCT

Merck Protocol MK08220-018, 032, 035

A Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium.

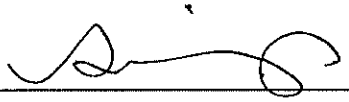
In conjunction with An Imaging Sub-Study and A Sub-Study to Explore Biomarkers of Physical Function

BioClinica Studies 1223, 1831, 1966

Date: 26-Sep-2014

Version: 2.0

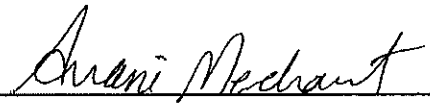
SPONSOR SIGNATURE PAGE



10/8/14

Sui Yan
Clinical Trial Manager

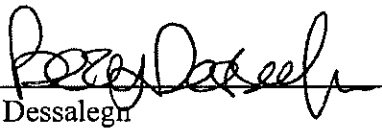
Date



10/8/14

Avani Desai_Merchant
Scientist, Clinical Research

Date



08-Oct-14

Bezy Dessalegn
Clinical Development Scientist

Date

BIOCLINICA SIGNATURE PAGE

Kenneth W Gaither, MS
Director, Osteoporosis
02 Oct 2014
Date

Klaus Engelke, PhD
Director, Special Imaging
04 Oct 2014
Date

Thomas Fuerst, PhD
Chief Science Officer
29-SEP-2014
Date

Gabriele von Ingersleben, MD, PhD
Senior Director Radiology
26 SEP 2014
Date

Emily Duncan
Clinical Project Manager
06-OCT-2014
Date

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations and Acronyms

BMC	Bone Mineral Content
BMD	Bone Mineral Density
BSQ	Binary Semi-quantitative
CD-ROM	Compact Disk – Read Only Memory
CFR	Code of Federal Regulation
eCRF	Electronic Case Report Form
CRO	Contract Research Organization
CT	Computed Tomography
CUSUM	Cumulative Sum
CV%	Percent Coefficient of Variation
DAP	Data Analysis Plan
DCF	Data Clarification Form
DICOM	Digital Imaging and Communications in Medicine
DVD	Digital Video Disc
DTS	Data Transfer Specification
DXA	Dual Energy X-ray Absorptiometry
EFP	European Forearm Phantom
ESP	European Spine Phantom
GCP	Good Clinical Practice
HAM	BioClinica Hamburg Office
HR-pQCT	High Resolution peripheral Quantitative Computed Tomography
HSQP	Hip/Spine QC Phantom
ID	Identification
MRI	Magnetic Resonance Imaging
NWK	BioClinica Newark Office
PDX	BioClinica Portland Office
QA	Quality Assurance
QC	Quality Control
QCT	Quantitative Computed Tomography
SAP	Statistical Analysis Plan
SOP	Standard Operating Procedure
SQ	Genant Semiquantitative
SSP	Study Specific Procedures

1 EXECUTIVE SUMMARY

The purpose of this charter is to describe the procedures for the acquisition, collection, handling, quality review, and analysis of DXA, X-ray, QCT and HR-pQCT images as dictated by the three Merck Protocols:

- MK0822-018 "A Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium"
- MK0822-032 "An Imaging Sub-Study of the Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium"
- MK0822-035 "A Sub-Study to Explore Biomarkers of Physical Function in the Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium"

Additional details can be found in the BioClinica standard operating procedures (SOPs), study specific procedures (SSPs) and site imaging manuals.

Imaging endpoints for the MK0822-018 study will be used to assess the effect of Odanacatib treatment compared to placebo on the risk of morphometrically assessed vertebral fractures; on the risk of hip fractures; and on the risk of clinical non-vertebral fractures. For MK0822-032, odanacatib treatment will be evaluated for the percent change from baseline in trabecular volumetric bone mineral density (vBMD) at the lumbar spine. Lastly, MK0822-035 will use whole body DXA to monitor longitudinal changes in appendicular lean body mass.

1.1 Protocol Summary

MK0822-018 –

A Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium.

- Primary objectives include assessing the effect of treatment with odanacatib 50 mg once weekly on the risk of morphometrically assessed vertebral fractures; on the risk of hip fractures: and on the risk of clinical non-vertebral fractures compared to placebo. (Note: for the purposes of this study, non-vertebral fractures **exclude fractures of the fingers, toes, face, and skull.**)
- Secondary objectives include assessing the effect of treatment with odanacatib 50 mg once weekly on the risk of clinical vertebral fractures and on the lumbar spine, total hip, femoral neck, trochanter and distal forearm BMD compared to placebo.

- Exploratory objectives include assessing the effect of treatment with odanacatib 50 mg once weekly on total body BMD and on skeletal microarchitecture as assessed by quantitative 2-D histomorphometry and 3-D μ-CT on transilial bone biopsy specimens compared to placebo

MK0822-032

An Imaging Sub-Study of the Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium.

- The primary objective is to evaluate the effect of treatment with MK-0822 50 mg once weekly (OW) on the percent change from baseline in trabecular volumetric bone mineral density (vBMD) at the lumbar spine (assessed by QCT) compared to placebo at Month 24.
- Secondary objectives include evaluating the effect of treatment with MK-0822 50 mg OW on the percent change from baseline in cortical vBMD at the hip (assessed by QCT) and BMD at the lumbar spine, hip and hip sub-regions and 1/3 distal forearm (assessed by DXA) compared to placebo at Month 24.
- The Exploratory Objectives are to estimate the effect of treatment with MK-0822 50 mg OW on cortical and trabecular parameters [e.g., cortical thickness (CTh) and trabecular thickness (TbTh)] at the 1/3 distal tibia and the ultra distal tibia (assessed by HR-pCT); at the non-dominant 1/3 distal radius and the non-dominant ultra distal radius (assessed by HR-pCT); the correlation between the percent change from baseline in cortical volume at hip (assessed by QCT) and the percent change from baseline in the 1/3 distal tibia and the ultra distal tibia cortical thickness (assessed by HR-pCT); the correlation between the percent change from baseline in vBMD at the non-dominant 1/3 distal radius and the non-dominant ultra distal radius (assessed by HR-pCT) and the percent change from baseline in aBMD at the 1/3 distal forearm (assessed by DXA); on strength (ultimate load and load-to-strength ratio [Φ]) at the hip, spine, non-dominant 1/3 distal radius, non-dominant ultra distal radius, 1/3 distal tibia and ultra distal tibia (assessed by finite elemental modeling [FEM] using data from QCT and HR-pCT images); and on skeletal microarchitectural parameters [e.g., trabecular thickness (TbTh)] assessed by 2-D histomorphometry and 3-D μ-CT on transilial bone biopsy specimens compared to placebo at Month 24.

MK0822-035

- A Sub-Study to Explore Biomarkers of Physical Function in the Phase III Randomized, Placebo-Controlled Clinical Trial to Assess the Safety and Efficacy of Odanacatib (MK-0822) to Reduce the Risk of Fracture in Osteoporotic Postmenopausal Women Treated With Vitamin D and Calcium.
- An exploratory objective is to perform multiple and partial correlation analyses to estimate the relationship among the following variables: longitudinal changes in appendicular lean body mass, measured by total body DXA,

longitudinal changes in the SPPB score, longitudinal changes in gait speed, and longitudinal changes in the chair stand test.

2 INTRODUCTION

2.1 Roles and Responsibilities

2.1.1 Sponsor

Merck will oversee the execution of the trials including all regulatory interactions and approvals, definition of study timelines, statistical analysis, and reporting of study results. Merck will provide the clinical trial protocols, and any related protocol clarification correspondences to include letters/memos to BioClinica.

Merck will provide additional information about the clinical design, outcomes, or execution when questions arise. The Merck study managers are the first point of contact for escalated issues. Any issues that cannot be resolved at the study team level will be raised to the Imaging Operational Governance Team.

2.1.2 Imaging Core Laboratory

BioClinica will have responsibility for the imaging components of the trial, including the following:

- Standardization of imaging protocols, procedure manuals, and training of the investigator site staff
- Collection, quality review, and archival of study images
- Digitization of film-based image data and preparation of digital image data for blinded review
- Central analysis of DXA scans
- Central reading of morphometric vertebral fracture
- Central reading of non-vertebral fractures
- Central analysis of QCT images
- Central analysis of HR-pQCT images
- Provide QCT and HR-pCT images to ON Diagnostics and Merck &Co. Inc for respective FEA Analysis.
- Central confirmation of subject eligibility
- Central Safety BMD monitoring
- Monitoring of DXA scanner calibration
- Monitoring of CT scanner calibration
- Transfer of analysis results to Merck

The following employees of BioClinica are key scientific and operational personnel who will supervise the activities of BioClinica's collection, quality control, and central reading of the imaging exams:

Name	Title
Thomas Fuerst, PhD	Chief Science Officer
Kenneth W. Gaither, MS	Director, Osteoporosis
Gabriele von Ingersleben, MD, PhD	Senior Director, Radiology
Elsa Griffith	Director, Clinical Operations, Portland
Klaus Engelke, PhD	Director, Special Imaging

These personnel will assemble a project team with the appropriate ability and experience to execute the activities outlined for the independent review of the imaging exams. The BioClinica project team will participate in the project from the pre-study planning stage to the completion of the image review and transfer of data to Merck.

All activities described by this charter will be conducted either at the investigator site (on-site training activity) or at the offices of BioClinica, Inc., with the exception of Investigator Meetings, if applicable. The BioClinica offices participating in the study are:

BioClinica Office	Address/Phone
Newark Office (NWK)	BioClinica, Inc. 7707 Gateway Boulevard Building 5, 3rd Floor Newark, CA 94560 USA Tel: +1.415.817.8900
Portland Office (PDX)	BioClinica, Inc 11731 NE Glenn Widing Dr. Portland, OR 97220 USA Tel: +1.503.528.7800
Hamburg Office (HAM)	BioClinica A/S Deutschland Kaiser-Wilhelm-Str. 89 20355 Hamburg Tel: +49 40 8740 8258

The BioClinica Global Project Manager is located in Portland, Oregon.

Imaging data will be submitted to, and processed and analyzed at the three BioClinica offices noted above. These BioClinica offices will also provide administrative, technical, and engineering support for the study.

Image collection and analysis will be conducted by region and modality as described in Table 1.

Table 1: Image Collection and Analysis by Region and Modality

BioClinica Office	DXA	X-ray	QCT	HR-pQCT
NWK	–	WW**	WW**	
HAM	–	EU*	EU*	WW
PDX	WW	–	–	–

**EU: all European countries from study start through Jan-2012; **WW: all investigator sites world-wide after Feb-2012.*

2.1.3 Miscellaneous Study Communication Procedures

Sections 2.1.3.1 and 2.1.3.2 are presented as guidelines with regard to communication regarding this document, the independent reviewers, and any regulatory authority. Additionally each study has its own separate Communication Plan.

2.1.3.1 Review and Approval of Charter

This charter was written by BioClinica in consultation with Merck and is consistent with the clinical trial protocol and the statistical analysis plan (SAP). BioClinica

agrees to follow the procedures described therein. This charter has been reviewed and approved by BioClinica and Merck. Additionally, it is anticipated that this charter may be submitted to regulatory authorities and approved according to local regulations.

2.1.3.2 Meeting Schedule

As needed, charter development meetings will be scheduled between BioClinica and Merck. Additionally, regularly occurring meetings between the BioClinica project manager and the Merck study team will be scheduled. These meetings will include discussions of: imaging data collection and analysis, study deliverables, schedule and data transfers.

2.2 Role of Imaging in Study Design

2.2.1 Trial Endpoints

In these three trials imaging endpoints will be used as surrogates for bone strength, fracture risk, and lean body mass changes. MK0822, a cathepsin-K inhibitor, is an antiresorptive therapy being investigated as a treatment for osteoporosis. MK0822 is expected to increase bone mineral density (BMD) and perhaps improve bone geometry and bone microarchitecture. These changes should improve bone strength and consequently reduce the risk of incident fracture.

X-ray imaging will serve as evidence for one of the primary endpoints, evaluating risk of morphometrically assessed vertebral fractures. Standardized lateral radiographs of the lateral lumbar and lateral thoracic spine will be read centrally using the Genant semi-quantitative scale. In cases in which an incident vertebral fracture is determined via semi-quantitative methods to be present then the spine radiograph will then be evaluated via a quantitative (morphometric) method.

Imaging will also play a role in the determination of the secondary endpoints of non-vertebral fracture and BMD change. Change in BMD from baseline will be assessed using DXA at the lumbar spine, hip, total body and forearm. DXA scans at defined study visits will be used to determine BMD at each time point and to evaluate percent change from baseline.

In addition, QCT will be used to measure changes in BMD and bone geometry at the spine and hip. Similarly, HR-pQCT will be used to measure changes at the radius and tibia in bone density, geometry and microarchitecture as well as estimated failure load from FEA.

2.2.2 Subject Eligibility

BioClinica will communicate to investigators the results of the central analysis of the screening DXA femur scans, and the screening lateral thoracic and lateral lumbar spine radiographs via fax or email. The results will be reported within 5 business days of receipt of a complete and correct DXA and x-ray package for each subject. The investigator will use the results of BioClinica's central analysis to determine subject eligibility. This data will be loaded into the Interactive Voice Recognition System (IVRS) during the enrollment period to safeguard against randomizing an ineligible patient (with regard to DXA T-scores and prevalent vertebral fracture status).

2.2.3 Subject Safety

Bone Loss

Investigators will be alerted if a subject experiences a BMD loss from baseline of 7% or more at the lumbar spine (total spine region of interest) or proximal femur (total hip region of interest) at any time during the study.

On-going during the study, BioClinica will monitor changes in BMD in order to identify subjects who meet the bone loss criteria defined by the clinical trial protocol. In the event a case is identified, BioClinica will review the DXA scans to confirm the validity of the BMD decrease and request a second scan for bone loss confirmation. The results from the second scan will be averaged with the first scan to confirm bone loss. Once confirmed, BioClinica will communicate these findings to the investigator and Merck. The investigator, study coordinator and monitor will be notified by fax or email within 30 days of receipt of the analyzable scan and all associated information. Merck will be notified by a weekly report sent to the Merck Study Manager. BioClinica will obtain acknowledgment from the study site of receipt of the bone loss notification.

When the investigator is notified of a subject who experiences an excessive decrease in BMD, the investigator is required to discuss alternative treatment options with the subject. Please see the clinical trial protocol for additional details.

Calculation of bone loss assumes that the subject has been scanned on the same scanner throughout the study. Changes in scanners require additional calibration checks to be implemented and correction factors applied when indicated. Additionally, if a replacement scanner of a different manufacturer is used, due to differences between manufacturers, safety checks should be calculated as the sum of the total percent change from the first to last scan on the original scanner and the total percent change from the first to last scan on the replacement scanner, when the replacement scanner is of a different manufacturer (see section 3.1.7.1).

For the End of Study (EOS) visit only single scan confirmations will be performed.

2.2.4 BioClinica Deliverables

BioClinica will deliver results of the independent review defined herein to Merck at a mutually agreed upon frequency and time using a pre-defined data transfer format. Details of the data transfer format can be found in the data transfer specification.

3 IMAGE ACQUISITION AND COLLECTION

3.1 Description of On-Protocol Imaging

All DXA, QCT and HR-pQCT scans acquired on approved scanners, whether they are at protocol defined visits scheduled for imaging examinations or at unscheduled, intermediate time points, are considered to be on-protocol imaging.

Similarly, all spine radiographs acquired at protocol defined visits or at unscheduled times to investigate subject reports of back pain are considered on-protocol imaging.

3.1.1 Acceptable Imaging Modalities

DXA will be used to measure BMD in this study. Lateral spine radiographs will be used to assess prevalent, new incident, and worsening incident vertebral fractures. QCT scans will be used to measure BMD and bone geometry at the spine and hip. HR-pQCT will be used to measure changes in bone density, geometry, microarchitecture and estimated failure load at the radius and tibia.

All images will be prospectively collected by BioClinica. DXA, QCT and HR-pQCT scans must be submitted as electronic images in approved digital formats. Plain radiographs are acceptable for lateral spine images.

For each subject, the same DXA, QCT or HR-pQCT scanner must be used for all scan acquisitions. If a scanner is changed during the course of the study, BioClinica will take steps, when possible, to quantify and address calibration differences between the new and old scanner.

3.1.1.1 Eligible DXA Scanners

Only sites with more recent model Hologic and GE Lunar DXA scanners will be allowed to participate in the trial (see Table 2). This will minimize problems such as repairs of aging scanners and investigator sites changing scanners during the course of the study. The decision to include a scanner model in the study is based on information from the manufacturer about plans for service and future support and BioClinica's experience with scanner reliability. Exceptions are possible in special circumstances and will be considered on a case-by-case basis. However, exceptions will be discussed between Merck and BioClinica before deciding if the exception will be allowed.

Table 2: List of Acceptable DXA Scanners

Hologic	GE Lunar	Norland
Delphi C	DPX-NT	XR-26
Delphi W	DPX-MD+	XR-36
Delphi SL	DPX-Bravo (DPX-BT)	
Delphi A	DPX-Duo	
Discovery C	DPX-Pro	
Discovery W	Prodigy	
Discovery SL	iDXA	
Discovery A		
Discovery Ci		
Discovery Wi		
Explorer		
Explorer w/Whole Body		
QDR 4500 Series		
QDR-1000W		
QDR-2000		
Horizon		

All scanners in the above table are acceptable models for reliably measuring BMD. This list will be evaluated and updated as-needed.

3.1.1.2 Eligible QCT Scanners

CT scanners will be evaluated and qualified by BioClinica prior to participation in the study. CT scanners used for QCT measurements should meet the following requirements:

- Spiral (helical) acquisition
- Four (4) detector rows or more
- Able to submit images in electronic form (DICOM 3.0)
- Slice thickness of 1 - 1.25 millimeters
- Able to execute the acquisition protocol outlined in the QCT procedure manual
- Able to store CT raw data up to 2 weeks (so that new reconstructions can be provided, if necessary)

3.1.1.3 Eligible HR-pQCT Scanners

Only XtremeCT scanners (SCANCO Medical AG, Switzerland) can be used to perform HR-pQCT scans in this study.

3.1.2 Anatomical Coverage

Radiographs of the lateral thoracic and lumbar spine will include coverage of T3 to S1. Radiographs acquired at unscheduled visits to investigate subject reports of back pain might be targeted to the area of discomfort and may have more limited coverage.

All subjects will undergo DXA bone density assessments of the lumbar spine and hip. Lumbar spine scans must include L1 through L4. Hip scans will include the entire proximal femur to about 2 cm below the lesser trochanter. For subjects enrolled in the DXA sub-study, DXA bone density assessments will also be performed at the distal radius and total body. Distal radius DXA scans will be performed in duplicate at all time points.

For proximal femur and distal radius, the left side must be used for all scans at all visits. If the right side must be used (e.g., hip implant) or is inadvertently used at baseline, then it must be used consistently throughout the study. If a subject fractures the hip (or radius) being examined during the study, no further DXA scans of either hip (or radius) will be acquired.

QCT scans will be acquired on the lumbar spine (L1-L2) and the hip. If at baseline L1 or L2 is fractured, has metal implants in the scan field or has another anatomical deformity such as a large Schmorl's node, scanning will be shifted up (T12-L1) or down (L2-L3) to avoid the affected vertebra. All follow-up scans will be performed on the same spinal segment or limb. In case one of the two vertebrae scanned at baseline becomes inaccessible during the study only the other vertebrae will be scanned after the event. If a subject fractures both vertebrae (and the hip) being examined, no further QCT scans of the spine (or hip) will be performed. By default the left hip will be analyzed unless a condition is present which interferes with the assessment, then the right hip will be analyzed. The same hip will be analyzed at all visits.

HR-pQCT scans will be done on the non dominant arm and left leg. If the target limb was previously fractured, the contralateral side should be selected, matching the same side scanned by DXA. All follow-up scans will be performed on the same limb. The default scan location as defined by the manufacturer will be analyzed. If a subject fractures the arm (or leg) being examined, no further HR-pQCT scans of either arm (or leg) will be performed.

3.1.3 Imaging Time Points / Schedule

Imaging time-points and visit schedule are shown in Table 3.

Table 3: Imaging Time Points and Visit Schedule

Imaging Modality/ Anatomical Area	Screening	Rand'n	Month 6	Month 12	Month 24	Month 36	Month 48***	Month 60	Month 72	Month 84	Month 96	Month 108	Month 120	Month 132	Month 144	EXCV	EX2CV	EOS
MK0822-018																		
DXA Hip	X		X *	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DXA Spine	X		X *	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DXA 1/3 Distal Forearm**		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
DXA Total Body**		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Lateral Spine X-ray (lumbar & thoracic)	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
MK0822-032																		
QCT Hip		X		X	X	X	X	X										X
QCT Spine		X		X	X	X	X	X										X
HR-pQCT Radius		X		X	X	X	X	X										X
HR-pQCT Tibia		X		X	X	X	X	X										X
DXA 1/3 Distal Forearm		X		X	X	X	X	X										X
MK0822-035																		
DXA Total Body		X		X	X	X	X											X

All study visits have a ±3-week window.

*Only when required by regulatory authority.

** Only completed for 10% substudy

*** Since this is an event driven study not all subjects will complete the Month 48 visit.

3.1.4 Summary of Image Acquisition Procedures

BioClinica will prepare a manual of image acquisition procedures customized to the study. The manual will include:

- Techniques for accurate and reproducible subject positioning, proper anatomical coverage, imaging parameters, and printing of film images
- Instructions for labeling electronic media with unique, anonymous subject identifiers
- Instructions for performing routine quality control
- Information on archiving of digital data
- Instructions on completing the transmittal form and data transfer to BioClinica

BioClinica will distribute the manual to the study technologists and study coordinators at each site. The BioClinica project team will ensure that all technologists participating in the study have received and read the procedure manuals and have completed the sign-off page and submitted it to BioClinica. However, in instances where technologists have been registered based upon previous BioClinica training and experience they will not be required to be re-trained by BioClinica on DXA or radiographic technical procedures for this specific protocol. Training on DXA, QCT, HR-pQCT and radiographic procedures will be conducted prior to commencing image acquisition at study start or when needed, before a new technologist begins to acquire images for the study.

BioClinica will also create a laminated quick reference guide for use by the technologists and study coordinators which will summarize the primary requirements for the acquisition of scans and the handling of imaging data.

3.1.5 Site Qualification Procedures

Before acquiring scans for this protocol, DXA, QCT, HR-pQCT and x-ray technologists will need to be registered with BioClinica. The registration process will involve completion of a site questionnaire and training provided by BioClinica (as needed).

3.1.5.1 Imaging Site Qualification Process

BioClinica will administer site questionnaires to each imaging site identified by Merck for participation in the study. The questionnaires will assess the following information for each site:

- Contact person
- Manufacturer and model of radiographic, QCT and DXA equipment to be used in the study
- Version of scanner software for DXA and HR-pQCT
- Image archiving and transfer capabilities at site
- For HR-pQCT, the scanner must be an XtremeCT scanner from SCANCO Medical AG (Switzerland).

3.1.5.2 Study Technologist and Coordinator Training

During the modality specific training, an experienced BioClinica trainer will instruct the site technologists and study coordinators in the following areas:

- Imaging aspects of the clinical trial protocol including imaging endpoints, inclusion and exclusion criteria related to DXA, x-ray, QCT or HR-pQCT, and the visit schedule for DXA, x-ray, QCT and HR-pQCT examinations
- Good clinical practice (GCP) documentation requirements
- Proper technique for acquiring scans and images, analysis of protocol defined regions for DXA scans and criteria that define a good quality scan and image
- Protocol for DXA, QCT and HR-pQCT instrument quality control
- Completion of study-specific labels and forms
- Procedures for shipping or electronically transferring scans and images to BioClinica (Synarc Connect)
- Critical timelines for data shipment and completion of imaging procedures, including those at study closeout

Three types of training: on-site training, telephone training, or interactive web-based training, can be used to train technologists with the study procedures. BioClinica plans to use the types of training for each imaging modality as shown in Table 4.

Table 4: Types of Training by Modality

	DXA	X-Ray	QCT	HR-pQCT
Telephone	X	X	X	X
On-site	X	X	X	X
Interactive web-based	X	X		

Telephone Training

Remote training by telephone will be used at the study start to train DXA, and x-ray technologists. In addition, new technologists joining the study after the start up training, or who were added to replace a previous technologist may be trained in this manner.

For QCT and HR-pQCT, telephone training will be used if the technologist has worked with BioClinica in clinical trials using these modalities before and has exhibited good performance.

The BioClinica trainer will send the study supplies and a training presentation to the investigator site. The BioClinica trainer will then telephone the site's technologist(s) and provide remote telephone training by discussing the presentation and answering questions.

Interactive Web-Based Training

A web-based training program will be available to guide technologists through each facet of training, including: procedure manual overview, review of the proper

technique for acquiring images/scans, shipping procedures, instrument quality control procedures for DXA, review of forms, and review of the critical timelines for data shipping and study completion.

On-Site Training

For DXA and x-ray technologists, on-site training will be used infrequently when more extensive training is required. For QCT and HR-pQCT technologists, on-site training will be used at study start for sites that have not previously worked with BioClinica. Additionally, it will be used as needed to improve data quality from sites that experience problems with image acquisition for any of the modalities (radiography, DXA, QCT and HR-pQCT).

3.1.6 Baseline DXA Instrument Quality Control

In order to monitor DXA scanner performance, the DXA technologists at the site will submit baseline instrument quality control information for any DXA scanner not currently monitored by BioClinica on another active clinical trial.

The baseline instrument quality control requirements are:

1. Completed QA Spine Phantom Form listing 25 most recent spine phantom scans and calculated acceptable range
2. Spine phantom scan printouts of any 2 of the 25 spine phantom scans recorded on the QA Spine Phantom Form
3. Electronic QC database file containing the data from the spine phantom scans
4. Completed DXA Service Record Form with record of any service performed in the past month

BioClinica will review the completed QA Spine Phantom and DXA Service Record forms for completeness and accuracy. The spine phantom scan printouts will be reviewed for correct scan acquisition and analysis. The data contained in the electronic QC database will be loaded into the BioClinica database for later analysis.

3.1.7 Phantom Imaging

3.1.7.1 DXA Scanners

Longitudinal Instrument Quality Control of DXA Scanners

To monitor the stability of each DXA scanner during the study, measurements of a quality control (QC) phantom (for Hologic scanners an anthropomorphic hydroxyapatite spine phantom, for Lunar scanners an aluminum spine phantom) will be collected and analyzed by BioClinica. The QC phantom should be measured each day a study subject is scanned but not less than three days per week for each scanner.

Throughout the duration of the study, investigative sites will submit monthly records of the QC phantom scan data. Upon receipt at BioClinica, the records will be visually reviewed to verify that all DXA systems are operating within 1.5% of the initial calibration. Sites will also submit service record forms, which include information about DXA scanner service, malfunctions, upgrades, and changes in personnel during the previous month.

In addition to the monthly review of forms submitted by investigator sites, BioClinica will collect electronic QC database transfers from each site on a quarterly basis. The QC database will include the results of the regular measurements of the QC phantom. The QC data will be loaded into the BioClinica database and analyzed using cumulative sum (CUSUM) tabular control charts to identify breakpoints in scanner performance during the previous three months (Lu et al., J Bone Miner Res 5:626-637, 1996).

In the event that a problem is identified by either the monthly review of QC data or quarterly CUSUM analysis, BioClinica will correspond with the site to determine if the problem is resolved or if additional corrective actions are necessary. Corrective actions could include a service visit from the manufacturer.

A detailed analysis of the quality control data from each DXA scanner will be completed prior to corrected data transfers. Based on this analysis, unstable scanner performance will be quantified and, if necessary, longitudinal correction factors will be developed from the QC data. Correction factors generated in this manner will be applied to the subject data collected on the respective DXA scanner. The evaluation of scanner calibration stability will be based on the electronic QC data (database archive file) collected from each DXA scanner used for the study. The initial evaluation of stability will be based upon the percent coefficient of variation (CV%) of QC phantom measurements over the time period of the study. Scanners with a CV% outside of normal specifications will undergo additional evaluation using CUSUM analysis to determine breakpoints in scanner performance. For each scanner, a baseline calibration reference point will be established based on the spine phantom QC data collected and the date of the first subject scan on that particular scanner. Change in scanner calibration from the baseline calibration will be assessed using CUSUM analysis. The change in scanner calibration at each identified breakpoint will be evaluated for both statistical and clinical significance. Statistical significance is defined as a p-value less than 0.05 for a t-test of mean phantom BMD before and after the break point. Clinical significance is a difference in BMD across the breakpoint greater than 0.5%. Non-significant breakpoints are considered normal variation within scanner specifications and are ignored. Significant breakpoints represent unacceptable changes in calibration and a correction is warranted. After the breakpoint analysis is complete, linear regression of phantom data versus time in each remaining interval (i.e., between breakpoints or, if no breakpoints were identified, over the entire time period) will be performed to determine gradual drift in scanner calibration. Statistical and clinical significance thresholds are again applied. Statistical significance is a p-value of the linear regression less than 0.05 and clinical significance is a change in BMD greater than 0.5% from the beginning of the interval to the end of the interval.

Longitudinal correction of the subject BMD is achieved in the following way:

Ex: Assume a site has L intervals. For the i interval ($i = 1, 2, \ldots L$), we use α_i to denote the intercept, β_i for slope, and *days* as the number of days since the beginning of the affected interval. The correction formula of subject BMD collected during the study period for the scanner is:

$$BMD_c = \left(\frac{\alpha_1}{(\beta_i)(days) + \alpha_i} \right) BMD$$

Here, BMD_c is the corrected subject BMD value so it is consistent with the baseline. *BMD* is the observed subject BMD value. The formula corrects BMD values within the interval *i* according to the mean phantom BMD at the beginning of the study (i.e., the baseline scanner calibration). This correction is applied to the subject BMD results only. Bone mineral content (BMC) and BMD are not corrected. BMD results from each skeletal site and region of interest will have the same correction. BioClinica will transfer to Merck both the raw and corrected subject BMD results. The corrected data will be provided for the final data transfers at the end of the study.

DXA Scanner Software Upgrades

DXA scanner software upgrades can be expected during the time period of the trial. Most software upgrades provide bug fixes or new features while retaining the existing functionality. Rarely are scan acquisition and analysis protocols changed. However, when protocols are changed, the result is usually improvement of the BMD measurement. Therefore, software upgrades will be permitted only after a case-by-case evaluation of each new software release. BioClinica will review the release notes of each new software release and discuss the changes with the manufacturer. Software upgrades will be permitted if the evaluation indicates there is no impact of the upgrade on BMD results or if the impact can be adequately managed. Several outcomes of the evaluation can result:

1. No Change in Scan Acquisition or Analysis Protocols: In general, upgrades of this nature will be permitted without special restrictions or actions. Advance approval is still required.

2. Scan Acquisition Changed and Original Protocols Still Available: Upgrades of this nature will be permitted but sites must continue to use the original scan acquisition protocols in order to maintain consistency with the baseline measurements.

3. Scan Acquisition Changed and Original Protocols No Longer Available: In general, upgrades of this nature will NOT be permitted.

4. Scan Analysis Changed and Original Protocols Still Available: Upgrades of this nature will be permitted but sites must continue to use the original scan analysis protocols in order to maintain consistency with the baseline measurements. Since analysis of follow-up scans is performed by BioClinica, BioClinica will ensure consistency of the scan analysis protocol. If the new analysis protocols are deemed beneficial to the study, it might be possible to switch (see #5 below).

5. Scan Analysis Changed and Original Protocols No Longer Available: In general, upgrades of this nature will NOT be permitted. However, if the upgrade has significant benefits or in a practical sense cannot be avoided, it might be possible to allow this upgrade. Generally, it will be possible to re-analyze the historical scans from all affected scanners using the new analysis

protocols thus establishing consistency in the analysis method and BMD results.

DXA Scanner Hardware Upgrades

DXA scans for an individual subject should be performed on the same DXA scanner at all visits. To minimize the possibility of DXA scanner changes during the course of the study, only recent model DXA scanners, which will be supported for the duration of the clinical trial, will be permitted to be used in this study. However, a change in DXA scanner might be unavoidable.

DXA scanner upgrades will be approved in advance by BioClinica in collaboration with Merck. Approved scanner upgrades will require, if possible, a cross-calibration experiment using 20 volunteers scanned on both the old and new systems at the relevant skeletal sites. If an in vivo cross-calibration is not possible, BioClinica will request phantom cross-calibration between the two scanners. BioClinica will collect the cross-calibration scans, analyze them to determine the bone mineral density calibration of each scanner and conduct the statistical analysis required to compare the relative calibration of the two systems. If a significant difference in calibration exists between the new and old scanner, BioClinica will develop correction factors and apply these corrections to the subject data from this study site.

In the event that a scanner is replaced with another scanner of a different manufacturer, correction factors will be developed for statistical analysis only. Due to measurement differences between manufacturers, safety calculations should be performed as the sum of the total percent change from the first to last scan on the original scanner and the total percent change from the first to last scan on the replacement scanner (see section 2.2.3).

A summary of DXA phantom procedures is provided in Table 5.

Table 5: Summary of DXA Phantom Scanning Procedures

Phantom	Quality Control	Cross-Calibration
Lunar spine	*Regular 'daily' scanning**	*Not Done*
Hologic spine	*Regular 'daily' scanning**	*Not Done*
Hologic linearity		*Not Done*

* See page 21

3.1.7.2 QCT Scanners

Longitudinal Instrument Quality Control of QCT Scanners

To monitor the stability of each QCT scanner during the study, measurements of the longitudinal quality assurance (QA) phantom will be collected and analyzed by BioClinica. The QA phantom should be measured within one week (±7 days) of the day a study subject is scanned. However, if there is a gap in subject visits the Longitudinal QA phantom should be scanned at least once a month. In addition, BioClinica will provide each site with a QRM Hip/Spine QC phantom

The QA phantom results will be analyzed for changes over time and adjustments made to correct for this. The adjustments will be made by developing correction factors that are applied to the BMD results before the data transfer to the sponsor. The QA phantom data will be analyzed for stability of calibration, identification of breakpoints and quantification of calibration shifts using CUSUM analysis and other statistical tests deemed appropriate for the quantity and character of data collected.

QCT Scanner Hardware Upgrades

QCT scans for an individual subject should be performed on the same QCT scanner at all visits. To minimize the possibility of on-study QCT scanner changes, sites will be advised of the need to use the same scanner for the entire study and asked about future plans to replace the scanner. However, a change in QCT scanner might be unavoidable.

In the event of QCT scanner replacement, phantoms will be scanned on the old and new scanner. Scanning on the old scanner may no longer be possible if a technical failure caused its replacement. The phantoms to be scanned will include the European Spine Phantom (ESP) and/or the Hip/Spine QC Phantom (HSQP) both manufactured by QRM (Möhrendorf, Germany). The phantom results from the original and new scanner will be compared to determine if calibration differences exist between scanners. If so, corrections will be developed and applied to the subject scans acquired in the new QCT scanner to match them to the original scanner. Correction factors will take the form

$$BMD_c = CF * BMD + CF_{const}$$

where BMD_c is the corrected subject BMD value, BMD is the observed or raw subject BMD, CF and CF_{const} represent multiplicative and additive scanner change correction factors.

Cross-Calibration of QCT Scanners

BioClinica will circulate a cross calibration phantom to all study sites participating in the study once during the course of the study. The cross-calibration phantom used for the study will be the same as those used in case of QCT scanner hardware upgrade. The ESP includes three hydroxyapatite vertebrae (representing the lumbar vertebrae from L1 to L3) embedded in a tissue mimicking torso. The vertebrae are of homogeneous density trabecular density which varies from 50 mg/cm^3 to 150 mg/cm^3 among the three vertebrae. The ESP also has a vertebral cortical shell which can be measured separately to determine if distinct cortical BMD differences exist among scanners. In addition the QRM Hip/Spine QC Phantom may be used for cross-calibration of study scanners. The QRM phantom has hydroxyapatite components representing a vertebra and the hips with varying trabecular BMD and cortical dimensions.

The cross-calibration phantoms will either be carried to the study site during the on-site training or they will be shipped to each QCT scanner in the study along with detailed instructions for performing the cross-calibration measurements. Each phantom will be scanned once on each QCT scanner. The cross-calibration scan will be copied to a transfer medium and sent to BioClinica for processing and statistical analysis. Each scan will undergo a quality review and central analysis to determine the BMD. The phantom

results will be used to assess the relative calibration among the study QCT scanners. For each scanner field inhomogeneity correction factors will be calculated based on the BMD differences between the calibration phantom and spinal or femoral inserts in the QA phantom. Thus correction factors for the spine and hip may differ. The correction factor will take the form:

$$BMD_c = XCF * BMD + XCF_{const}$$

where BMD_c is the corrected subject BMD value, BMD is the observed or raw subject BMD, XCF and XCF_{const} represent the multiplicative and additive cross-calibration correction factors. If appropriate corrections can be derived, they will be applied to the subject data to remove differences among scanners.

Table 6: Summary of QCT Phantom Scanning Procedures

Phantom	Quality Control	Cross-Calibration
Mindways QA	*Within 1 week of subject scans; at least monthly*	*NA*
QRM Hip/Spine QC Phantom	*Within 1 week of subject scans; at least monthly*	*Once during study*
European Spine Phantom		*Once during study*

*** In case of major scanner repair or scanner replacement additional scans will be required.**

3.1.7.3 HR-pQCT Scanners

Longitudinal Instrument Quality Control of HR-pQCT Scanners

To monitor the stability of each HR-pQCT scanner during the study, measurements of the XtremeCT Forearm QC phantom (supplied by the manufacturer with each HR-pQCT scanner) will be scanned and analyzed locally on a daily basis or at a minimum each day a subject is scanned. The results of these measurements will be collected by BioClinica. In addition, BioClinica will provide each HR-pQCT facility with a QRM Forearm QC phantom (QRM GmbH, Moehrendorf, Germany). This phantom should be measured within one week (±7 days) of the day a study subject is scanned. However, if there is a gap in subject visits the QRM Forearm QC phantom should be scanned at least once a month. These scans will be sent to BioClinica for analysis and archival.

The XtremeCT Forearm QC phantom results as well as the QRM Forearm QC phantom results will be analyzed for changes over time and adjustments made to correct for this. The phantom data will be analyzed for stability of calibration, identification of breakpoints and quantification of calibration shifts using CUSUM analysis and other statistical tests deemed appropriate for the quantity and character of data collected.

HR-pQCT Scanner Hardware Upgrades

HR-pQCT scans for an individual subject should be performed on the same HR-pQCT scanner at all visits. To minimize the possibility of on-study HR-pQCT scanner changes, sites will be advised of the need to use the same scanner for the entire study and asked

about future plans to replace the scanner. However, a change in HR-pQCT scanner might be unavoidable. In addition, HR-pQCT scanner x-ray tubes might fail and need to be changed. This major service could cause relevant calibration changes and will need to be investigated.

In the event of scanner or x-ray tube replacement, phantoms will be scanned on the old and new scanner. Scans on the old scanner may no longer be possible if a technical failure caused its replacement. The phantoms to be scanned will include the XtremeCT Forearm QC phantom and the QRM Forearm QC phantom. The phantom results from the original and new scanner will be compared to determine if calibration differences exist between scanners. If so, corrections will be developed and applied to the subject scans acquired in the new HR-pQCT scanner to match them to the original scanner. Correction factors will take the form

$$BMD_c = CF * BMD + CF_{const}$$

where BMD_c is the corrected subject BMD value, BMD is the observed or raw subject BMD, CF and CF_{const} represent multiplicative and additive scanner change correction factors.

Cross-Calibration of HR-pQCT Scanners

BioClinica will circulate cross calibration phantoms to all study sites participating in the study.. The cross-calibration phantom used for the study will be the QRM Forearm QC phantom. The QRM Forearm QC phantom is comprised of a European Forearm Phantom (EFP) coupled to a QRM calibration phantom. The EFP has different segments with varying trabecular BMD and cortical dimensions.

The cross-calibration phantom will either be carried to the study site during the on-site training or they will be shipped to each site with an HR-pQCT scanner along with detailed instructions for performing the cross-calibration measurements. Each phantom will be scanned once on each HR-pQCT scanner. Cross-calibration scans will be copied to transfer tape and sent to BioClinica for processing and statistical analysis. Each scan will undergo a quality review and central analysis to determine the BMD and geometry parameters. The phantom results will be used to assess the relative calibration among the study HR-pQCT scanners. Slope and intercept between the nominal values of the QFP and the measured values are calculated. The scanner for which the slope is closest to one and the intercept is closest to zero is selected as gold standard reference scanner and the remaining HR-pQCT scanners compared to it. If the slope of a scanner differs by more than 5% from the reference scanner and the intercept by more than $10 mg/cm^3$, a cross-calibration correction factor will be developed and applied to the subject BMD results from that scanner; these corrected data will be provided at select times during the study. The BMD correction factors will take the form:

$$BMD_c = XCF * BMD + XCF_{const}$$

where BMD_c is the corrected subject BMD value, BMD is the observed or raw subject BMD, XCF and XCF_{const} represent the multiplicative and additive cross-calibration correction factors. If appropriate corrections can be derived, they will be applied to the subject data to remove differences among scanners.

Table 7: Summary of HR-pQCT Phantom Scanning Procedures

Phantom	Quality Control	Cross-Calibration
XtremeCT Forearm QC	*Daily**	*N/A*
QRM Forearm QC	*Within 1 week of subject scans; at least monthly*	*N/A*

*** Daily before subject scans; also scanned weekly for mechanical stability.**

3.2 Collection of Unscheduled On-Protocol Imaging for Central Review

Unscheduled imaging examinations will be collected and read in the same fashion as scheduled examinations, provided they are collected on the appropriate equipment or scanner and follow the imaging protocol.

3.3 Collection of Off-Protocol Imaging for Central Review

Off-protocol images will neither be collected nor centrally reviewed. Any off-protocol imaging data that are received will be returned to the study sites.

4 RECEIPT, TRACKING AND QUALITY CONTROL OF IMAGE DATA

The flowcharts below (Figures 1-3) summarize the overall data handling process from imaging site preparation through final image preparation prior to the independent review, and transfer to Merck.

Figure 1: X-ray & QCT Data Handling Process

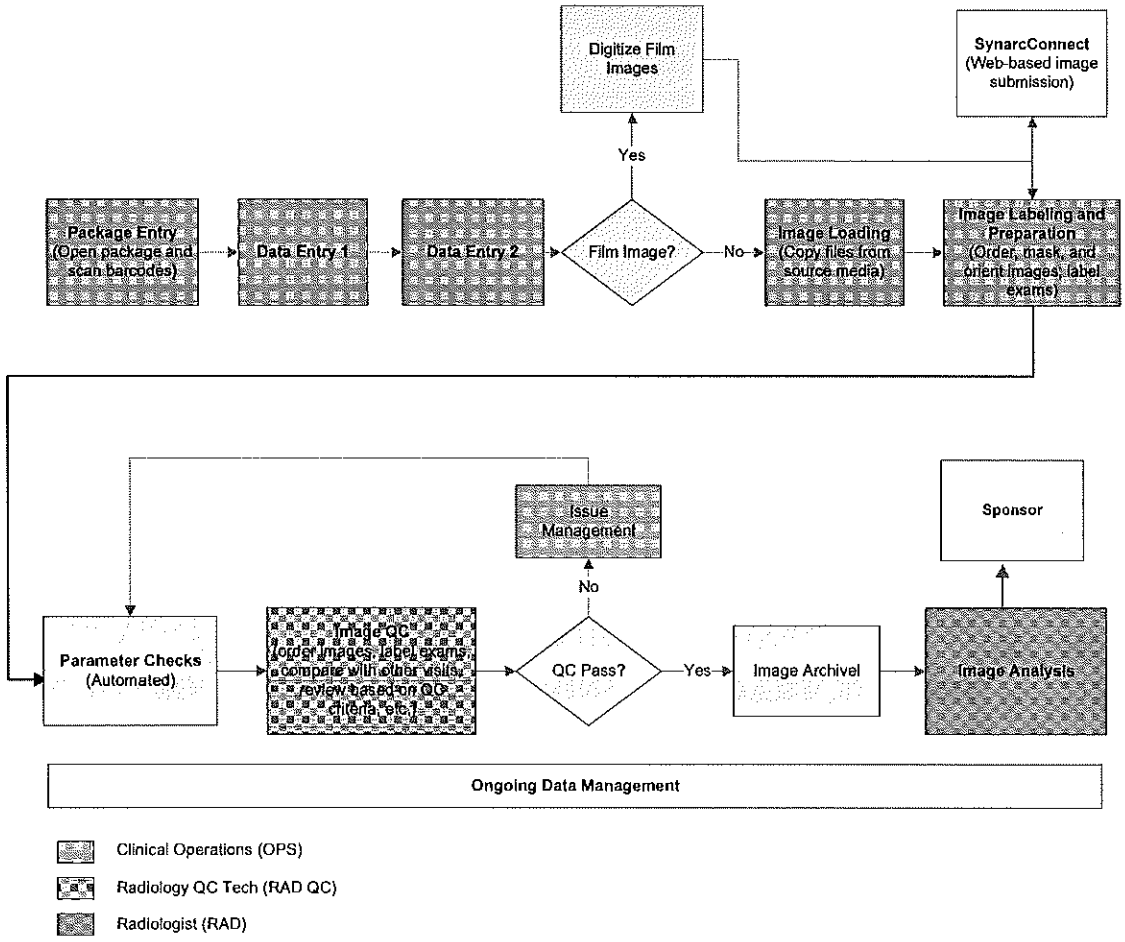


Figure 2: HR-pQCT Data Handling Process

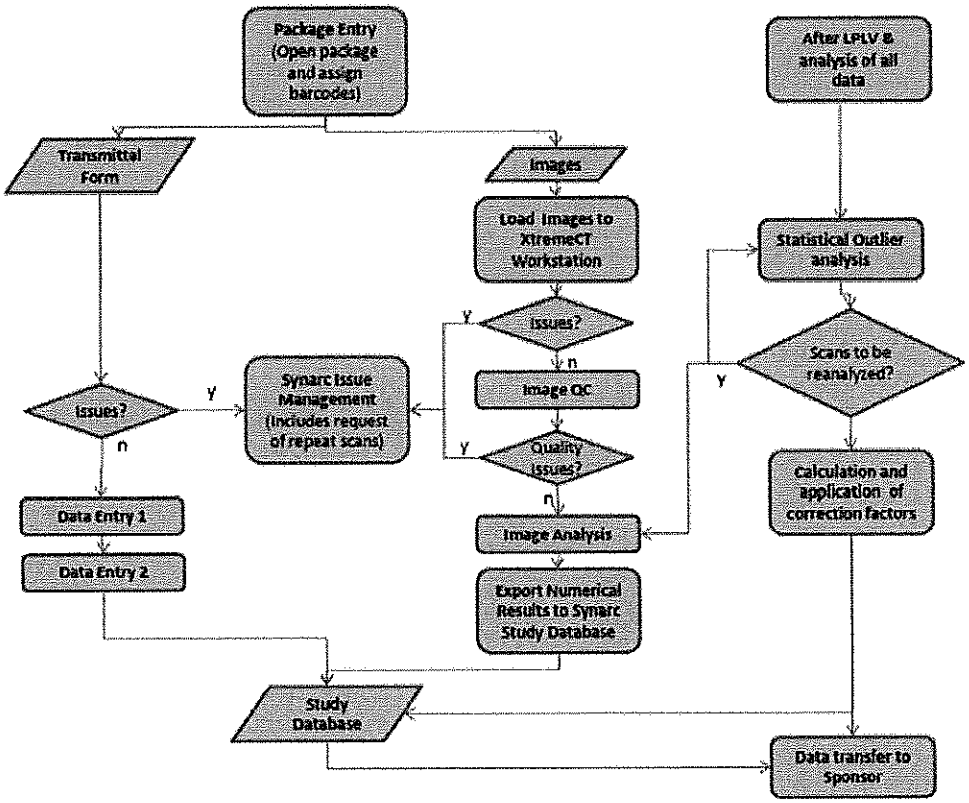
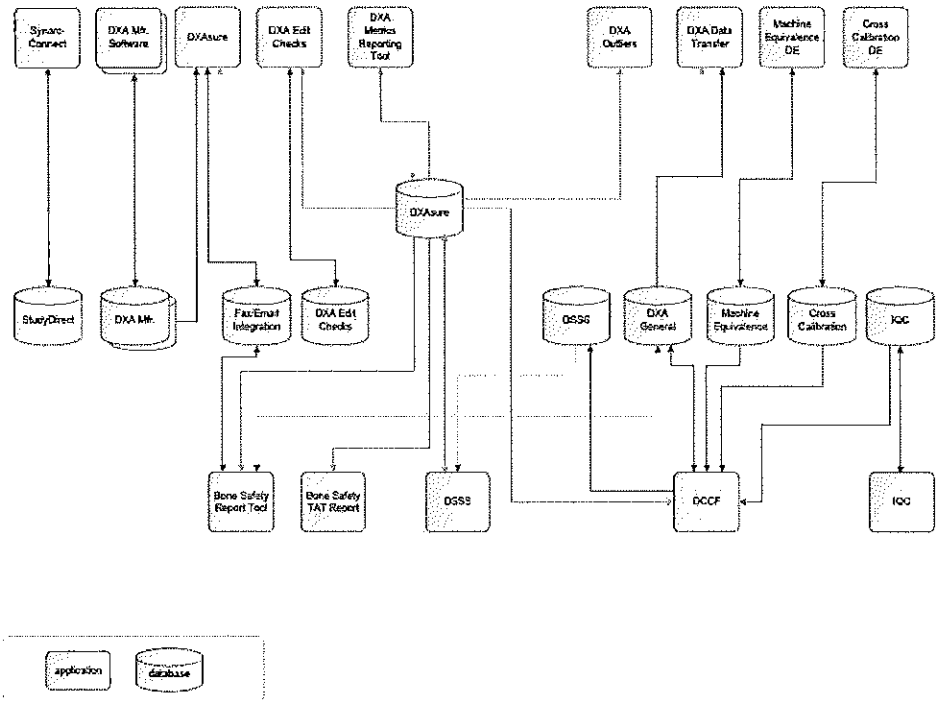


Figure 3: DXA Data Handling Process



4.1 Receipt, Storage, and Back-up of Data in Image Archive

Processes detailing image collection, management, tracking and image digitization/loading, and archival at BioClinica are documented in BioClinica's SOPs.

4.1.1 Package Receipt, Processing, Data Entry

The investigator sites will prepare complete sets of imaging data as they are acquired. Each subject's data for a given modality will be prepared as a single package (one complete visit). BioClinica will provide instructions for labeling original films and electronic media with unique, anonymous subject identifiers and instructions for completion of the transmittal form which accompanies each package submitted to BioClinica.

BioClinica will receive images (digital or hard copy) from the investigator sites in a variety of formats. Electronic images will be transmitted to BioClinica either on physical media (CD-ROM, DVD, Tape) or via BioClinica's secure web portal.

BioClinica's operations team will review each individual subject data package (images and transmittal form) received. The operations team will check the completeness of contents, the accuracy of information on the transmittal form, and the appropriateness of the timing of the visit (scheduled visits versus unscheduled visits) as well as any additional correspondence such as comments received from the clinical site.

Problems with packages will result in queries sent to the sites. All queries, in addition to issues with the data, will be recorded into BioClinica's database. Data will be entered into the database and verified by double-data entry. The database will maintain all necessary audit trails. Information captured in the database will include:

- Subject demographic information (site and subject ID numbers, date of birth)
- Visit information (visit name, visit date)
- Exam information (date of scan and subject anatomy examined)
- Date package received and identity of data entry personnel
- Results of package review including annotations, status of completeness and correctness, and query issuance and closure
- Results of quality image review including annotations

4.1.2 Digitization of Images

Hard copy x-ray films received at BioClinica will be digitized at a resolution of 200 micron pixel size using a high-resolution digitizer and converted to a digital image in the DICOM format. The use of digitized images permits computer assisted evaluation of images and provides a means for electronic archival of the study hard copy films.

An operations assistant will evaluate the quality of each digitized image. Digitized images of poor quality are discarded and the films are reprocessed. The sites will be queried if any film cannot be converted into an acceptable digitized image. The assistant will mask any information (i.e., labels or subject, institution identifiers and exam date) from the digital image to allow blinding of subject identity, institution name, and exam date.

4.1.3 Image Archival

After the process of conversion or digitization and quality review, each digital image will be copied to the BioClinica image archive. The image archive will be backed up nightly. Each image will be archived as a digital image onto a server. When needed, images will be retrieved from the archive and transferred over the network to reading workstations for analysis.

4.2 Data Tracking Procedure

Semi-automated tracking of all x-ray, QCT and HR-pQCT images and transmittal forms will be accomplished using BioClinica's proprietary tracking database and barcode labeling. Data received at BioClinica will be logged into the tracking database and a unique barcode is associated with each image/scan and each transmittal form. As the data is processed, the status will be updated in the system database such that the

location and status of the data at any time can be determined. For DXA, the receipt of each DXA will be tracked using DXAsure™ BioClinica's DXA scan tracking program, thereby recording the date of receipt.

Data discrepancies identified either during the package review or image quality review will be reported to the sites for correction and resolution.

BioClinica will issue standard reports both that confirm receipt of the image data and on the results of the quality assessment.

4.3 Quality Assessment of Images

Processes detailing image quality review at BioClinica are documented in BioClinica's SOPs.

All images will be checked by specially trained technologists familiar with the technical aspects of image quality and with factors that can affect the specific image evaluations that will be performed. Images will be checked visually for:

- Appropriate anatomical coverage
- Correct subject positioning and subject side (left/right)
- Follow-up scans consistent with baseline for subject positioning and subject side
- Scan mode selection correct and follow-up scans consistent with baseline
- Scanner consistent with baseline (same scanner used throughout study)
- Presence of artifacts (metal, clothing, subject motion, etc.)
- Proper use of film/screen combination and exposure parameters for hardcopy films
- Proper spatial and contrast resolution for digital images

Image quality reviews will be performed during the scan analysis step. The results of these evaluations will be recorded in the study database. Sites that encounter problems with image quality or protocol adherence will be contacted directly by BioClinica to discuss the problem and potential solutions.

If the scan quality is unacceptable and the problem can be corrected, a repeat examination to replace the rejected scan will be requested.

If the quality of an image is insufficient to allow proper interpretation, a query will be issued to the site requesting either a repeat of the imaging exam, a reprint of hard copy film(if available), or a retransmission of digital media.

4.3.1 Summary of Repeat Examination Guidelines

Table 7 summarizes the protocol BioClinica uses to request repeat examinations and is listed according to the different scanning modalities and skeletal sites. Note that due to frequent movement artifacts, poor quality HR-pQCT scans are typically repeated at the

site before transmission to BioClinica. However, BioClinica will request that sites repeat poor quality scans if the number of repeats has not been exceeded.

Table 8: Guidelines for Repeat Examinations

Examination	BioClinica Repeat Guideline (per visit)
DXA (any site)	Repeat as necessary, up to 2 requests
X-ray (any site)	Repeat as necessary, up to 2 requests
QCT (hip)	Repeat not allowed
QCT (spine)	Repeat once
HR-pQCT(radius)	Repeat up to 2 times, by site or BioClinica initiative
HR-pQCT(tibia)	Repeat up to 2 times, by site or BioClinica initiative

4.4 De-identification and Labeling of Image Data

4.4.1 Blinding of Data

Sites will be instructed and trained to remove all personal-identifying information from images prior to sending the data to BioClinica. Following the initial review of data received from the sites, BioClinica will convert all imaging data to DICOM using customized software designed especially for use in clinical trials. Original image header information will be retained in the DICOM header, with the exception of personal identifying information. If such information was inadvertently left in by the site, it will be automatically and permanently removed upon loading the data into the tracking database.

During image analysis, readers will be blinded to treatment assignments and clinical information that may bias the readers' assessments. The BioClinica database will hold subject identifiers (e.g., Site ID and Subject ID number) and these keys will be used to link reading results to the same subject in Merck's clinical database.

All personnel at BioClinica, including the operations staff, will also be blinded to treatment assignments and other clinical data from study subjects. There will be no communication between BioClinica and Merck regarding treatment codes. Unblinding information will not be shared with BioClinica even after the BioClinica database is locked.

4.4.2 Labeling of Data

The investigator site will label all imaging data submitted to BioClinica in a manner that unambiguously identifies the nature of the data. This will be done by using a study and site identifier as well as anonymous identifiers for the subject. Imaging data will be labeled with the appropriate visit code as well as the modality and the examination date for each image/scan. Finally, the anatomical location and projection (e.g., lateral lumbar spine radiograph) will be either be assigned by the investigator site and confirmed by BioClinica or assigned by BioClinica technicians or radiologists.

4.5 Procedures for Tracking, Documenting and Resolving Missing Image Data

BioClinica will collaborate with Merck to identify and collect all missing data. Each protocol required image will be tracked until it is collected or the site has confirmed that the scan was not done or is no longer available.

If there is a discrepancy with the actual package contents and the transmittal form, sites will be sent a query in an attempt to resolve the following issues:

- Missing baseline

  BioClinica will use enrollment reports provided by Merck and data completeness checks to determine if baseline scans are missing.

- Missing follow-up time points

  BioClinica will project expected follow-up visit dates and report to Merck and investigational sites any scans that are not received in a timely fashion after the expected date.

- Missing anatomical views

  If required anatomy is missing from a scan or image, BioClinica will ask the site to repeat the examination. If a scan cannot be repeated, it will not be analyzed unless the available data can be used to generate a subset of results (e.g., if one vertebral body is cutoff, the remaining vertebrae can be analyzed and reported).

- Missing tomographic reconstructions

  Several tomographic reconstructions of the same CT scan raw data representing large and small field of view will be submitted by the site. Each reconstruction is used in the QCT analysis. If one of these is missing, BioClinica will ask the site to perform the reconstruction from the existing raw data stored at the site. Should these data no longer be available then a repeat scan may be requested. Depending on the nature of the missing reconstructions some or all of the QCT parameters will be missing.

Queries generated by BioClinica will be forwarded directly to the investigator site. Query resolutions will be sent directly to BioClinica from the site.

4.6 Deviations from Imaging Protocol

Deviations from the imaging protocol involving image acquisition or timing will be evaluated on a case-by-case basis and a decision made as to whether or not the scans are useable. In order to avoid missing data, scans will be used whenever possible. However, scans will not be used if the deviation from the imaging protocol would alter results.

During the course of the study, a list of deviations will be maintained and it will contain information about which data have been excluded from analysis or the transfer along with the reasons such as:

- Image data not acquired per protocol
- Missing image data
- Poor quality image data
- Excluded or incomplete reading due to data receipt after deadline for inclusion in data transfer

5 DESIGN OF INDEPENDENT REVIEW

5.1 Purpose

Independent review will play a role in evaluating subject eligibility, drug efficacy, and subject safety. Details are provided in the clinical study protocol and summarized below. Refer to the current version of the study protocol for additional details.

5.1.1 Subject Eligibility

BioClinica will communicate to investigators via fax or email the results of the central analysis of the screening DXA spine and femur scans and the screening lateral thoracic and lateral lumbar x-ray images. The results will be reported within 5 business days of the receipt of a complete and correct DXA and x-ray package for each subject. The investigator will use the results of BioClinica central analysis to determine subject eligibility. Eligibility was only performed for the base study and all results were loaded in the IVRS system used for subject randomization.

5.1.2 Efficacy

Central reading and analysis of imaging data will be used to evaluate drug efficacy, including the primary endpoint of the trial and numerous secondary and exploratory endpoints. Specifically, the primary endpoint of the trial, the incidence of new vertebral fractures, will be assessed from the lateral spine radiographs. Secondary endpoints evaluating the effects on BMD will use DXA scans of the spine and hip in all subjects and the forearm and total body in a subset of subjects. Exploratory endpoints to investigate the effects on volumetric BMD, bone geometry and microarchitecture as well as estimated failure load will be assessed using QCT and HR-pQCT in a subset of subjects.

5.1.3 Safety

During enrollment, DXA will be used to ensure subject safety by identifying subjects with BMD that is too low and below protocol defined thresholds. These individuals are not eligible for the trial due to their excess risk for future fractures. In addition, DXA will be used to monitor subject safety during follow-up by identifying subjects with excessive bone loss at proximal femur (total hip region of interest) and lumbar spine.

Details of the safety criteria can be found in the clinical trial protocol. Any confirmed safety findings related to excessive bone loss will be reported by BioClinica to the investigator site.

5.2 Description of Review Paradigm

5.2.1 Prevalent Vertebral Fracture Assessment for Eligibility

Screening spine radiographs will be read for prevalent vertebral fracture using the Genant Semiquantitative (SQ) scoring method as they are received. The reading is done by a single reader from a pool of qualified readers. If necessary, the reader will consult with other readers on questionable cases. This central assessment of vertebral fracture will be reported back to the investigator and used to determine subject eligibility.

5.2.2 Incident Vertebral Fracture Assessment

Follow-up visit spine radiographs will be read centrally as they are received starting with the Month 24 follow-up visit or early termination visit (note: Month 6 and 12 visits will be read according to the reading protocol described below) by a reader from a pool of SQ readers. If the SQ reader identifies an incident vertebral fracture, a reader from a pool of QM readers will assess the case independently, blinded to the results of the SQ reading. If the SQ and QM readers agree that an incident vertebral fracture occurred at a particular vertebra and visit, the fracture will be confirmed. If the QM assessment identifies an incident vertebral fracture at a particular vertebra and visit where SQ does not, then a binary SQ (BSQ) reading will be performed by a third independent reader to adjudicate the discrepancy.

When necessary SQ, QM and BSQ readers will review screening and follow-up visits side-by-side without blinding to time order. In addition, any previously recorded scores or markings for measurement of vertebral body height will be accessible and editable by the respective reader. As noted above, the SQ, QM and BSQ reader will be blinded to the results of the other reading types (e.g., SQ reader will not have access to QM or BSQ results, etc.).

5.2.3 Modification of Prevalent Vertebral Fracture Status

When reading spine radiographs from post-screening visits, the reader might have additional information (i.e., a clearer view of the vertebral body) that will lead to a change in the original prevalent fracture assessment. The SQ readers may get input from other SQ readers when making these changes to the prevalent fracture assessment. Both original and new scores are kept in the study database.

5.2.4 Clinical Fracture and Delayed Fracture Union Adjudication

BioClinica radiologists will perform central adjudication of clinical fracture and delayed fracture union events that occur during the study. The details for the collection, handling and reading of these events are described in a separate charter created in collaboration with Merck and Parexel: Protocol 018 Fracture Events Clinical Adjudication Committee. Note that clinical vertebral fracture events will be confirmed by the process outlined in that charter and not the morphometric vertebral fracture assessment methods described in this document.

5.2.5 DXA and HR-pQCT

DXA and HR-pQCT scans will be analyzed in an on-going basis as they arrive from the investigator sites. When analyzing a particular scan, the reader will have access to all other scans and scan results for that subject.

5.2.6 QCT

QCT scans will be analyzed in batches when the last follow-up visits are arriving. All visits for a subject will be analyzed together to maximize consistency of the analysis. When analyzing a particular scan, the reader will have access to all other scans and scan results for that subject.

5.3 Assessment Criteria

Assessments will be performed according to clinical standards, the recommendations of the software manufacturers, and BioClinica's SOPs and analysis guidelines. No exceptions to standard methods are expected.

5.3.1 Vertebral Fracture Assessment

Vertebral fracture assessment will be performed using three different methods: semiquantitative assessment (SQ), quantitative morphometry (QM) and binary semiquantitative assessment (BSQ). SQ reading is the primary method of vertebral fracture assessment and will be performed on all subjects at all visits. QM will be used to confirm incident vertebral fractures found by the SQ method and BSQ will be used to adjudicate discrepancies in the SQ and QM results. Whenever one of these reading methods is applied to a visit, all 13 vertebrae from T4 to L4 will be evaluated.

This section describes the individual reading methods. Details of the fracture reading process and method of combining the individual assessments into a final fracture outcome are given in Section 6.

Semiquantitative Assessment (SQ)

The lateral spine radiographs will be assessed for prevalent and incident vertebral fractures in the range T4 to L4 using the Genant Semiquantitative Scoring method.

- Grade 0: normal (approximately < 20% reduction in anterior, middle, and/or posterior height)
- Grade 1: mild fracture (approximately 20% - 25% reduction in anterior, middle, and/or posterior height)
- Grade 2: moderate fracture (approximately 25% - 40% reduction in anterior, middle, and/or posterior height)
- Grade 3: severe fracture (approximately > 40% reduction in anterior, middle, and/or posterior height)

Incident vertebral fractures occur in vertebrae with an increase in SQ score ≥ 1 at a follow-up visit. New incident vertebral fractures occur in vertebrae that are not fractured (i.e., SQ score of 0) at baseline. Worsening incident vertebral fractures occur in vertebrae with prevalent (i.e., pre-existing) fractures at baseline.

Binary Semiquantitative Assessment (BSQ)

Binary SQ reading follows the same visual assessment protocol as SQ reading with the exception that fracture outcomes are simply 'No' (i.e., no fracture present) or 'Yes' (i.e., an vertebral fracture is present) and severity is not recorded. Thus, at the Screening visit the BSQ reader will assign a vertebra without a prevalent fracture (i.e.,

an equivalent SQ score of 0) the result 'No' while a vertebra with a prevalent fracture (i.e., equivalent SQ score of 1, 2 or 3) will be assigned the result 'Yes'. On follow-up visits the BSQ reader evaluates whether the vertebral body lost height and the equivalent SQ grade changed by 1 or more from the prior visit and scores those vertebrae without a change as 'No' (i.e., no incident fracture present) and those with a change as 'Yes' (i.e., an incident vertebral fracture is present).

Quantitative Morphometry (QM)

In quantitative morphometry the shape of the vertebral bodies from T4 to L4 will be analyzed on the digital images using a 6 point morphometric technique. Experienced readers will place three points on the inferior and superior endplates of each vertebral body. The points will define the anterior, middle and posterior aspects of the endplate. The location of each point defining the vertebral contour is saved and can be retrieved for inspection at a later time. Anterior, middle, and posterior vertebral body heights will be determined from the relative point positions and the heights stored in an electronic database. Incident vertebral fractures will be defined as a 20% reduction from Screening in any vertebral body height that also shows an absolute change that is at least 4 mm.

5.4 Types of Assessments or Measurements

5.4.1 Description of Measurements or Assessments

Evaluation of morphometric vertebral fractures will be conducted by subjective evaluation by trained radiologists (SQ and BSQ) and trained radiologists and/or technologists (QM) as described above. Results will be recorded and stored in the study database.

Measurements of BMD will be performed using algorithms coded into computer software supervised by trained human operators. The software is provided by various third parties. Specifically, DXA scanner software produced by:

GE Lunar (Madison, WI)

Hologic, Inc. (Bedford, MA)

Norland, (Fort Atkinson, WI)

For QCT, scans will be analyzed using MIAF (Medical Image Analysis Framework, University of Erlangen, Germany).

For HR-pQCT, XtremeCT software produced by SCANCO Medical AG (Switzerland) will be used.

Finite element analysis will be performed by third parties. Finite element analysis of spine and hip QCT will be conducted by O.N. Diagnostics, LLC (Berkeley, CA). Finite element analysis of HR-pQCT scans will be conducted by Merck & Co.

5.4.2 Limitations of Image Acquisition Parameters on Measurements

DXA and HR-pQCT scanners use standardized acquisition protocols, but each has several "scan modes" to select from. Inconsistency in the scan acquisition parameters can cause measurement results to be invalid; they cannot be compared to

measurements made with a different scan mode. Variation in scan mode will lead to a repeat request. QCT scans will be acquired according to a strict acquisition protocol, customized to a site's specific CT scanner, as necessary. Variation in parameters (e.g., slice thickness, kVp, reconstruction kernel, etc.) can invalidate results and lead to a repeat request, as appropriate.

Limitations in x-ray are due to variable positioning of the spine which affects the projection of the vertebral body in the radiograph as well as tube current and voltage settings. All of these can cause over/underexposed images or low contrast images.

Variation in acquisition parameters will be evaluated during the scan quality review and scan analysis steps, and inadequate scans will be repeated or invalid measurement results will be excluded.

5.5 Number and Classification of Reviewers

For morphometric vertebral fracture assessment, at least 6 readers will be identified so that each of the three roles (primary SQ reader, QM reader, and BSQ adjudicator) will have at least two readers in the pool to manage the volume of work.

At least 6 reviewers will be selected for the analysis of the DXA scans collected in this trial. Similarly, at least 2 reviewers will be selected for the analysis of the QCT scans collected in this trial.

Based on the expected volume, 2 reviewers will be selected for the analysis of the HR-pQCT scans collected in this trial.

5.6 Reviewer Roles &Responsibilities

The reviewers will be selected from a pool of BioClinica's in-house radiologists and technicians, who will be further trained and qualified on the specifics of this study as described in Section 7.0.

Readers will be responsible for evaluating radiographs and other medical images and analyzing DXA, QCT and HR-pQCT scans for all subjects in the trial. Readers will also review scans flagged by edit checks and outlier analyses. After this investigation, the readers will make exclusions or corrections as necessary.

5.7 Delineation of Independent Review Population

All on protocol images of screened or randomized subjects determined to be of good quality will be included in the independent review.

5.8 Reviewer Blinding

During image analysis, readers will be blinded to treatment assignment. No clinical information is required for central reading of DXA scans, QCT scans, HR-pQCT scans or vertebral fracture. However, clinical information may be provided from time to time. This includes medical history information that might explain artifacts or abnormalities in the DXA, QCT and HR-pQCT scans. Access to clinical information will not bias the readers' assessments.

If diagnostic markings or measurements made by site investigators are noted on the films at the time of the quality check and prior to digitization, BioClinica will ask the investigator site to produce images without markings. If there is no possibility of obtaining a non-marked image, an effort will be made to remove them without damage to the underlying film.

5.9 Description of Adjudication Paradigm

Adjudication of incident vertebral fracture discrepancies between the SQ and QM readers will be adjudicated by a third reader (BSQ reader) at BioClinica who is blinded to the results of the two other readers (SQ and QM). The BSQ adjudication will be performed when the QM assessment identifies an incident vertebral fracture at a particular vertebra and visit where SQ does not. This adjudication is separate and distinct from clinical fracture adjudication performed in collaboration with Parexel.

For QCT, HR-pQCT and DXA scans, there will be a single reading of each subject's images/scans and therefore no adjudication is planned for these examinations.

5.10 Interim Analysis

The first formal interim analysis for both efficacy and safety will be performed at the time at which approximately 70% of all pre-specified hip fracture events (first adjudicated osteoporotic hip fracture per patient) have occurred and is expected to be after approximately 3 years of total study duration (including 1 year recruitment). The second formal efficacy interim will be performed when approximately 85% of the pre-specified hip events have been observed, and the Final Analysis will occur when all pre-specified fracture events (hip, morphometric vertebral and non-vertebral) have been observed.

5.11 Role of Independent Review in the Interim Analysis

The results of central reading of incident vertebral fracture as detailed in this charter will be used at the two interim analyses to decide on early termination or continuation of the trial (see the clinical trial protocol and Statistical Analysis Plan for details).

Early Termination for Efficacy: If all three p-values (for vertebral, hip, non-vertebral fractures) are lower than the corresponding boundary based on the alpha-spending function at either the first or second interim look, the Data Monitoring Committee may recommend termination of the trial before all pre-specified events are seen.

Early Termination for Futility: The trial may be terminated in the first (70% of all pre-specified hip fractures) or second (85% of all pre-specified hip fractures) formal interim analysis if the p-value for vertebral fractures is higher than the corresponding cut-off for futility.

The DXA, QCT and HR-pQCT data will be collected and analyzed at each interim analysis but will not play a direct role in deciding the early termination or continuation of the trial.

6 METHODOLOGY FOR THE INDEPENDENT REVIEW

6.1 Description of Image Analysis Systems

BioClinica will customize its proprietary electronic central reading systems, V-FRACT and SynarcConnect™ for the radiological assessment of the imaging data collected in this study. Analysis of DXA scans will be performed with the DXA scanner software from each manufacturer and DXA scan tracking and management of DXA results will be performed with BioClinica's proprietary DXAsure™ software. The review and analysis of spine and hip QCT scans will be performed using MIAF (Medical Image Analysis Framework, University of Erlangen, Germany). The review and analysis of radius and tibia HR-pQCT scans will be performed using XtremeCT software (SCANCO Medical, Switzerland).

6.1.1 Radiographic Image Analysis

The BIOCLINICA reading system will allow readers to see the spine radiographs from selected visits for a subject side-by-side on high resolution monitors (The selection of visits is described in Section 6.2). Readers can customize the image display including the hanging protocol as well as image brightness/contrast and magnification. Images are displayed without temporal blinding. The radiologist will assess the Screening images for prevalent vertebral fractures and the follow-up visit images for incident vertebral fractures. These results will be recorded by the radiologist through the electronic reading system and saved with an electronic signature into the study database. When reading, results from all previously read visits will be displayed and readers can record new results or edit existing results via the user interface.

BioClinica radiologists or specially trained technicians will mark the location of each vertebral level on the lateral spine radiographs to facilitate scoring by the radiologist and comparison of vertebrae between visits. The vertebral level will be determined by identifying L5 according to anatomical landmarks and counting up to T4 in the thoracic spine (L5 to L1, T12 to T4).

6.1.2 DXA Image Analysis

DXA scans received by BioClinica will be centrally analyzed following standard procedures documented in BioClinica's DXA analysis manual. These analyses are based upon the manufacturer's recommendations and clinical best practices. Trained clinical staff will utilize the DXA manufacturer software to analyze each scan by loading it onto the DXA analysis workstation where it can be assessed using the appropriate DXA software. The clinical staff member will analyze the scan by reviewing and modifying the bone edges and positioning markers to define the necessary regions of interest. BioClinica will allow the software to automatically detect bone edges; however, if needed, the BioClinica reviewer can intervene to manually identify the bone edges. In these cases, the clinical staff member will follow procedures outlined in the procedural manual, and all changes will be documented in the system. Once analyzed, the scan will be saved and archived with the BioClinica analysis included. In addition, the analyzed scan will be printed and filed in study binders.

BioClinica may exclude scans for various reasons including: poor subject positioning, improper scan parameters, subject movement, and internal or external artifacts. Whenever possible a repeat scan will be requested to replace the inadequate scan. If a good quality scan cannot be obtained the data will be flagged for exclusion from the final analysis, but will remain in the database for audit purposes.

For all anatomical regions BMD will be calculated by the standard convention: BMC divided by area. Specific conventions are as follows:

6.1.2.1 Lumbar Spine Conventions

Analysis will include four vertebral levels from L1 to L4. Vertebral labels will be assigned, using anatomical landmarks, from the bottom up starting with L5.

Individual vertebral levels may be excluded due to artifact on incident vertebral fracture. All confirmed incident vertebral fractures in the vertebrae from L1 to L4 will be reconciled with the DXA data to exclude the fractured levels from the total spine BMD. A vertebral level excluded from one visit will be excluded from all visits with the total spine BMD calculated based on the evaluable vertebral levels.

6.1.2.2 Femur Conventions

The left side should be scanned at all study visits. If the left side is not evaluable (e.g., hip osteoarthritis, deformity) or if the right side is inadvertently scanned at baseline, then the right side must be used throughout the study.

If a subject fractures the hip being examined during the study, no further scans of either hip will be acquired. If the fractured hip or contralateral hip is submitted to BioClinica, the scan will be entered into DXAsure and marked as unusable.

The femur scans will be analyzed to measure BMD of the total femur, femoral neck, trochanter, intertrochanter/shaft, and Ward's regions of interest.

6.1.2.3 Forearm Conventions

The non-dominant side should be scanned. If the left (or non-dominant) side is not evaluable (e.g., metal implant, previous fracture, deformity) or if the right (or dominant) side is inadvertently scanned at baseline, then the right (or dominant) side must be used throughout the study.

If a subject fractures the arm being examined during the study, no further scans of either arm will be acquired. If the fractured arm or contralateral arm is submitted to BioClinica, the scan will be entered into DXAsure and marked as unusable.

The forearm scans will be analyzed to measure BMD of the ultradistal radius, mid region, and distal 1/3 radius as well as the total radius.

BioClinica may exclude scans from analysis and/or from the transfer file for various reasons ranging from poor subject positioning, improper scan parameters, subject movement and internal or external artifacts. These data will be flagged for exclusion from the final analysis but will remain in the BioClinica database for audit purposes.

6.1.2.4 Total Body Conventions

The presence of metal in the body will be noted and its location recorded

Total body BMD is calculated both including and excluding the head. The total body scans will be analyzed to measure BMD of the arms, legs and trunk.

6.1.3 QCT Scan Analysis

The review and analysis of spine and hip QCT scans will be performed using MIAF (Medical Image Analysis Framework, University of Erlangen, Germany).

QCT scans collected by BioClinica will be centrally analyzed using standard procedures that will be outlined in BioClinica's QCT analysis guidelines. In general, QCT scans will be analyzed in batches as the last visit for each subject is received.

If a subject fractures the vertebrae (or hip) being examined during the study, no further scans of the spine (or hip) will be acquired. If a scan of a fractured location is submitted to BioClinica, the scan will be entered into the BioClinica database and marked as unusable.

BioClinica may exclude scans from analysis and/or from the transfer file for various reasons ranging from poor subject positioning, improper scan parameters, subject movement and internal or external artifacts. In some cases the analysis software may also fail. These data will be flagged for exclusion from the final analysis but will remain in the BioClinica database for audit purposes.

Spine and Hip

Specially trained technicians will utilize the QCT software to analyze each scan. To do so, the scans will be loaded onto the QCT analysis workstation. The technician will analyze the scan by using the software to segment the periosteal and endosteal surface. The software is designed to perform some of these processes in an automated fashion with adjustment and correction by the operator when needed. Once analyzed, the results of the analysis will be loaded into the BioClinica database.

Finite Element Analysis

Finite element analysis of spine and hip QCT will be conducted by O.N. Diagnostics, LLC (Berkeley, CA). BioClinica will provide O.N. Diagnostics with the necessary images. The results of the FEA will be returned to BioClinica and added to the BioClinica study database. These data will be transferred to Merck along with the data generated by BioClinica.

6.1.4 HR-pQCT Scan Analysis

The review and analysis of radius and tibia HR-pQCT scans will be performed using the XtremeCT software (SCANCO Medical, Switzerland).

HR-pQCT scans collected by BioClinica will be centrally analyzed using standard procedures that will be outlined in BioClinica's HR-pQCT analysis guidelines. These will be based upon the manufacturer's recommendations as detailed in the XtremeCT User's Manual and clinical best practices. In general, HR-pQCT scans will be analyzed as they are received.

If a subject fractures the arm (or leg) being examined during the study, no further scans of either arm (or leg) will be acquired. If the fractured limb or contralateral limb is submitted to BioClinica, the scan will be entered into the BioClinica database and marked as unusable.

BioClinica may exclude scans from analysis and/or from the transfer file for various reasons ranging from poor subject positioning subject movement and internal or external artifacts. These data will be flagged for exclusion from the final analysis but will remain in the BioClinica database for audit purposes.

Radius and Tibia

Specially trained BioClinica technicians will utilize the XtremeCT software to analyze each scan. To do so, the scans will be loaded onto the XtremeCT analysis workstation. The technician will analyze the scan by using the software to perform the following steps: segmentation of the periosteal surface. The software is designed to perform some of these steps in an automated fashion with adjustment and correction by the operator when needed. Once analyzed, the results of the analysis will be loaded into the BioClinica database.

The XtremeCT software provides evaluation of BMD, geometry and microarchitecture of the radius and tibia.

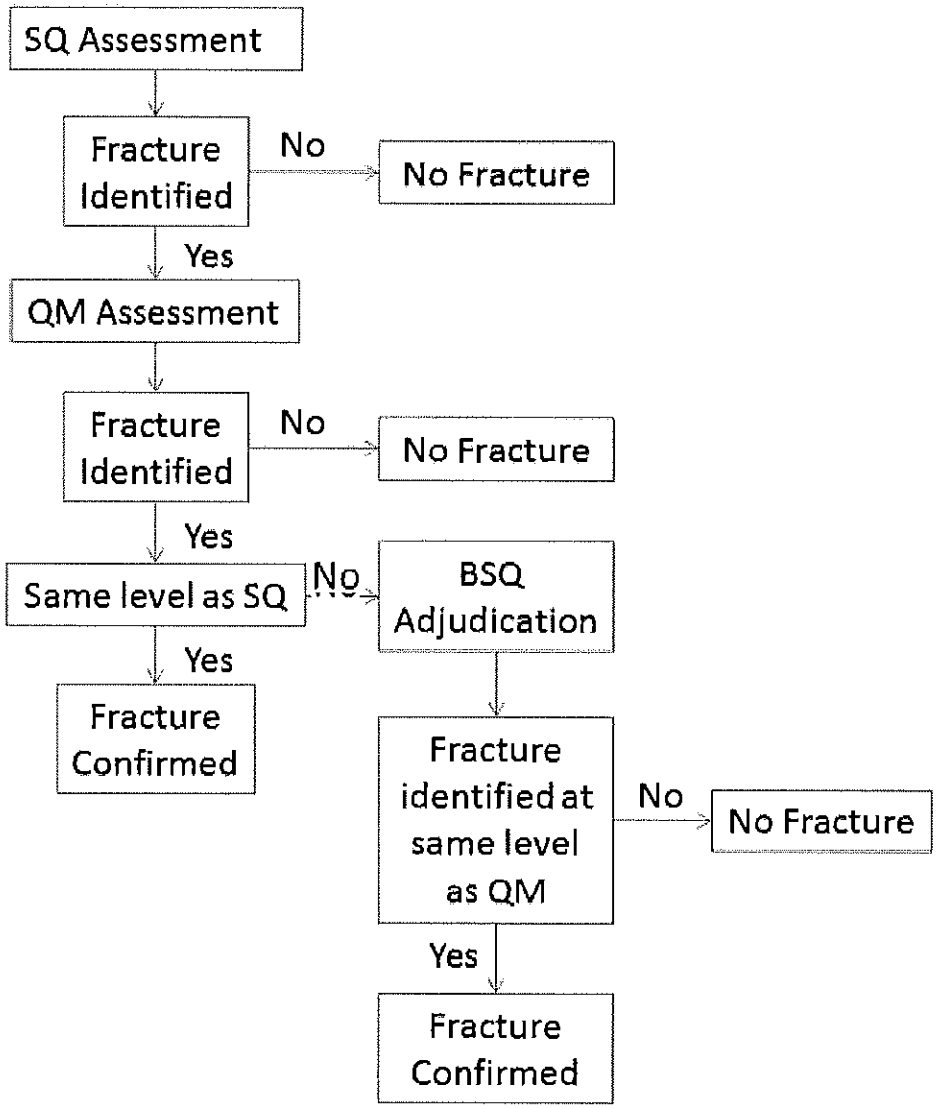
Finite Element Analysis

Finite element analysis of HR-pQCT scans will be conducted by Merck & Co. BioClinica will provide Merck with the necessary images. The results of the FEA will be returned to BioClinica and added to the BioClinica study database. These data will be transferred to Merck along with the data generated by BioClinica.

6.2 Temporal Presentation of the Images During Analysis

The presence or absence of incident vertebral fractures will be assessed by BioClinica on spine radiographs using both semiquantitative and quantitative (morphometric) criteria. All visits of all subjects will be read by the SQ method to detect prevalent and incident vertebral fractures. QM reading will be performed on all subjects who have an incident fracture identified by the SQ method. BSQ reading will be performed on all subjects where there is a discrepancy between SQ and QM on the presence of an incident vertebral fracture (specifically, when the SQ reader does not find an incident fracture but QM reader does). Figure 4 contains a flow chart showing how the three reading methods are used together.

Figure 4: Flow Chart of the Vertebral Fracture Reading Schema



All subjects are required to have lateral spine radiographs at Screening, Month 6, and Month 12 and at subsequent yearly intervals. Reading will begin when the Month 24 radiograph arrives. This section describes how the visits are read by each method. Table 9 summarizes this reading choreography.

6.2.1 Reading Visits Up to Month 24

The first assessment of incident vertebral fracture will be made once a randomized subject has reached one of the following milestones: 1) completed their Month 24 visit; 2) discontinued prior to the Month 24 visit; 3) had an unscheduled spine radiograph examination. In the nominal situation comprising subjects with Screening, Month 6, Month 12, and Month 24 lateral spine radiographs, BioClinica will first conduct a

semiquantitative (SQ) assessment of incident vertebral fracture on the Screening and Month 24 radiographs. If an incident vertebral fracture is identified by the SQ reader in the Month 24 visit, the SQ assessment will be extended to both the Month 6 and Month 12 visits to determine whether the fracture first appeared in one of these earlier visits. The same process will be followed if the triggering event is an early termination visit or an unscheduled visit rather than the scheduled Month 24 visit.

If an incident vertebral fracture is found by the SQ reading, the same visits that are read by SQ will be displayed and read by the QM reader. If there is a discrepancy between the SQ and QM outcome for incident fracture (see Section 6.5 for details), the same visits will be read by BSQ.

In summary, the following rules describe which visits and which vertebrae are read by QM and BSQ for a given subject.

- When required by the presence of an incident fracture found by SQ at any visit, QM will be performed at all visits that have had an SQ reading.
- When required by the presence of a discrepancy between SQ and QM at any visit, BSQ will be performed at all visits that have had an SQ and QM reading.
- QM is performed on all 13 vertebrae at each visit and not just the vertebra(e) found to have an incident fracture by SQ.
- BSQ is performed on all 13 vertebrae at each visit and not just the vertebra(e) found to have a discrepancy between SQ and QM.

The first two rules provide consistency and ensure that the SQ, QM and BSQ readers are all presented with the same information (i.e., images) when making fracture assessments. It also provides for correction of false negative SQ results at other visits ensuring that the timing of the incident vertebral fracture is accurately recorded. The second two rules provide for blinding of the QM and BSQ reader to the other reading results and allow correction of false negative SQ results at other vertebrae in the subject.

Handling Unreadable Vertebrae and Indeterminate SQ Results

From time to time image quality might be suboptimal due to patient anatomy or problems with the image acquisition making some vertebrae unreadable. In these cases the incident fracture status might be indeterminate. When reading the Screening and Month 24 visits, these limitations can be overcome by superior image quality in the radiographs of the follow up visits visit. Therefore, an indeterminate incident fracture outcome at Month 24 will trigger SQ reading of follow up visits in the same was as an incident fracture would. As above, QM reading will only be initiated if the SQ reading identifies an incident vertebral fracture at one of these visits.

6.2.2 Reading Visits after Month 24

Regardless of the incident fracture results, all visits received after Month 24 will be presented to the reader(s) and read in the same manner as the previous visits. This may include only SQ reading, SQ and QM reading or all three reading methods. The radiologist reading the images from the current visit will not necessarily be the same radiologist who read previous visits.

If an incident vertebral fracture is not identified at the Month 24 visit, subsequent visits will be read by the same procedure except that if an incident vertebral fracture is confirmed after Month 24 (e.g., at Month 36), earlier visits (i.e., those between Screening and Month 24) will not be read by any method. The only exception is when these visits were read by SQ due to indeterminate incident fracture outcomes at Month 24. See Table 9 for details.

Table 9: Vertebral Fracture Reading Choreography.

	SCR	M06	M12	M24	M36	M48	M60	M72	M84	M96	M108	M120	M132	M144	COMMENT
SQ	X			N											If no SQ incident fracture is found at M24, do not read prior visits and, do not perform QM.
QM	-			-											
SQ	X			N	N	N									If no SQ incident fracture is found at any visit, do not perform QM.
QM	-			-	-	-	N	N	N	N	N	N	N	N	
							-	-	-	-	-	-	-	-	
SQ	X			N	Y	Y									If SQ incident fracture is found after M24, do not read M06 or M12 visits. Perform QM on all visits read by SQ.
QM	X			X	X	X	Y	Y	Y	Y	Y	Y	Y	Y	
							X	X	X	X	X	X	X	X	
SQ	X	X	X	Y											If an SQ incident fracture is found at M24, read M06 and M12 visits by SQ and QM to determine if fracture was present at an earlier time.
QM	X	X	X	X											
SQ	X	X	X	Y	Y	Y									Once QM reading is triggered by an SQ incident fracture, continue reading QM at all subsequent visits.
QM	X	X	X	X	X	X	Y	Y	Y	Y	Y	Y	Y	Y	
							X	X	X	X	X	X	X	X	
SQ	X	N	N	U											If the SQ incident fracture status is indeterminate at M24 because a vertebra was unreadable, read the prior visits by SQ to check for incident fracture. If none found, do not perform QM.
QM	-	-	-	-											
SQ	X	N	Y	U											If the SQ incident fracture status is indeterminate at M24 because a vertebra was unreadable, read the prior visits by SQ to check for incident fracture. If SQ incident fracture found, perform QM on all visits.
QM	X	X	X	X											
SQ	X	N	N	U	N	N									If no SQ incident fracture is found at any visit, do not perform QM.
QM	-	-	-	-	-	-	N	N	N	N	N	N	N	N	
							-	-	-	-	-	-	-	-	
SQ	X	N	N	U	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	If SQ incident fracture found at a visit after M24, perform QM on all visits.
QM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
SQ	X	N	Y	U	X	X									If the SQ incident fracture status is indeterminate at M24 because a vertebra was unreadable, read the prior visits by SQ to check for incident fracture. If SQ fracture found, perform QM read on all visits.
QM	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
							X	X	X	X	X	X	X	X	

Y = Yes, SQ fracture present; N = No, SQ fracture not present; U = SQ incident fracture status unknown due to unreadable vertebra

X = Perform SQ or QM at this visit

SCR = Screening visit; MXX = Month XX visit

DXA and HR-pQCT scans will be presented to the reader one-by-one as they are received. The reader will compare each scan to its corresponding baseline scan at the time of reading.

QCT scans will be analyzed in batches, with all spine scans or hip scans from a subject analyzed at the same time. The analysis will begin soon after the first subject reaches the last QCT visit and their data begin to arrive at BioClinica.

6.3 Recording Results of Image Assessment

6.3.1 Image Quality Assessments

Image quality assessments will not be performed at the time of review since this assessment will have been performed at the time of image receipt. Reviewers will have the option to comment on image quality findings.

6.3.2 Description of Assessments in Relation to Study Design

Details of the image analysis and measurements in relation to study design are provided in Section 5.1.

6.4 Description of the Reading Schedule

The analysis for DXA will be performed on-going as the imaging time points are collected and made available for a given subject. For spine radiographs, the Month 24 is the first post-baseline visit to be read along with all subsequent annual visits. For QCT and HR-pQCT visits will be grouped and batched for analysis prior to the requested data transfers.

6.5 Implementation of Adjudication Process

All subjects identified by the SQ reader to have an incident vertebral fracture will also be read using the QM method on all vertebrae (T4 to L4) at all visits assessed by the SQ reader. The QM reader will not know which vertebra was found to have an incident fracture by the SQ reader. If the QM reading agrees with the SQ reading on the presence of an incident vertebral fracture in a particular vertebral body, then the incident fracture is confirmed. If the QM reading does not confirm the SQ incident fracture, the SQ result is overruled and no fracture is reported.

If the QM reader finds an incident vertebral fracture in a vertebra where the SQ reader did not, then the BSQ reading will be performed on all visits assessed by SQ and QM in order to adjudicate the discrepancy. If the BSQ reading agrees with the QM reading on the presence of an incident vertebral fracture in a particular vertebral body, then the incident fracture is confirmed. If the QM reading does not confirm the BSQ incident fracture, the BSQ result is overruled and no fracture is reported.

In summary, an incident vertebral fracture must be identified by the QM reading and either the SQ reading or the BSQ reading in order to be confirmed. A QM result of no incident fracture will overrule any combination of positive SQ and BSQ incident fracture outcomes. The various scenarios for SQ, QM and BSQ incident fracture results and how they lead to the final adjudicated outcome are detailed in Table 10.

Table 10: Final Adjudicated Incident Vertebral Fracture Outcomes Based on the SQ, QM and BSQ Reading.

SQ	QM	BSQ	ADJ	COMMENT
N	-	-	N	No incident fracture by SQ at any vertebral level; no further reading required.
Y	Y	-	Y	QM confirms the SQ incident fracture; adjudicated outcome is YES.
Y	N	-	N	QM does not confirm the SQ incident fracture; adjudicated outcome is NO.
N	Y	N	N	QM finds an incident fracture that SQ did not; BSQ does not confirm the incident fracture.
N	Y	Y	Y	QM finds an incident fracture that SQ did not; BSQ confirms the incident fracture.
N	N	-	N	QM not required but performed due to SQ=Y at another level; adjudicated outcome is NO.
Y	N	Y	N	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.
Y	N	N	N	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.
Y	Y	Y	Y	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.
Y	Y	N	Y	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.
N	N	Y	N	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.
N	N	N	N	BSQ not required but performed due to SQ-QM discrepancy at another level; incident fracture outcome determined by SQ & QM.

Y = Yes, incident vertebral fracture present; N = No, incident vertebral fracture not present.

There will be no adjudication of the measurements by QCT, HR-pQCT and DXA.

6.6 Process for Database Lock of Image Assessments

Once completed, the image reading and analysis results will be locked into the database and the reviewers will not be able to amend them unless a re-assessment of the review is needed; for example, new imaging data being made available or changed by the site, Merck, or BioClinica as a result of data reconciliation. These deviations will be documented.

Prior to transferring the data to Merck, BioClinica will perform various data cleaning procedures. These will include edit checks and consistency checks as outlined in the

Data Validation Plan. Edit checks will include checks of the completeness and correctness of the data. Discrepancies will be identified and corrected.

After all edit checks have been applied and all issues resolved, the BioClinica database will be frozen and audited. Corrective action that results from the audit will be executed and verified by the quality assurance department. Once the BioClinica database passes an internal audit, it will be locked and the final transfer will be sent to Merck. BioClinica will obtain signed approval from Merck if it is necessary to make changes to the data in final locked database.

6.7 Reviewer Performance and Quality Metrics

6.7.1 Inter-reader Variability

Inter-reader variability for vertebral fracture is assessed annually in the SQ and QM reader qualification process described below.

DXA, QCT and HR-pQCT inter-reader variability is assessed via the reader qualification procedure.

6.7.2 Intra-reader Variability

Intra-reader variability will not be assessed in the context of this trial.

6.8 Methodology for Export of Results to Sponsor and Regulatory Authorities

6.8.1 Database Set-up and Maintenance

The customization and initialization of the database by BioClinica will include the creation of all necessary validation documents required by regulatory authorities. These validation documents may include requirement specifications, design documentation, source code (written to coding standards) and the validation test plan and report.

After customization and initialization the database will be maintained throughout the study. User accounts will be assigned and the database will be periodically backed up and archived to protect against data loss.

6.8.2 Data Transfer to Sponsor

BioClinica will transfer the results of the readers' assessments to Merck according to a predefined schedule. Prior to the first transfer, BioClinica and Merck will develop a Data Transfer Specification (DTS) to define the content and format of the data transfer file. Data transfers for final analysis will be subject to additional review and/or audit prior to transfer. For each transfer, BioClinica will create, validate and send the file as outlined in the DTS. Merck will load the transferred file into the study database.

Merck may generate and send queries to BioClinica requesting resolution of any remaining data discrepancies. BioClinica will investigate and correct the data in the BioClinica database as necessary. The corrected data will be sent to Merck in the subsequent transfer, except for the last scheduled transfer in which the changes will be completed prior to sending the final transfer. The records of all transfer files sent to Merck from BioClinica will be stored by BioClinica.

6.8.3 Data Storage and Back-up

All data will be stored on BioClinica's servers which have full daily back-ups and are archived offsite.

7 SELECTION AND TRAINING OF INDEPENDENT RADIOLOGISTS

7.1 Reviewer Selection

BioClinica will identify radiologists as well as DXA and HR-pQCT technicians for this study from the pool of qualified BioClinica employees. Both the radiologists and the technicians are qualified by BioClinica and have experience in the assessment of vertebral fracture, in the measurement of BMD with DXA and the analysis of QCT and HR-pQCT images. The radiologists and technicians will not be affiliated with Merck in any way other than their participation in this central reading. This includes current employment at Merck, present consulting arrangements with Merck or equity ownership in Merck. BioClinica will maintain the CVs and qualifications of each of the radiologists and technicians performing any work under this charter.

7.2 Reviewer Training and Qualification

Prior to initiating readings, the radiologists, QCT,HR-pQCT and DXA technicians will undergo formal training on the clinical protocol, imaging charter and reading system to be used. Radiologists, QCT, HR-pQCT and DXA technicians will also be trained on the reading/measurement methodology in one or more joint sessions using image data from a training set in order to standardize reader performance and reduce variability. Radiologists, QCT, HR-pQCT and DXA technicians must participate in these training activities and achieve a specified level of inter-reader consistency before being allowed to work on this study.

The qualification of the radiologists, QCT, HR-pQCT and DXA technicians will be performed in 2 phases: standardization and qualification.

7.2.1 Standardization Step

A training set of spine radiographs or DXA, QCT and HR-pQCT scans of subjects will be reviewed and evaluated in consensus in a joint session by the participating reviewers using the appropriate software and workstation. Analysis criteria will be discussed in detail. Special circumstances and anatomical variants are reviewed. The prospective reviewer will practice reading, will analyze images and will watch another reviewer perform this work to gain additional training and experience.

7.2.2 Qualification Step

After the training activity and practice sessions, the prospective reviewers will independently read or analyze a qualification set of images. The prospective reviewer will be evaluated against a gold standard qualification that was set by comparing the results of other reviewers.

7.2.3 Analysis of the Qualification Review Results

The qualification test results of each prospective reviewer will be collected and analyzed using appropriate statistical methods (e.g., the kappa statistic for vertebral fracture; coefficient of variation or intra-class correlation coefficient for DXA, QCT and HR-pQCT results). Each reviewer must achieve a specified level of agreement with the gold standard results before being allowed to work on this study. These performance thresholds vary by the type of assessment and are defined in BioClinica's SOPs.

If the specified level of agreement is not achieved by the reviewer, retraining sessions will be organized to improve the reviewer's performance. After re-training, the reviewer must evaluate the same qualification set again and must show an acceptable performance before being allowed to work on this study.

7.3 Procedures for Replacing Reviewers

In the event that a radiologist or technician is unable to continue his/her readings, the readings will either be completed by the remaining reviewer(s) or a new reviewer will be added to the pool. All new reviewers will go through the same training and qualification process before working on this study. If a reviewer is replaced due to inadequate performance, data produced by that reviewer will be evaluated to determine if none, some, or all of the cases must be reviewed or completely re-read.

8 REVISION HISTORY

Version No.	Author	Date	Description
0.1	Tim Gibson	15-Mar-2012	First draft
0.2	Ken Gaither	16-Mar-2012	Review and updates
0.3	Tom Fuerst	23-Mar-2012	Review and updates; expanded on vertebral fracture assessment.
0.4	Tim Gibson	29-Mar-2012	Incorporate additional comments from review.
0.5	Tom Fuerst and Tim Gibson	06-Apr-2012	Incorporate and address comments from Merck
0.6	Tom Fuerst and Ken Gaither	09-Jul-2012	Add DXA machine change info, and additional reading/analysis clarity
0.7	Tom Fuerst and Tim Gibson	21-Sep-2012	Address comments from Merck
0.8	Tom Fuerst and Tim Gibson	28-Oct-2012	Address comments from Merck
0.9	Tom Fuerst and Tim Gibson	19-Nov-2012	Incorporate additional comments from Merck
0.91	Tom Fuerst	18-Dec-2012	Incorporate Merck Comments
1.0	Tim Gibson	18-Jan-2013	General edit, format and review; finalize for signatures
1.1	Emily Duncan	18-Sep-2014	Update to BioClinica update for MK0822-018 second extension
1.2	Emily Duncan	25-Sep-2014	Included updated office locations
2.0	Emily Duncan	26-Sep-2014	Updated Merck contact information and finalized for signatures

9 TECHNICAL/SUPPLEMENTAL SUPPORTING DOCUMENTATION AVAILABLE ON REQUEST

Image Acquisition Guidelines

- *Imaging manual and data transmittal forms*
- *Image site technical evaluations (Site quality procedures)*
- *Digital image data translation*
- *Film digitization*
- *Electronic CRF Screenshots*
- *Programming Specification Documents*
- *Read application*
- *Data export*
- *Database QC Procedure*
- *Database validation plan*

Sample Supplemental Documents

- *Radiation dosage*
- *More detailed depiction of Image review system configuration*
- *Communication plan*
- *Database QC and transmission to client*
- *Data transfer specification document*
- *Data tracking design*
- *Database design and programming*