

Insights from the conduct of a device trial in older persons: low magnitude mechanical stimulation for musculoskeletal health

Douglas P Kiel^a, Marian T Hannan^a, Bruce A Barton^b, Mary L Bouxsein^c, Thomas F Lang^d, Kathleen M Brown^b, Elizabeth Shane^e, Jay Magaziner^f, Sheryl Zimmerman^g and Clinton T Rubin^h

Background Osteoporosis is a common complication of aging. Alternatives to pharmacologic treatment are needed for older adults. Nonpharmacologic treatment with low magnitude, high frequency mechanical stimulation has been shown to prevent bone loss in animal and human studies.

Methods The VIBES (Vibration to Improve Bone Density in Elderly Subjects) study is a randomized, double-blind, sham-controlled trial of the efficacy of low magnitude, high frequency mechanical stimulation in 200 men and women aged 60 years and older with bone mineral density T-scores by dual X-ray absorptiometry between -1 and -2.5 at entry. Participants are healthy, cognitively intact residents of independent living communities in the Boston area who receive free calcium and Vitamin D supplements. They are randomly assigned to active or sham treatment and stand on their assigned platform once daily for 10 min. All platforms have adherence data collection software downloadable to a laptop computer. Adverse events are closely monitored. 174 participants were randomized and will be followed for 2 years. Almost all active subjects have attained 1 year of follow-up. Bone mineral density is measured by both dual X-ray absorptiometry and quantitative computed tomography at baseline and annually. The main analysis will compare mean changes from baseline in volumetric bone density by quantitative computed tomography in active and sham groups. Adherence and treatment effect magnitude will also be evaluated. Secondary analyses will compare changes in two biochemical markers of bone turnover as well as longitudinal comparisons of muscle and balance endpoints.

Results The VIBES trial has completed its first year of data collection and encountered multiple challenges leading to valuable lessons learned about the areas of recruitment from independent living communities, deployment of multi-user mechanical devices using radio frequency identification cards and electronic adherence monitoring, organization of transportation for imaging at a central site, and the expansion of study aims to include additional musculoskeletal outcomes.

Conclusions These lessons will guide future investigations in studies of individuals of advanced age. *Clinical Trials* 2010; 0: 1–14. <http://ctj.sagepub.com>

^aInstitute for Aging Research, Hebrew SeniorLife, Boston, MA, USA, ^bMaryland Medical Research Institute, Baltimore, MD, USA, ^cCenter for Advanced Orthopaedic Studies, Beth Israel Deaconess Medical Center, Boston, MA, USA, ^dRadiology, UCSF, San Francisco, CA, USA, ^eDepartment of Medicine, Columbia University College of Physicians & Surgeons, New York, NY, USA, ^fEpidemiology & Preventive Medicine, University of Maryland, Baltimore, MD, USA, ^gCecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, NC, USA, ^hBiomedical Engineering, SUNY Stony Brook, Stony Brook, NY, USA

Author for correspondence: Douglas P Kiel, Institute for Aging Research, Hebrew SeniorLife, 1200 Centre Street, Boston, MA 02131, USA. E-mail: kiel@hrca.harvard.edu

Introduction

To test the potential benefits of low-level mechanical intervention in a clinically relevant population at risk for osteoporosis, we designed a 2-year, double-blind, randomized, sham-controlled clinical trial of low magnitude mechanical stimulation (LMMS) in 200 women and men 60 years and over with low bone mineral density (BMD). Entitled Vibration to Improve Bone Density in Elderly Subjects and referred to as VIBES, the clinical trial is designed to evaluate the efficacy of LMMS in the treatment of low BMD, and other indices of musculoskeletal health, including balance, muscle mass, and muscle strength.

Osteoporosis, a disease characterized by the progressive loss of bone, is one of the most common complications of aging [1]. To date, prevention of bone loss has been approached principally through pharmacologic interventions, which have been shown to significantly reduce fracture risk, yet these drugs target the skeleton without affecting other risk factors for fracture such as muscle, balance, and falls. Also despite the proven efficacy of these drugs, many patients are reluctant to commit to long-term therapy because of significant side effects [2], and compliance is very poor [3]. Particularly in elderly individuals, it is desirable to treat bone loss without pharmacologic intervention whenever possible because older persons frequently take multiple medications, and thus are at an increased risk for drug interactions and polypharmacy effects.

In addition to bone loss, older patients also may suffer from a loss of muscle strength and age-related changes in neuromuscular control, which also contribute to the risk for falls and fractures. There has been considerable interest in the role of environmental stimuli, such as exercise, in preserving bone mass and morphology as well as muscle strength [4]. The most effective approach to fracture prevention in older patients may be one that targets BMD and reduces the risk for falls [5] by improving balance, muscle mass, and neuromuscular control. The VIBES trial was designed specifically to test a unique device that delivered low magnitude mechanical stimulation capable of stimulating bone formation [6,7] and potentially increasing muscle mass [8] in vulnerable older people living independently in residential communities.

Methods

Eligibility and recruitment

The VIBES trial involves men and women over the age of 60 years (with no upper limit on age) with

moderately low BMD. One of its unique aspects is the recruitment of individuals over the age of 80 years who typically are not included in trials of osteoporosis prevention. Another unique aspect of recruitment is the use of independent living communities (ILCs). ILCs make it possible to: (1) install LMMS devices in 'common' areas that attract the attention of residents who may want to participate; (2) share equipment (a key feature to optimize resources); and (3) promote adherence to an intervention through communal social relations. ILCs may be a single building or a campus of buildings where large numbers of older persons reside. As congregate communities, they typically offer private apartments, shared meals, housekeeping, laundry, transportation, and access to medical care. ILCs are a thriving housing option for older adults and are expected to house as many as 18% of adults age 75 and older by 2020. They already have become popular study sites for health-related research [9,10]; however, they have not been used to conduct randomized, controlled, clinical trials. The experience gained from the VIBES trial should be valuable to other investigators planning clinical trials in these popular residential settings.

Inclusion and exclusion criteria ensure that trial participants do not have advanced osteoporosis and are healthy enough to complete a 2-year follow-up. Eligible trial participants are not taking pharmacologic treatment for osteoporosis, and have osteopenic BMD with T-scores between 1 and 2.5 standard deviations below young normal values in either the total hip, femoral neck, trochanter, or lumbar spine. To maximize adherence with the assigned intervention, candidates with cognitive dysfunction (scores greater than 12 on the short blessed test) are not enrolled [11]. ILC residents with BMD T-scores worse than -2.5 or with fragility fracture within the past 5 years were excluded unless they had no pharmacologic treatment options. Table 1 presents a complete list of inclusion and exclusion criteria.

Participants were recruited only within ILCs. ILC enrollment required a formal discussion of the study between community leadership and VIBES investigators. Prior to contact with community residents, an advertising campaign using flyers, public announcements, and community TV presentations was conducted. Following the campaign, the investigators and the VIBES staff presented a community-specific information session with the residents. Residents were presented with a study overview, given a chance to ask questions, and encouraged to participate. For residents unable to attend the information session, a second, informal session with refreshments was held later.

Table 1 Inclusion and exclusion criteria used for the VIBES randomized sham-controlled trial of Low Magnitude Mechanical Stimulation**Inclusion criteria:**

English speaking: Ability to speak and understand English.

Age/Gender: Men and women ≥ 60 years.

General health: Absence of terminal illness (no illness expected to result in death within the next 12 months).

Bone mineral density: A BMD gender specific *T*-score in either the total hip, femoral neck, trochanter, or L1–L4 spine between -1 and -2.49 inclusive. Subjects with a *T*-score of -2.5 or less are included if they have no options for osteoporosis medications.

Cognitive status: Capable of following protocol and providing informed consent; scoring <12 on the Short Blessed Test of cognitive function.

Exclusion criteria:

Ponderosity: Subjects who weigh 250 pounds or more.

Physical activity: Immobilization within the last year, nonambulatory, or prolonged bed rest for greater than 3 months in the last year.

Malignancy or renal disease: Malignancy other than cured thyroid cancer or skin cancer. Estimated glomerular filtration rate below 30.

Orthopedic: Bilateral hip replacement ever and hip replacement or internal fixation, total knee replacement, both hip and spine (L1–L2) surgical hardware, or lower limb amputation, or a history of a fragility fracture within the past 5 years except if the subject has no options for osteoporosis medications.

Medications: Glucocorticoids, suppressive doses of thyroid hormone as determined by screening TSH, phenytoin, phenobarbital, carbamazepine, estrogen/testosterone replacement, SERMs, PTH, or bisphosphonates >1 month in the past year, calcitonin therapy within the preceding month, fluoride therapy at any time, rosiglitazone, pioglitazone, inhaled corticosteroids greater than a prednisone equivalent dose of 5 mg per day.

Laboratory values: 25-hydroxyvitamin D levels less than 15 ng/mL, calcium level greater than the upper limit of the laboratory range.

Availability: Not available for ≥ 3 months per year because of travel to warmer climate or other seasonal residence.

Other: Paget's disease, hyperparathyroidism, rheumatoid arthritis or other connective tissue disorders requiring systemic treatment with disease modifying drugs, or a history of Cushing's syndrome by history.

Screening was a multistage process, beginning with simple questions such as maintaining residence in the community for the full calendar year, over the next 2 years, interest in the study, willingness to stand on a vibrating platform for 10 min per day, and use of osteoporosis drugs. Pre-screening was followed by the administration of informed consent and a detailed review of remaining eligibility criteria. Residents who were judged eligible after further questioning and cognitive function screening were scheduled for a dual-energy X-ray absorptiometry (DXA) scan at the centrally located hospital, which performed these measures for all ILCs. Transportation was provided and VIBES staff members accompanied groups of residents. Following the DXA scan, eligible trial participants had phlebotomy for screening labs, including a 25 hydroxy-vitamin D concentration, renal and liver function, and thyroid stimulating hormone concentration. Residents with vitamin D-values less than 15 ng/mL were offered vitamin D supplements and re-screened at a later date. Regardless of eligibility, residents with abnormal findings at screening are immediately brought to the attention of their regular physician. Residents who were eligible on all criteria in Table 1 were randomized and then underwent a baseline quantitative computed tomography (QCT) scan of the hip and spine. All randomized participants were provided with free calcium (1000 mg) and vitamin D (800 IU) tablets for the duration of the trial.

Clinical measures

Clinical measures collected at baseline, 12 and 24 months are listed in Table 2. The primary outcome variables are QCT measures of volumetric trabecular bone density of the spine and hip. DXA measures are used only for screening and safety follow-up. Secondary endpoints include biochemical markers of bone turnover and musculoskeletal health measures including standing balance, leg extension strength, and self-reported falls.

Volumetric bone density and geometry by quantitative computed tomography

QCT is being used to evaluate volumetric bone density (vBMD) and geometric indices at the lumbar spine and proximal femur. Volumetric (or 3D) QCT imaging offers advantages over traditional DXA BMD measurements because: (1) it can assess trabecular and cortical bone compartments separately; (2) it can assess bone geometry, and (3) measurements are not confounded by bone size [12]. Previous studies have shown that the responses to therapeutic intervention are greater when assessed by QCT-based vertebral trabecular bone measurements than by DXA-based spine BMD [13,14], and thus QCT may be able to detect changes in trabecular bone that would not be apparent by DXA-based BMD measures. QCT is

Table 2 Schedule of data collection for participants in the VIBES randomized sham-controlled, trial of low magnitude mechanical stimulation

	Screen	Baseline	Follow-up month									
Month	-2	0	1	3	6	9	12	15	18	21	24	
Assessment												
Telephone call			x	x	X	x		x	x	x		
Eligibility assessment	x											
Consent	x											
Physical activity, dietary calcium intake, medical history		x					x				x	
Bone markers		x	x	x	X		x				x	
DXA BMD (hip and spine)	x						x				x	
QCT (hip and spine)		x					x				x	
Muscle strength knee extensors		x					x				x	
Quiet standing balance		x					x				x	
Performance measures		x					x				x	
Dispense calcium, vitamin D		x					x					
Vitality Plus Questionnaire		x					x				x	
Adverse events			x	x	x	x	x	x	x	x	x	
Falls/hospitalizations (bi-monthly)		x	▶									
Adherence			x	x	x	x	x	x	x	x	x	

particularly useful for this trial since LMMS may preferentially impact trabecular bone [15], and QCT measures not only bone density, but also bone morphology.

Trial participants have QCT scanning of the proximal femur and lumbar spine (L1, L2) at baseline, 12, and 24 months using a helical CT scanner operating at 120 kVp, 150 mAs, 48 mm field of view and 1 mm slice thickness. They are scanned simultaneously with a bone mineral reference phantom (Image Analysis, Columbia, KY) that allows conversion of Hounsfield units to BMD in mg/cm³ hydroxyapatite. All QCT image data are evaluated using semi-automated methodology [13,16–18]. Outcomes include integral, trabecular and cortical volume, BMC and vBMD at these regions. The coefficient of variation (CV) for the ‘total hip’ trabecular volumetric BMD is 2.93% [19]. For the spine (L1, L2), outcomes include integral, trabecular and cortical tissue volume (cm³), bone mineral content (BMC, g), and vBMD (g/cm³) as well as cross-sectional area (Figure 1). The CV for mid-vertebral trabecular BMD is 2.93% [19]. For follow-up measurements, we are planning to use a version of the QCT analysis software in which the densitometric and structural analysis is integrated with a three-dimensional image registration program leading to improved precision [19]. The radiation exposure for QCT scans is higher than for DXA measurements; however, the exposure received from all of the radiographic procedures is equivalent to a uniform whole body exposure of 3.7 rem, or 74% of the annual limit allowed for a radiation worker such as an X-ray technologist. The exposure received in VIBES is comparable to other every day risks.

Dual-energy X-ray absorptiometry

DXA is used to screen the BMD of trial candidates as well as to monitor for significant bone loss 1 year after enrollment. At screening, residents undergo hip and spine scans to determine whether the BMD T-score is osteopenic (lower than -1 but greater than -2.5). The DXA is repeated after 1 year and the change in total hip BMD is calculated. Bone loss of greater than 8% at the total hip is a safety-monitoring trigger for the referral of the participant to his/her personal physician.

In the first version of the VIBES study design, we considered a multicenter approach with DXA-derived BMD as the primary outcome. The peer review of the design suggested that it was premature to use a multicenter approach. Consequently, by designating CT-derived bone density with its greater sensitivity to change over short time intervals as the primary outcome instead of DXA we reduced the scope of the trial to a single recruitment center with a smaller sample size. One of the advantages of having both DXA and QCT in the same individuals is that changes observed during the study in one imaging modality can be compared with the other. Previous osteoporosis clinical trials have relied exclusively on DXA for measurement of skeletal responses to treatment.

Biochemical markers of bone turnover

An important aim of this study is to use serum assays of biochemical markers of bone formation and resorption to understand the mechanism by

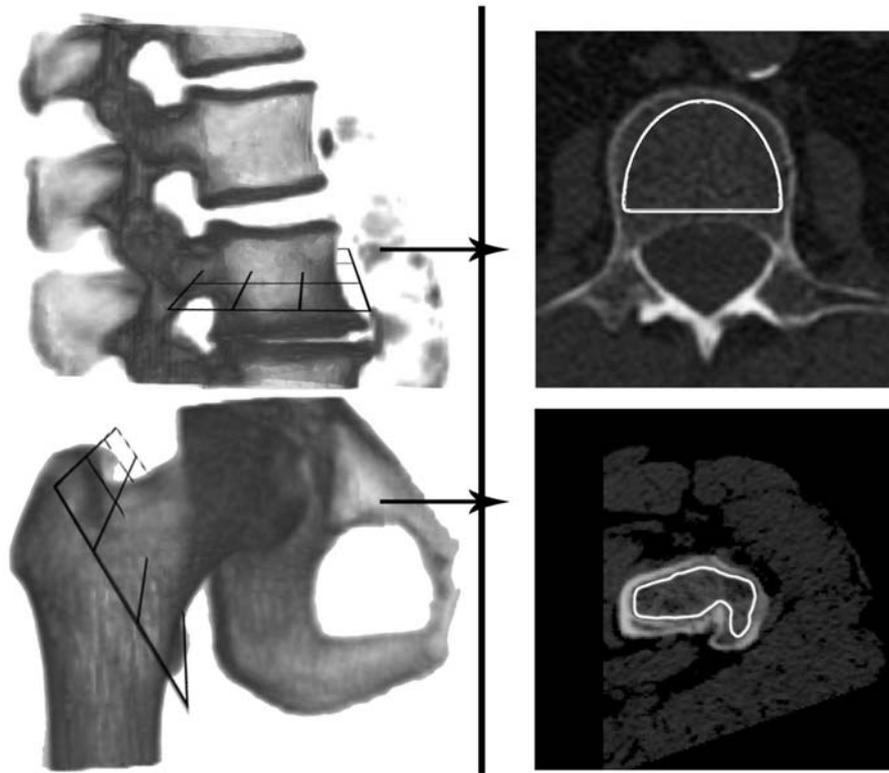


Figure 1 Representative images from quantitative CT scans performed on the lumbar spine and proximal femur. The planes shown for each region on the left are displayed in cross-section on the right

which LMMS increases bone mass. LMMS is anabolic to bone and may increase bone mass through an increase in bone formation. However, because there are few data on the effects of LMMS on bone turnover, our hypothesis is based on the effects of a well-established anabolic agent, parathyroid hormone (PTH), on bone turnover markers in studies of PTH in human volunteers with osteoporosis. PTH is associated with early and rapid increases in markers of bone formation, followed by increases in both bone resorption and bone formation [20]. The early phase when PTH stimulates bone formation to a greater extent than bone resorption, has been termed the ‘anabolic window’ [21].

To study the effects of LMMS on bone remodeling, we are storing serum at -70°C to measure bone formation and resorption markers. Procollagen type 1N-terminal peptide (P1NP) is cleaved from newly formed collagen type 1 polypeptide and reflects the earliest phase of bone formation. It is also the formation marker that rises earliest in response to the anabolic effects of intermittent PTH injections [21]. The bone resorption marker, serum C-terminal telopeptide of type I collagen (CTX), originates from the nonhelical region of the collagen molecule. Serum P1NP and CTX are measured

at baseline, and at 1, 3, 6, 12, and 24 months in both treatment groups. Specimens will be thawed once and processed batch-wise in single assay runs. If LMMS functions as an anabolic agent, we expect to observe increases in both markers in the following order: P1NP first, followed by CTX. An increase of 20%, 1/3 to 1/6 of that observed with PTH, will be clinically important and the study is powered to detect treatment group differences of this magnitude.

Postural stability

In the original VIBES design, postural stability was not included. Subsequently, more information had emerged suggesting that LMMS may have salutary effects on the neuromuscular system by providing the higher frequency of muscle stimuli that typically declines due to age-related loss of type II muscle fibers. The type II muscle fiber is involved in maintaining upright posture. Therefore, we added relatively inexpensive and noninvasive measures of postural stability at baseline, and at 1- and 2-year follow-up examinations, as a means of assessing whether LMMS helps to slow the loss of postural

stability that typically parallels aging. Center of pressure measurements are performed on an eight-channel force plate (Kistler, Winterthur, Switzerland) using an eight-channel amplifier, an analogue-digital converter, and Bioware 3.2.6.104 software. Data are sampled at 1000 Hz, low-pass filtered at 50 Hz, and stored for later analysis.

Participants are instructed to stand as still as possible with hands at sides, eyes open and directed ahead, and feet placed at shoulder width. Four minutes of data are collected during this quiet stance. The collection of postural stability data is automated and variables are defined using a custom MATLAB program (version 7.0.1, The MathWorks, Natick, Massachusetts). Operators are blinded to participant treatment group to minimize bias.

Muscle strength by isometric hand-held dynamometry

Based on observations in younger individuals that LMMS increases muscle area [8], we use an inexpensive measure of muscle strength with good reliability and minimal participant burden. An isometric hand-held dynamometer (Lafayette Manual Muscle Test System Model 01163) is used to measure right leg extension strength at baseline, 12 and 24 months. It is placed on the anterior surface of the tibia 6 cm above the lateral malleolus. Participants practice one leg extension before three measures of muscle strength are recorded.

Self-reported falls

Various approaches to ascertainment of falls were considered. Based on previous experience from a longitudinal study of falls [22], it was recognized that monthly falls calendars would create undue workloads on the VIBES staff and trial participants. We collect self-reported falls every other month using a brief self-administered questionnaire that is returned by mail. When fall report questionnaires are not returned, VIBES staff members contact study participants to remind them. We initially were concerned that trial participants would be reluctant to divulge their personal experience of falling, for fear that their report could lead to placement in a more supervised residential setting. Instead we found that we have almost perfect compliance with the fall forms, and that many of the participants openly discuss their falls and their fears of falling and injury resulting from them.

Randomization

Trial participants are enrolled into the study and randomized to active or sham intervention after successfully completing all of the screening

measures. Treatment group assignment and platforms are color-coded so that study participants and VIBES staff remain blinded to treatment assignment.

Randomization is stratified by ILC, gender, and BMI to ensure balance between treatment groups. Two BMI randomization strata are based upon the median BMI for NHANES III participants 65 years and older with the same T-scores used in VIBES (BMI = 24.0). Because of the potential for relatively few individuals in a community to pass screening, the randomization schedule was designed with several small blocks at the beginning of each stratum to avoid potential treatment imbalance because too few participants were assigned to one or more of the four stratification groups.

Current status of the vibes trial

Enrollment

As of May 2009, 21 independent living communities had been enrolled, 304 participants had been screened and 174 were randomized.

Treatment with low magnitude mechanical stimulation

All participants stand on their assigned platform for one 10-min session each day. At least two platforms are available in each ILC and they are used by study participants for their sessions at any time throughout the day. It has been shown that daily 10-min periods of a high frequency, low magnitude mechanical signal were sufficient to stimulate bone formation in humans [8,23], and that compliance remained high given this challenge, even in an elderly population similar to the VIBES participants [24].

Platforms require access to a standard electrical outlet. The top plate of each device is suspended over the base by a set of compression springs. Force input to the platform is provided by a single, low force (18 N) electromagnetic linear actuator (BEI model LA18-18, San Marcos, CA) capable of imposing acceleration of up to 1 g (9.8 m/s²) on individuals weighing up to 100 kg over 400 Hz. Accelerometer feedback from the plate surface adjusts vertical displacement of the platform to compensate for participant motion or positional changes.

Active LMMS and sham control platforms are engineered to look and sound the same, and are programmed to activate sessions when the participant swipes his/her radio frequency identification



Figure 2 The low magnitude mechanical stimulation plate with a surrounding frame for ensuring stability while standing

card (RFID) card (Figure 2). The engineering and presentation of sham platforms is one of the unique achievements of the VIBES trial. Because the high frequency (0.3 g) and low magnitude (30 Hz or cycles per second) of LMMS vibration is barely perceptible to participants standing on the active device, it was not difficult to construct sham platforms emitting similar audible humming sounds indistinguishable from active devices. In our pilot studies, participants could not discriminate between the active and sham platforms when queried at the end of a 6-month trial [24]. In addition to the configuration of the platforms, color-coded RFID cards ensure that each participant uses the correct platform. The RFID cards activate only the platform to which the subject is assigned. Nevertheless, there is a chance that if individuals become unblinded, they could share this with others at the site using the same assigned platform. To avoid this as much as possible, the investigators and staff inform the trial participants that they should not try to figure out which platform is active and which one is sham. In this age group, we have

observed that participants take these recommendations to heart and do not attempt to unblind others.

Adherence assessment

Adherence information (participant ID, time on, time off, total session time) is recorded electronically, stored by the platforms, and downloaded at regular intervals. This method of collection of electronic adherence data was developed by VIBES and has not been used in previous studies.

During the first week of study intervention in each ILC, research assistants were present throughout the 10-min sessions to familiarize participants with the safe operation of the equipment and use of RFID cards, and to reinforce daily use of the platforms and sign-in logs. To maximize adherence, participants are encouraged to choose the most convenient time of day to complete their 10-min session. They are told that the device automatically records the date, start and end time of their sessions

because of their unique RFID card. They are instructed not to exchange RFID cards with other participants or to use platforms not coded with their assigned color. Sign-in logs also are used to detect nonadherence. Participants whose adherence has declined over time are contacted to discuss reasons for nonadherence and options to improve adherence to the trial regimen. At each annual study examination, adherence to the recommended daily intakes of calcium and vitamin D are encouraged.

Adverse event monitoring and notification

Vibration, particularly in the frequency domain of 5–15 Hz where resonance of the spine can occur [25], is considered a key etiologic factor in low back pain [26], and a causal factor in circulatory disorders such as Raynaud's Syndrome, blurred vision, and hearing loss [27]. Research has focused on *attenuating* the transmission of whole body vibration to the skeleton, with the widely held presumption that high frequency vibrations are pathogenic to the musculoskeletal system [28], and may even cause percussive injuries to the brain [29]. In cases where vibration is unavoidable [30], exposure limits have been recommended by the National Institute of Occupational Safety and Health, (NIOSH), Centers for Disease Control (CDC), and the International Organization for Standardization (ISO) [31]. However, the level of vibration used in VIBES is very low (e.g., generating less than 10 microstrain on the cortical bone surface) [32], compared to the strain generated by walking (>1000 microstrain) [33]. The vibration level is 50× less than vibration levels used in studies of muscle building in athletes, where devices delivered 8–15 g of vibration, far exceeding the 1-min threshold limits mandated by ISO and NIOSH [28,34]. Therefore, adverse events in VIBES which can be attributed to the vibrating platform are not expected to be frequent or severe; however, participants aged 60 years and older are expected to have many unrelated comorbidities.

Balancing the low anticipated rate of adverse events due to LMMS with the relatively high rate of expected events related to the age and comorbidities of participants, the VIBES study uses a conservative definition of an adverse event as an undesirable medical occurrence (sign, symptom, or diagnosis) or worsening of a pre-existing medical condition that occurs after the beginning of treatment. Early experience with reported adverse events highlighted the potential for very large number of events in the VIBES age group. The workload created by having to record the large number of common age-related events, such as

musculoskeletal pain, taught us that a more specific pre-study definition of reportable events for a low risk device like the LMMS platform is preferable to a more standard broad definition used in trials of devices and drugs that have a greater potential for unexpected or unknown adverse events. The advanced imaging used in VIBES resulted in some incidentally discovered abnormalities on CT scans (i.e., tumor masses and aortic aneurysms) that necessitated the development of a systematic approach to the notification of health care providers. Nonserious adverse events for each participant are tracked from the first report to subsequent resolution with an event/report numbering system.

Serious adverse events (SAEs) are defined as significant hazards or side effects that may be fatal or life threatening; require in-patient hospitalization or prolongation of an existing hospitalization; be persistent or significant disabilities; or be other significant medical hazards. Detailed information is collected on SAEs, including a description of symptoms, possible causes, related medical conditions, relationship to study intervention, and narrative comments and assessments.

For both serious adverse events and nonserious adverse events, their relationship to the LMMS device is classified by the study physician as: definitely unrelated (virtually no possibility), unlikely to be related (very little probability), possibly related (a small chance), probably related (a reasonably high likelihood), or definitely related (a very strong likelihood). For all events related to the LMMS device, medical records are obtained using a signed release from the participant. All platform-related events are followed until they are resolved. An independent Data and Safety Monitoring Committee meets every 6 months to review adverse events and data quality.

Organization of the VIBES trial

Central imaging

The organization of the VIBES trial to transport all participants to a single central imaging facility was an important lesson learned. QCT imaging of volumetric bone density is subject to considerable measurement error due to machine differences and technologist differences, making the central imaging preferable to the use of multiple imaging centers. The use of multiple imaging centers might be perceived as having less participant burden in terms of travel time; however, we learned that even transportation times of 45 min are acceptable to participants, especially when groups of participants travel together. Group travel also

helped to create rapport with VIBES staff and among participants from a given ILC.

Quality assurance

All VIBES personnel attended a central training session prior to collecting of data and were certified in standard data collection techniques. To minimize the variability in the primary QCT outcomes, one primary technologist was designated to perform scans, with three back up technologists also trained in the protocol of QCT scan acquisition.

To ensure that all active LMMS platforms are delivering the appropriate mechanical signal, and to be able to replace faulty equipment, all devices are calibrated at baseline, and at monthly intervals throughout the study to monitor for the unlikely event of drift. Calibration is performed by an unblinded 'device monitor' who has no contact with study participants. We originally did not anticipate the need for such a person in the VIBES trial; however, we now recognize the importance of such personnel when technical monitoring of devices is required in a clinical trial.

Quality control of QCT scan analyses is carried out in the QCT analysis program at the central imaging hospital. As with analysis of DXA images, data are added to the program database only after a quality assurance stage in which the operator confirms that the bone regions determined by the analysis are acceptable, and the key BMD variables, calibration phantom slopes and Hounsfield Unit values are found to be within pre-determined constraints. To relate the BMD measurements from VIBES to other studies, the QCT machines were calibrated prior to beginning the study to reference spine phantoms with anatomically correct contours (European Spine Phantom, QRM, Erlangen Germany and CIRS Torso Phantom, CIRS, Norfolk, Virginia). Weekly quality control measures are obtained by scanning the hydroxyapatite density phantom (Image Analysis, Columbia, KY). Results from these quality control scans are processed and 'red-flag' values are established that initiate corrective action whenever deviation from accepted limits occurs.

The DXA technologist was trained in all scanning and analysis procedures using manufacturer training standards. Quality control procedures include scanning a Hologic spine phantom on each day that participants are scanned. Longitudinal DXA machine performance is also monitored for drift using this phantom. The study physician reviews results from the daily DXA quality control scan at monthly intervals to ensure that the values are within accepted limits as specified by the manufacturer. All DXA analyses

are reviewed by one of the investigators, and are re-analyzed if problems are found. Finally, the study physician reviews the longitudinal assessment of BMD for all participants without being aware of treatment assignment.

Statistical analysis

Analysis of changes in vBMD

The main analysis for the primary study outcome will be to compare the mean changes from baseline in vBMD measured by QCT in the active and sham groups at the end of 2 years. If the variances are statistically equal, the two-group *t*-test (equal variances) [35] will be used, and if they are not equal the two-group Satterthwaite *t*-test [36] will be used instead. Intention-to-treat will be used for treatment effect assessment of the primary outcome. Subjects unable to continue the assigned intervention are asked to have the annual assessments. For participants who are unable to carry out the assessments, the last observation will be carried forward.

Several secondary analyses of the primary outcome will be performed. An assessment of the relation between treatment adherence and the magnitude of the vBMD treatment effect (i.e., the relationship between the number of sessions completed, and cumulative minutes spent on the platform, and bone density treatment effect) will be performed using generalized estimating equations (GEE) models [37]. In addition, an adjustment of the primary study outcome for potentially confounding factors will be made using a GEE model so that the treatment effect is adjusted for other factors, such as age, gender, or physical activity. Finally, an assessment of the change in vBMD within the active group will use random-effects models [38] stratified by treatment group.

We want to emphasize the importance of assessing change within both groups. If a significant difference in the vBMD change between active and sham groups is observed at 24 months, the active group may have improved significantly while the sham group was unchanged or declined, or the sham group may have declined substantially but the active group remained the same. These two very different scenarios have different implications for the intervention. For this reason, we will evaluate the magnitude and significance of the change from baseline to follow-up in the active group.

Longitudinal models, such as GEE [37] or random effects models [38] will be used to analyze the serial vBMD measures of each participant across the 2-year follow-up period. GEE models will be used when we focus on differences in

population-averaged outcomes (i.e., active versus sham group differences at follow-up). Random effects models will be used when the emphasis is on change in a participant's outcome across time (i.e., trend over time) [38–40].

An adaptation of the Lan-DeMets [41] procedure will be used for assessing the primary study outcome for interim looks' at the data. It is based on the work of Pampallona *et al.* [42,43] and allows for flexible interim monitoring while simultaneously preserving both the type-I and type-II errors of the study. Both alpha and beta spending functions will be used; alpha for efficacy and beta for futility [44]. The rate at which the alpha and beta are spent is a function of the total outcome available at the time of the interim analysis. A stopping boundary that preserves the spirit of the O'Brien–Fleming stopping boundary [45] will be used. The software package East [46] will be used for the monitoring of the primary study outcome. No interim looks at the data have been requested by the DSMB to date.

Analysis of biochemical markers

Biochemical markers will be analyzed using approaches similar to those described above for vBMD. The main analysis for bone formation and resorption markers will compare the mean change from baseline in each biochemical marker in the active and sham groups using a two-group *t*-test (using Satterthwaite adjustment if the variances are unequal).

Several secondary analyses will be performed, including an assessment of the relation between treatment adherence and the magnitude of the biochemical marker treatment effect (i.e., the relationship between the number of sessions completed, and cumulative minutes spent on the platform, and biochemical marker of bone turnover). Similar to the primary outcome, secondary analyses will also include an assessment of the change in each biochemical marker within the active group and adjustment of the biochemical marker outcomes for potentially confounding factors. Nonlinear associations between vBMD and factors such as cumulative minutes spent on the platform might suggest a dose–response relationship, that is, there may be 'diminishing returns' in vBMD gains after a certain time spent on the platform.

Sample sizes

All sample size and power calculations were computed using the software package nQuery [47].

To determine the sample size for the primary outcome of trabecular vBMD as determined by QCT, we assumed a two-group *t*-test with equal group sizes and an 8% difference in outcomes between the active group and the sham group at the end of 2 years. This estimate was based on a conservative 3% per year increase in vBMD (6% over 2 years) in the active group and a 1% per year decrease in vBMD (2% over 2 years) in the sham group. The 3% per year increase expected in the active group was based on conservative extrapolations from a clinical trial of parathyroid hormone in osteoporosis [48]. In the placebo group, data from a drug study supported our estimated 2-year loss of 2% in a placebo group treated with calcium and vitamin D [49]. With 200 participants (100 active and 100 sham), and an attrition rate of 18%, we expect 82 participants per study group (i.e., a total sample of 164 at the end of the study), giving us 91% power to detect an 8% difference in outcome between the active group and the placebo group (two-sided $\alpha = 0.05$). With 82 participants in the treatment group, we expect 97% power (one-sided $\alpha = 0.05$) to detect a 6% increase of BMD from baseline in the active group to the end of 24 months. The 18% attrition rate was estimated based on our pilot study [24].

Table 3 presents the projected differences for each of the secondary outcomes and the power associated with that difference. The treatment group comparison assumes a two-group *t*-test with equal sample sizes and a two-sided $\alpha = 0.05$. The test of difference from baseline to the 24 months of follow-up assumes a one-group *t*-test of the differences from baseline to 24 months with a one-sided $\alpha = 0.05$.

Lessons learned

Recruitment

The challenges encountered by VIBES investigators during recruitment have led to methodologic and substantive innovations in the original design of the study. Perhaps the most important problem faced was one of recruiting osteopenic seniors not taking osteoporosis medications in an environment where prompt treatment of even moderately low BMD is increasingly common. Investigators found it necessary to recruit more than three times the number of ILCs initially expected instead of the planned six ILCs, with minimal increase in the number of VIBES research staff. Some efficiencies of scale were built into the original study design since participation in the intervention took place within the ILCs. The VIBES team obtained informed

Table 3 Projected differences and power for statistical tests of secondary outcomes in the VIBES randomized, sham-controlled trial of low magnitude mechanical stimulation

Outcomes	Treatment group comparison (%)		Difference from baseline to 2nd year of follow-up in treatment group (%)	
	Difference	Power	Difference	Power
Biochemical markers of bone turnover				
CTX	20	84	20	90
BSAP	20	90	20	95
P1NP	20	95	20	98
Musculoskeletal health				
Quadriceps isometric strength	15	84	10	86
Falls	15	88	10	90
Balance				
Mean velocity	15	91	10	92
Root mean square distance	15	88	10	92
Mean frequency	10	88	7.5	83
Centroid frequency	10	86	7.5	82

consent and screened all residents within a short time frame and within a few weeks were ready to screen at the next community. Initial recruitment sessions were held in a group setting in each ILC. Small groups from each ILC traveled to the imaging center for baseline DXA and CT scans, greatly reducing transportation costs. In some cases, transportation has been supplied by the ILC using their own vehicles and drivers. Screening phlebotomy was also performed on-site, which was convenient for participants. Because many ILCs had medical services, ILC medical staff sometimes helped with difficult phlebotomies. Finally, many fewer platforms were required because ILC residents shared devices placed in common areas.

As recruitment progressed and enrollment numbers lagged behind expectations, a number of strategies were developed to further increase screening yields. Given a choice of ILCs, VIBES targeted larger communities in the Boston area in an effort to enroll the largest number of participants. Pre-screening participant information, including medications, was collected at introductory information sessions in order to eliminate residents taking exclusionary medications as early in the screening process as possible. Investigators also learned to make more effective use of ILC personnel and publicity distribution systems during recruitment, including the production of short TV programs and *infomercials* for airing on ILC broadcast systems. Body mass index eligibility criteria and a 2-week adherence run-in period were dropped within several months of study initiation. Investigators also re-screened residents of platform-equipped ILCs in search of eligible new residents or residents who had not been screened initially. These changes greatly improved the efficiency of VIBES recruitment so that a relatively small VIBES staff working

at 21 ILCs were able to enroll and randomize 174 participants as of May 2009 with minimal increase in resources. Sharing baseline BMD results with residents also served as an incentive to participate.

Technical development and electronic data collection

Other lessons that were learned were about the use of the vibrating platforms and the electronic recording and downloading of adherence data. To our knowledge, these methods have no precedent in previous studies. Significant effort was contributed by the device manufacturer in engineering platforms with data collection equipment that permits files to be downloaded to a laptop computer. Platforms are either active or sham and are programmed to operate only for trial participants using an RFID card linked to their treatment group. RFID cards are assigned to each eligible subject at randomization by the Data Coordinating Center. Cards contain ID information and were attached to lanyards to minimize loss. Platforms also were equipped with electronic read-outs providing participants with cues on device status. When the electricity failed or the plug was removed from the electric wall outlet, the equipment had to be reset manually. This problem was discovered early and participants were trained to reset the equipment to ensure optimal functioning. Other problems occurred when software errors in the electronic adherence programs of some platforms led to inaccurate recording of session clock times. Regular monitoring of adherence data detected an unusual number of duplicate sessions on the same day accompanied by an unusual number of

missing sessions. These errors were rectified and the lesson learned was that these type of data requires careful monitoring.

Special technical problems were encountered during the engineering of the sham platforms to make them difficult to distinguish from active vibrating platforms. They were designed to emit auditory cues, identical to those emitted by active platforms. Platform read-outs on session status were identical in both treatment groups.

Problems emerged related to the greater number of platforms required to equip the greater number of ILCs enrolled. Investigators developed methods of re-cycling unused platforms in ILCs with small numbers of randomized participants. Unused platforms are sent back to the manufacturer for re-programming and re-numbering and are reused in another community. Devices that malfunction are repaired and similarly re-cycled.

The groupings of platforms in ILC common areas allowed multiple users per machine without threatening the blinding of participants. Because platforms are not easily moved, they were placed in a central location affording access to all participants of the ILC; however, placement nearby or in the same building was not always possible due to space considerations, the need for an electrical outlet, and safety of others who might be near the equipment. The administrative staff of one ILC designated a single area for participants from two adjacent buildings rather than using each building as a site. As a result, potentially eligible participants withdrew during the screening process. In another ILC, the distance between the platforms and one of the subject's apartments was too far for him to walk. Thus flexibility in the use of shared equipment may be the best approach when resources permit.

Outcome measures – sensitivity to change

As previously mentioned, the VIBES trial was initially designed to be a multicenter clinical trial with each center enrolling its own ILCs. However, when peer reviewers recommended a single recruitment center with a smaller sample, we changed our original primary outcome measure from DXA-derived BMD to QCT BMD. This change in the primary outcome afforded greater sensitivity to change such that our power was adequate with a smaller sample size. Using the QCT outcome measure also allowed us to examine specific bone compartment changes. These advantages came at the cost of the additional transportation to the hospital for the CT scans, additional radiation exposure which resulted in a small number of potential participants who declined to participate, the expense and special training of CT

technologists, and the adaptation of image analysis software for longitudinal analyses.

Addition of musculoskeletal health outcomes

Finally, other lessons that were learned during the start-up phase of the trial leading to the addition of important muscle and balance secondary outcomes to the Protocol. The VIBES design, as originally conceived, evaluated a potentially bone active intervention. Endpoints focused on bone density, architecture, and turnover alone. However, investigators quickly recognized that the intervention may affect muscle as well as bone and that use of the QCT technology leads naturally to the measurement of muscle and balance related endpoints at little extra cost. Consideration of the literature on musculoskeletal health and postural stability related to falls in older people provided other reasons for the addition of muscle and postural stability endpoints to VIBES, provided that they could be measured at minimal cost to the study and time burden to participants. Consequently, investigators expanded the study design to include measurements of: (1) muscle mass by QCT in the area of the lower spine from the same scans used to read spinal bone density and architecture; (2) muscle strength in the right leg by a handheld dynamometer done at the time of the QCT measurement; (3) lower extremity function and balance using the Short Physical Performance Battery done at the time of the QCT measurement; and (4) postural stability with a 4-min trial of double-legged stance on a donated force plate done at the time of the QCT measurement. This expansion of scope also had a considerable impact on the level of interest expressed by potential participants. We encountered multiple questions about balance and equilibrium during our recruitment, and many individuals agreed to participate in the hope that they might learn more about their musculoskeletal health and improve their balance. Their interest underscored the appeal of more broadly defined musculoskeletal endpoints in clinical trials of older persons.

Conclusions

VIBES is the largest study of LMMS in participants over the age of 60 years with state of the art QCT outcome measures of bone health, muscle, and balance outcomes. Data analysis will compare QCT with DXA primary outcome measures in the VIBES population. Our results will contribute important information on the use of QCT primary endpoints and on whether the use of LMMS for 10 min daily will reduce falls and fracture risk factors in older persons. It includes participants of both genders over the age of 80, who are most likely to fall and

who are concerned about osteoporosis, balance, and polypharmacy effects. The VIBES investigators have solved many of the practical problems associated with measurements made on independent-living residents of retirement communities, and the use of active and sham vibrating platforms. The trial also has eliminated the need for using costly per-person intervention equipment by developing flexible methods for centralized intervention and electronic adherence data collection.

We expect that future designers of bone health studies in older participants will follow the same logical progression to a less bone-centered, broader picture of musculoskeletal health, and encounter many of the same problems as VIBES. Our experience should provide a valuable guide to future efforts.

Acknowledgments

We are grateful for the generous assistance of the platform manufacturer for the donation and maintenance of equipment; to our field team, Judith Stone, Dawn Dewkett, Maliaka Lindsey, Erin Kivell, Brianne Jeffrey, Danette Carroll and to the members of our Data and Safety Monitoring Committee, Dennis Black, PhD of the University of California San Francisco, Lawrence Raisz, MD of the University of Connecticut Health Center, Gail Greendale, MD of the University of California Los Angeles, Susan Greenspan, MD of the University of Pittsburgh, and Ronald Zernicke, PhD, DSc of the University of Michigan, for their valuable guidance and encouragement. This work was supported by the National Institutes of Health grant R01AG025489. The VIBES trial was registered at clinicaltrials.gov in September 2006 (NCT00396994). Dr. Kiel has received grant support from Merck, Novartis, Pfizer, Amgen, Wyeth, Hologic; he has served as a consultant for Merck, Novartis, Amgen, Wyeth, GSK, Procter & Gamble, and has served as a speaker for Merck and Novartis. Dr Shane has received grant support from Merck, Lilly, Novartis. Dr. Rubin is a founder of Marodyne Medical. Dr Magaziner has received grant support from Novartis, Merck, and Eli Lilly and has served as a consultant for Amgen, Eli Lilly, Novartis. Dr Zimmerman has received grant support from Merck, Ortho-McNeil Janssen.

References

1. NIH (ed). Osteoporosis prevention, diagnosis and therapy. In: *NIH Consensus Development Conference*, Natcher Auditorium, NIH, Washington, DC, 2000.
2. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002; **288**: 321–33.
3. Siris ES, Harris ST, Rosen CJ, et al. Adherence to bisphosphonate therapy and fracture rates in osteoporotic women: Relationship to vertebral and nonvertebral fractures from 2 US claims databases. *Mayo Clin Proc* 2006; **81**: 1013–22.
4. Warden SJ, Fuchs RK, Castillo AB, et al. Exercise when young provides lifelong benefits to bone structure and strength. *J Bone Miner Res* 2007; **22**: 251–9.
5. Rosenberg IH. Sarcopenia: Origins and clinical relevance. *J Nutr* 1997; **127**: 990S–91S.
6. Burr DB, Forwood MR, Fyhrrie DP, et al. Bone micro-damage and skeletal fragility in osteoporotic and stress fractures. *J Bone Miner Res* 1997; **12**: 6–15.
7. Frost HM. Skeletal structural adaptations to mechanical usage (SATMU): 1. Redefining Wolff's law: The bone modeling problem. *Anat Rec* 1990; **226**: 403–13.
8. Gilsanz V, Wren TA, Sanchez M, et al. Low-level, high-frequency mechanical signals enhance musculoskeletal development of young women with low BMD. *J Bone Miner Res* 2006; **21**: 1464–74.
9. Jenkins KR, Pienta AM, Horgas AL. Activity and health-related quality of life in continuing care retirement communities. *Res Aging* 2002; **24**: 124–30.
10. Wendel VI, Durso SC, Remsburg RE. Studying coronary artery disease in the retirement community. *Nurse Pract* 2001; **26**: 48–55.
11. Blessed G, Tomlinson BE, Roth M. The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. *Br J Psychiatry* 1968; **114**: 797–811.
12. Bouxsein ML. Technology insight: Noninvasive assessment of bone strength in osteoporosis. *Nat Clin Pract Rheumatol* 2008; **4**: 310–18.
13. Black DM, Greenspan SL, Ensrud KE, et al. The effects of parathyroid hormone and alendronate alone or in combination in postmenopausal osteoporosis. *N Engl J Med* 2003; **349**: 1207–15.
14. McClung MR, San Martin J, Miller PD, et al. Opposite bone remodeling effects of teriparatide and alendronate in increasing bone mass. *Arch Intern Med* 2005; **165**: 1762–8.
15. Rubin C, Turner AS, Mallinckrodt C, et al. Mechanical strain, induced noninvasively in the high-frequency domain, is anabolic to cancellous bone, but not cortical bone. *Bone* 2002; **30**: 445–52.
16. Lang T, LeBlanc A, Evans H, et al. Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. *J Bone Miner Res* 2004; **19**: 1006–12.
17. Lang TF, Leblanc AD, Evans HJ, et al. Adaptation of the proximal femur to skeletal reloading after long-duration spaceflight. *J Bone Miner Res* 2006; **21**: 1224–30.
18. Marshall LM, Zmuda JM, Chan BK, et al. Race and ethnic variation in proximal femur structure and BMD among older men. *J Bone Miner Res* 2008; **23**: 121–30.
19. Li W, Sode M, Saeed I, et al. Automated registration of hip and spine for longitudinal QCT studies: Integration with 3D densitometric and structural analysis. *Bone* 2006; **38**: 273–9.
20. Canalis E, Giustina A, Bilezikian JP. Mechanisms of anabolic therapies for osteoporosis. *N Engl J Med* 2007; **357**: 905–16.
21. Bilezikian JP. Combination anabolic and antiresorptive therapy for osteoporosis: Opening the anabolic window. *Curr Osteoporos Rep* 2008; **6**: 24–30.
22. Leveille SG, Kiel DP, Jones RN, et al. The MOBILIZE Boston Study: Design and methods of a prospective cohort study of novel risk factors for falls in an older population. *BMC Geriatr* 2008; **8**: 16–23.

23. Ward K, Alsop C, Caulton J, et al. Low magnitude mechanical loading is osteogenic in children with disabbling conditions. *J Bone Miner Res* 2004; **19**: 360–9.
24. Hannan MT, Cheng DM, Green E, et al. Establishing the compliance in elderly women for use of a low level mechanical stress device in a clinical osteoporosis study. *Osteoporos Int* 2004; **15**: 918–26.
25. Goel VK, Park H, Kong W. Investigation of vibration characteristics of the ligamentous lumbar spine using the finite element approach. *J Biomech Eng* 1994; **116**: 377–83.
26. Magnusson ML, Pope MH, Wilder DG, et al. Are occupational drivers at an increased risk for developing musculoskeletal disorders? *Spine* 1996; **21**: 710–17.
27. Dandanell R, Engstrom K. Vibration from riveting tools in the frequency range 6 Hz-10 MHz and Raynaud's phenomenon. *Scand J Work Environ Health* 1986; **12**: 338–42.
28. Bernard B, Nelson N, Estill CF, et al. The NIOSH review of hand-arm vibration syndrome: Vigilance is crucial. National Institute of Occupational Safety and Health. *J Occup Environ Med* 1998; **40**: 780–5.
29. Collins MW, Lovell MR, Iverson GL, et al. Cumulative effects of concussion in high school athletes. *Neurosurgery* 2002; **51**: 1175–9, discussion 80–81.
30. Bongers PM, Boshuizen HC, Hulshof CT, et al. Long-term sickness absence due to back disorders in crane operators exposed to whole-body vibration. *Int Arch Occup Environ Health* 1988; **61**: 59–64.
31. Griffin MJ. Predicting the hazards of whole-body vibration—considerations of a standard. *Ind Health* 1998; **36**: 83–91.
32. Judex S, Boyd S, Qin YX, et al. Adaptations of trabecular bone to low magnitude vibrations result in more uniform stress and strain under load. *Ann Biomed Eng* 2003; **31**: 12–20.
33. Rubin CT, Lanyon LE. Limb mechanics as a function of speed and gait: A study of functional strains in the radius and tibia of horse and dog. *J Exp Biol* 1982; **101**: 187–211.
34. ISO. *Evaluation of Human Exposure to Whole-Body Vibration*. International Standards Organization, Geneva, 1985 (Report No.: ISO 2631/1).
35. O'Brien R, Muller K. *Applied Analysis of Variance in Behavior Sciences*. Marcel Dekker, New York, 1993.
36. Moser B, Stevens G, Watts C. The two-sample test versus Satterthwaite's approximate F test. *Commun Statist Theory Meth* 1989; **18**: 3963–75.
37. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**: 13–22.
38. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–74.
39. Hu F, Goldberg J, Hedeker D, et al. Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *Am J Epidemiol* 1998; **147**: 694–703.
40. Park T. A comparison of the generalized estimating equation approach with the maximum likelihood approach for repeated measurements. *Stat Med* 1993; **12**: 1723–32.
41. Lan K, DeMets D. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70**: 659–63.
42. Pampallona S, Tsiatis A, Kim K. Spending functions for type I and type II error probabilities of group sequential trials. Technical Report, Boston, Harvard School of Public Health, MA, 1995.
43. Pampallona S, Tsiatis A, Kim K. Interim monitoring of group sequential trials using spending functions for type I and type II error probabilities. *Drug Inf J* 2001; **35**: 1113–21.
44. Pampallona S, Tsiatis A, Kim K. Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *J Stat Plan Inf* 1994; **42**: 19–35.
45. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**: 549–56.
46. Cytel Software Corporation. East (5 edn). Cytel Software Corporation, Cambridge, MA, 2003.
47. Statistical Solutions. nQuery (5 edn). Statistical Solutions, Saugus, MA, 2002.
48. Black DM, Bilezikian JP, Ensrud KE, et al. One year of alendronate after one year of parathyroid hormone (1–84) for osteoporosis. *N Engl J Med* 2005; **353**: 555–65.
49. Genant HK, Lang T, Fuerst T, et al. Treatment with raloxifene for 2 years increases vertebral bone mineral density as measured by volumetric quantitative computed tomography. *Bone* 2004; **35**: 1164–8.