

Exercise effects on bone mineral density in older adults: a meta-analysis of randomized controlled trials

Elisa A. Marques · Jorge Mota · Joana Carvalho

Received: 7 May 2011 / Accepted: 24 August 2011 / Published online: 16 September 2011
© American Aging Association 2011

Abstract The purpose of the study was to assess the effects of exercise interventions with different impact loading characteristics on lumbar spine (LS) and femoral neck (FN) bone mineral density (BMD) in older adults. We searched electronic databases and hand searched selected journals up to February 2011 for randomized controlled trials (RCTs) investigating the effects of impact exercise interventions on LS and FN BMD in older adults. Exercise protocols were categorized according to impact loading characteristics. Weighted mean difference (WMD) meta-analyses were undertaken. Heterogeneity amongst trials and publication bias was tested. Random-effects models were applied. Trial quality assessment was also undertaken. Nineteen RCTs, including 1577 subjects, met the inclusion criteria. Twenty-two study group comparisons reported BMD data at the LS. Meta-analysis showed a significant change in BMD at this site (WMD 0.011 g/cm², 95% CI 0.003 to 0.020; $p=0.007$), although results were moderately inconsistent ($I^2=52.2\%$). BMD data at the FN were available from 19 study group comparisons among older adults. Results were inconsistent ($I^2=63.6\%$) in showing a significant positive effect of exercise on BMD at this site (WMD 0.016 g/cm², 95% CI 0.005 to 0.027; $p=0.004$). Combined loading studies of impact activity mixed

with high-magnitude joint reaction force loading through resistance training were effective at LS (WMD 0.016 g/cm², 95% CI 0.002 to 0.036; $p=0.028$), and no inconsistency existed among these trials. Odd-impact protocols were also effective in increasing BMD at LS (WMD 0.039 g/cm², 95% CI 0.002 to 0.075; $p=0.038$) and FN (WMD 0.036 g/cm², 95% CI 0.012 to 0.061; $p=0.004$), although heterogeneity was evident ($I^2=87.5\%$ and $I^2=83.5\%$, respectively). We found consistency among results for low-impact and resistance exercise studies on LS and FN, although non-significant BMD changes were evident amongst these types of protocols at any site and amongst the RCTs that provided a combined loading impact exercise at FN. Funnel plots showed no evidence of publication bias. Trial quality was moderate to high. The findings from our meta-analysis of RCTs support the efficacy of exercise for increasing LS and FN BMD in older adults.

Keywords Systematic review · Meta-analysis · Bone density · Exercise · Aging

Introduction

Aging is linked to a decreased osteoblast activity, increased osteoclast activity and diminished differentiation potential of bone marrow stem cells due to a relative decline in trophic factors (e.g. oestrogen, IGF-1, vitamin D) favoring local expression of molecules such as interleukins and TNF- α (Khosla and Riggs 2005). As

E. A. Marques (✉) · J. Mota · J. Carvalho
Research Centre in Physical Activity, Health and Leisure,
Faculty of Sport, University of Porto,
Rua Dr. Plácido Costa 91,
4200-450 Porto, Portugal
e-mail: emarques@fade.up.pt

a result, the amount of bone tissue is reduced, which consequently motivates bones to become weaker, commonly leading to osteoporosis. This is a common, serious, and disabling condition due to the inherent association with low-energy trauma or fragility fractures. Hip and vertebral fractures have serious complications such as chronic pain, disability, diminished quality of life and premature death (Dhanwal et al. 2011; Johnell and Kanis 2005) and therefore have become a major and growing problem as elderly population is increasing in every geographical region (Dhanwal et al. 2011; Cooper et al. 2011). Like women, older men become susceptible to age-related bone mineral loss, which continues for the remainder of life, and vertebral and hip fracture rates also increase with advancing age.

The number of hip fractures that occur each year in the world has been estimated by Gullberg et al. (1997) to be 1.25 million (338,000 in men and 917,000 in women) in 1990 and is predicted to rise by 310% in men and 240% in women by 2025. Additionally, there are severe economic consequences of fragility fractures, as the combined annual cost has been estimated to be \$20 billion in the USA and €30 billion in the European Union (Cummings and Melton 2002). Thus it is worthwhile to prevent major osteoporotic fractures, namely hip and vertebral fractures, with intervention.

Physical exercise has been advised as a preventive and therapeutic strategy against aging-induced bone weakness (Schwab and Scalapino 2011), although it has been also described that osteogenic responsiveness to mechanical loading declines with age (Lanyon and Skerry 2001). The effects of exercise on bone mass appear to be attributed to the activation of osteocytes, which in turn alter the balance between bone resorption and formation, favoring modeling, if mechanical loading creates strains of sufficient magnitude (Hsieh et al. 2001). The current American College of Sports Medicine (ACSM) position stand on exercise and physical activity for older adults (Chodzko-Zajko et al. 2009) summarizes published research with respect to the known benefits of exercise on bone health mostly based on postmenopausal women (mean age are commonly set between 55 and 65 years), although different results may be expected in elderly subjects. Recommendations include aerobic exercise training such as walking (low-intensity weight bearing activity) and stair climbing/descending, brisk walking, walking with weighted vests or jogging (higher-intensity bone loading activities) which may be effective in counter-

acting age-related declines in bone mineral density (BMD) in postmenopausal women, and high-intensity resistance exercise training as it seems to preserve or improve BMD relative to sedentary controls (Chodzko-Zajko et al. 2009; Nelson et al. 2007). During the past years, meta-analyses of exercise effects on bone mass have focused on premenopausal and postmenopausal women (Martyn-St James and Carroll 2009; Berard et al. 1997; Martyn-St James and Carroll 2008; Kelley et al. 2001). Results confirmed that exercise may have a positive influence on the skeleton, by increasing or maintaining BMD at the loading sites, although the type of exercise, the skeletal loading characteristics of the different impact exercise interventions and the bone site measured lead to contrasting results (Martyn-St James and Carroll 2009; Berard et al. 1997; Martyn-St James and Carroll 2008; Kelley et al. 2001). Despite the importance of bone density in the elderly, according to our knowledge no previous systematic literature reviews and meta-analyses of the efficacy of physical exercises exclusively on older adults have been performed. Thus, we conducted a systematic review and meta-analysis of randomized controlled trials (RCTs) to assess the effects of exercise on lumbar spine and femoral neck BMD in older adults.

Methods

We carried out our systematic review using a pre-specified protocol, devised according to the guidelines of the Cochrane Collaboration (Higgins and Green 2009).

Criteria for considering studies for this review

The inclusion criteria for this study were (1) randomized controlled trials (RCTs), (2) exercise as the only intervention, (3) older adults aged ≥ 60 years or whose mean age is ≥ 65 years, (4) data for one or more of the following variables provided: lumbar spine BMD and femoral neck BMD, (5) studies published in English language journals, (6) comparative control group and (7) exercise intervention lasting a minimum of 16 weeks.

Only information that met the above criteria was included in our analysis. We did not include abstracts and conference papers from national meetings because of the paucity of data provided as well as the inability to obtain complete data from the authors. Studies published in

non-English-language journals were also not included due to potential errors in the translation and interpretation of findings. Intervention exercise or physical activity trials were defined as weight-bearing exercise (meaning a structured, force-generating activity which loads the skeletal regions above the stimuli provided by activities of daily living (MacKelvie et al. 2002; Hind and Burrows 2007)). Weight-bearing exercises can include aerobics, resistance, endurance training, circuit training, jogging, jumping and other modalities that generate impact to the skeleton. Habitual recreational activity without any specific intervention or supervised activity known not to affect bone (sham exercise) was accepted as activity for the control participants.

We excluded uncontrolled trials, cross-sectional and case-control studies and animal investigations. Among RCTs, we excluded studies in which exercise was combined with other interventions or treatments, such as anti-osteoporotic medication and nutritional or hormonal therapies, as both effects could not be separated. We also excluded studies that included participants (even some) already taking/engaged in those treatments/interventions. Interventions involving multiple behaviors, such as diet plus exercise, were not included although the use of supplements of calcium and vitamin D was acceptable, unless equally distributed between study arms in a given trial. Studies of exercise interventions for individuals institutionalized and/or with specific diseases or conditions or with severe physical disabilities or presence of severe frailty were not included. Extremely low-impact exercise programs such as chair aerobics or yoga and exercise training with a training frequency less than 2 days/week were also not included. Since Tai Chi is a weight-bearing exercise, beneficial effects may be expected, which validate its eligibility as a training program.

Outcome measures for this review were defined as BMD (grams per square centimeter) at the lumbar spine and femoral neck measured by radiographic techniques (single photon absorptiometry (SPA), dual photon absorptiometry (DPA) or dual X-ray absorptiometry (DXA)) with standard deviations (SD).

Data were collected at baseline and the most distal data collection point available when reports presented outcomes at different intervals (unless it would severely reduce the sample size) given the higher possibility of more marked results (changes).

For studies that met our inclusion criteria but did not provide appropriate information on changes in

BMD, we personally tried to contact the authors to retrieve such information. Studies were excluded when authors did not respond or the data were no longer available.

Search methods for identification of studies

A computerized literature search of the MEDLINE, PubMed, Academic Search Complete, CINAHL plus, Scopus, Sport Discus and Web of Knowledge databases was conducted from their inception to February 2011 by one reviewer (EAM). The text words, key words and subject headings used were (exercise* OR “physical activity” OR training) AND (“bone density” OR “bone mineral density”). No limits of the search (including language restrictions, human studies or age) were used at this stage. The reference lists of all identified retrieved studies and some review articles were carefully checked aiming to identify potential interesting studies not found in the primary electronic search. In addition, hand searching of key peer-reviewed journals (*Bone*, *Calcified Tissue International*, *Journal of Bone and Mineral Metabolism*, *Journal of Bone and Mineral Research*, *Osteoporosis International*, and *Medicine and Science in Sports and Exercise*) was also performed to identify possible missed RCTs in database searches. Citations were entered into the reference management software EndNote, version X2 (Thomson Reuters, Carlsbad, CA, USA).

Data collection and analysis

Selection of studies

The titles and abstracts of studies identified in the computerized searches (7727 reports) were examined by two authors (EAM and JM) in order to remove obvious irrelevant reports. An over-inclusive policy was adopted at this stage, which implied that in the absence of any information to the contrary, each article was forwarded to the next stage of the screening process. Criteria for inclusion were titles, abstracts, and/or articles that did not meet the exclusion criteria. Full-text of 113 reports was closely screened to identify those studies complying with the eligibility criteria by two authors independently (EAM and JC). In the case of disagreements, decisions were made by joint consensus between the authors.

Multiple publication bias (trials reporting BMD data for the same participants in more than one publication) was examined by analyzing each study to ensure that data from only one of the articles were included to avoid double-counting participants.

Clarification of data and missing results was obtained by correspondence with authors. Blinding of the investigators to the name of the author, institutional affiliation, journal of publication and study results were not performed because it has been shown that these procedures have no clinically or statistically significant effect on results (Berlin 1997).

Data extraction and management

Data extraction was completed using a pilot-tested and revised coding frame to record information on a range of details. The major categories of variables coded included source characteristics (e.g. country, publication year and presence of funding), study design (e.g. number of allocated participants and number of participants followed up, follow-up length, attrition, compliance, exercise supervision, any adjuvant pharmacological or nutritional therapy affecting bone and intent-to-treat analysis), sample characteristics (e.g. gender, age, health and/or functional status), intervention (e.g. type, frequency and duration of the exercise interventions), scanning technique and outcome measures (BMD values with standard deviations). All data were coded and reviewed for accuracy and consistency by the first author.

Assessment of risk of bias in included studies

The methodological quality of the studies was assessed with the Cochrane risk of bias assessment tool (Higgins and Green 2009), which addresses six specific domains, namely sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other potential sources of systematic bias. Each domain includes one or more specific entries in a ‘Risk of bias’ table. Within each entry, the first part of the tool involves the description of what was reported to have happened in the study. The second part of the tool involves assigning a judgment related to the risk of bias for that entry, such that a judgment of ‘Yes’ indicates low risk of bias, ‘No’ indicates high risk of bias and ‘Unclear’ indicates unclear or unknown risk

of bias. In cases of possible disagreement between the reviewers (EAM and JM), after reanalysis of the article a joint decision was made. Particularly, we were concerned in about whether intention-to-treat (ITT) rather than a per-protocol approach was used in analyzing the data of original articles and if attrition and exclusions from the analysis were reported (item 4). Moreover, as the capacity to detect an osteogenic response is closely associated with the dose of stimuli, the sample size and the follow-up length (item 6), these issues were considered as other potential sources of bias, which were not addressed in the other domains in the tool. Thus, based in this empirical evidence of bias, incomplete outcome data and other source of bias were the two domains with higher relevance in the present review. A risk of bias summary graph was generated according to recommendations by the Cochrane guidelines (Higgins and Green 2009).

Continuous data

Given that our outcome (BMD) was reported on a meaningful scale and all studies in the analysis used the same scale (grams per square centimeter), the effect size (ES) was computed directly on the raw (unstandardized) difference in means (D). We combined trials reporting mean final values with trials reporting mean changes in the same meta-analysis. When studies reported both final and change we used the change score to compute D . Thus, D was calculated from studies that used independent groups as the difference between treatment group vs. comparison group final value means or the difference between absolute change-from-baseline values. A positive D reflects more favorable outcome scores for treatment groups than for the comparison group. Due to variable reporting from the data summaries presented in the individual studies, standard deviation was obtained from the standard error (SE) by multiplying SE by the square root of the sample size; confidence intervals for means were also used to calculate standard deviations using

$$SD = \sqrt{N} \times (\text{upper limit} - \text{lower limit}) / 95\% \text{ CI}^*$$

*calculated using a value from a t distribution (with degrees of freedom equal to the group sample size minus 1) as explained with detail in the *Cochrane Reviewers’ Handbook* (Higgins and Green 2009).

Studies with multiple treatment groups

For trials that randomized participants to one of several relevant experimental intervention groups with a common control group (multi-arm studies), each pair-wise comparison was included separately, although with the *shared* control group divided into two groups with a smaller sample size. Nevertheless, the means and standard deviations were preserved unchanged. This process partially overcomes the unit-of-analysis error and ensures that control participants are not counted more than once within the meta-analysis, representing a practical method of performing investigation of heterogeneity (Higgins and Green 2009).

Dealing with missing data

Where final values were not available from the original article or author, post-means were extracted from mean percent change, and SD were imputed using the pre-training SD as it is reasonable to assume that the intervention does not alter the variability of the outcome measure (Higgins and Green 2009). Reported results regarding missing individual participants were not imputed, and thus we included data for only those participants whose results were known.

Assessment of heterogeneity

In order to express informative heterogeneity indices, a measure of both the magnitude and of uncertainty were presented. Magnitude was represented by both the degree of true variation (the between-studies variance) on the scale of the effect measure (T^2) and the degree of inconsistency (I^2 , *I*-squared, the ratio of true heterogeneity to total observed variation). Uncertainty over whether apparent heterogeneity is genuine was expressed using the *p*-value (<0.10 was considered significant (Higgins and Green 2009) since the *Q* statistic tends to suffer from low differential power) for *Q* statistic (a measure of weighted squared deviations) and using confidence intervals for I^2 .

The *Q* statistic has the advantage of being expressed on a standard scale and is sensitive to the number of studies; T^2 is independent of the number of studies but is expressed on the original metric. Finally, the I^2 statistic is not directly affected by the number of studies in the analysis and expresses the result as a ratio (proportion of the observed variance reflects real

differences in effect size). Generally, values of 25%, 50% and 75% are considered to be indicative of small, moderate and large amounts of inconsistency, respectively (Higgins and Green 2009).

Assessment of reporting biases

It is possible that the studies included in our meta-analysis may overestimate the true effect size as they are based on a biased sample of the target population of studies. Thus, to examine if there is evidence of any bias, we tested the relationship between standard error on the vertical axis and effect size on the horizontal axis (funnel plot) using the Egger's linear regression method (Sterne et al. 2000; Egger et al. 1997).

Another approach to deal with publication bias was applying the trim-and-fill procedure of Duval and Tweedie (Duval and Tweedie 2000). This approach essentially addresses the question, "*What is our best estimate of the unbiased effect size?*" (p. 289) (Borenstein et al. 2009). Trim-and-fill is an iterative non-parametric method used to investigate the number of "missing" studies in a meta-analysis, as indicated by funnel plot asymmetry, and calculates an adjusted pooled estimate with the addition of those "missing" studies (Duval and Tweedie 2000). The probability level of $p < 0.05$ was used to indicate statistical significance.

All the approaches used (funnel plot, the regression test and trim-and-fill) are based on a model which assumes that if the effect size is higher in the smaller studies, then publication bias is the reason. In addition, the procedures to address publication bias are subject to a number of caveats; mostly they tend to have lower power, and trim-and-fill approach can be influenced by one or two aberrant studies (Borenstein et al. 2009).

Data synthesis

Outcomes were analyzed as continuous using a random-effects meta-analysis as studies apparently differ in the mixes of participants and in the implementations of interventions, and thus there may be different effect sizes underlying different studies. We calculated a weighted mean, where the weight assigned to each study is the inverse of that study's variance, and the variance includes the original (within-studies) variance (*V*) plus the estimate of the between-studies variance, T^2 (tau-squared). T^2 was estimated

using the method of moments (or the DerSimonian and Laird). Study weights were assigned with the goal of minimizing both sources of variance, which are more balanced under the random-effects model than under the fixed-effect model. Additionally, large studies are assigned less relative weight, and small studies are assigned more relative weight as compared with the fixed-effect model.

Thus, the computed summary effect is our estimate of the mean of the distribution of all relevant true effects, and the null hypothesis is that the mean of these effects is zero. To establish the statistical significance of our results 95% confidence intervals (CIs) were used. All analyses were conducted with Comprehensive Meta Analysis software version 2.2.048 (Biostat, Englewood, NJ, USA).

Sensitivity analysis

In order to test the impact of alternative decisions or ranges of values for decisions that were arbitrary or unclear during the process of undertaking the present systematic review (which is inevitable), a sensitivity analysis was performed to assess the following decisions: Should analyses be based on change scores and final values simultaneously? Should both genders be combined? What range of dose (training frequency) should be included? Should small sample sizes be included? Should studies which add supplements of calcium and/or vitamin D be combined with no supplementation studies? Can different impact loading exercise protocols be combined in a single summary effect? Exercise protocols were categorized according to the impact classifications described by Nikander et al. (2005) and acceleration forces observed by Vainionpaa et al. (2006), and previously described by Martyn-St James and Carroll (2009). Exercise protocols exclusively based on resistance training were labeled as a new impact group.

Results

Description of studies

Results of the search

The initial database search yielded 18845 possible articles which were screened against the inclusion and exclusion

criteria (Fig. 1). Of these publications, 7727 abstracts were identified without duplicates, and 7614 were rejected at the title and abstract stage. One hundred and thirteen studies were identified for potential inclusion and full-text reports were analyzed, and 19 were entered in the analysis (Brooke-Wavell et al. 1997; Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Korpelainen et al. 2006; Lau et al. 1992; Lord et al. 1996; Marques et al. 2011a, b; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; Taaffe et al. 1999; Villareal et al. 2004; Vincent and Braith 2002; von Stengel et al. 2011a, b; Woo et al. 2007).

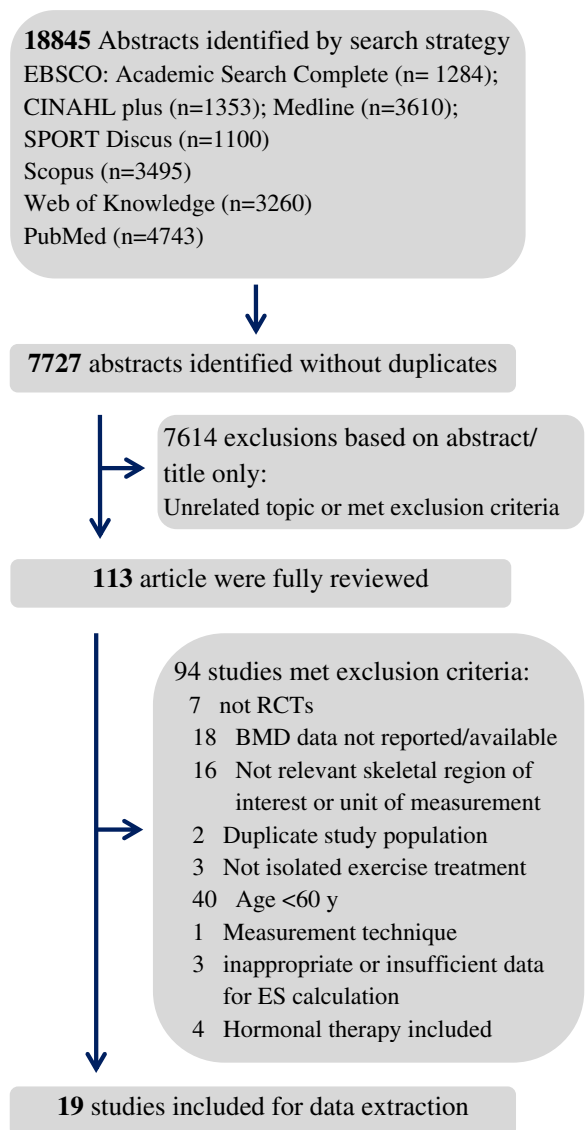


Fig. 1 Flow chart depicting the trial flow for selection of randomized controlled trials (RCTs) to be included

Included studies

Study characteristics are shown in Table 1. Taking the 19 studies together, results were reported from 26 interventions, and the overall sample size was 1577 participants. The sample sizes of the individual studies varied. While small sample sizes had approximately 20 subjects, the large sample sizes ranged between 112 and 227 subjects. The median value for mean age was 69 years, and the mean age ranged from 65 to 83 years. Fifteen studies (Brooke-Wavell et al. 1997; Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Korpelainen et al. 2006; Lau et al. 1992; Lord et al. 1996; Marques et al. 2011a, b; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; von Stengel et al. 2011a, b) were focused exclusively on women ($n=1339$, 85%). Participants were mostly community-dwelling and living independently, excepting one study which included subjects with mild-to-moderate physical frailty (Villareal et al. 2004) and another study that recruited subjects from a hostel for elderly subjects (Lau et al. 1992). One study (Korpelainen et al. 2006) included women with low BMD (hip BMD 2 SD below the reference value). Minority inclusion was infrequently reported: four studies included only Caucasian subjects (Chuin et al. 2009; Jessup et al. 2003; Marques et al. 2011a, b), 3 included Asians (Lau et al. 1992; Park et al. 2008; Woo et al. 2007) and only one study included a small proportion ($\approx 15\%$) of non-Caucasians (Villareal et al. 2004). Three trials (Lau et al. 1992; Park et al. 2008; Woo et al. 2007) were carried out in Asia, while 8 (Brooke-Wavell et al. 1997; Englund et al. 2005; Kemmler et al. 2010; Korpelainen et al. 2006; Marques et al. 2011a, b; von Stengel et al. 2011a, b) were in Europe and 8 (Chuin et al. 2009; Jessup et al. 2003; Lord et al. 1996; Nichols et al. 1995; Rhodes et al. 2000; Taaffe et al. 1999; Villareal et al. 2004; Vincent and Braith 2002) in North America. A few reports appeared before the year of 2000, and 12 studies were published between 2002 and 2011. Attrition was typically small (mean=14.2%), though four studies experienced more pronounced losses (around 25%). Attrition was similar between treatment and control groups. Most studies measured BMD outcomes immediately after completing the intervention, and only one report (Lord et al. 1996) measured the outcomes within 3 weeks after intervention. Six studies (Marques et al. 2011b; Taaffe et al.

1999; Vincent and Braith 2002; von Stengel et al. 2011a, b; Woo et al. 2007) included multiple treatment groups which enabled us to calculate an ES for 12-treatment-group vs. control-group comparisons. For those studies in which dropout information was discriminated, 11 used a per-protocol approach to analyze their data (Brooke-Wavell et al. 1997; Englund et al. 2005; Jessup et al. 2003; Lau et al. 1992; Lord et al. 1996; Nichols et al. 1995; Rhodes et al. 2000; Taaffe et al. 1999; Villareal et al. 2004; Vincent and Braith 2002; Woo et al. 2007), six used the ITT approach (Kemmler et al. 2010; Korpelainen et al. 2006; Marques et al. 2011b; Park et al. 2008; von Stengel et al. 2011a, b) and one study used both approaches for data analysis (Marques et al. 2011a). One study did not report having any dropouts (Chuin et al. 2009).

With respect to intervention characteristics, the most common training modalities were either resistance exercise or a combination of a resistance/strength exercise component with endurance and/or balance exercises. Thus, 15 interventions (62.5%) included a strength or resistance exercise component (Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Lord et al. 1996; Marques et al. 2011a, b; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; Taaffe et al. 1999; Villareal et al. 2004; Vincent and Braith 2002; von Stengel et al. 2011a; Woo et al. 2007). Four studies used a combination of strength, aerobic (odd-impact loading) and balance exercises (Englund et al. 2005; Kemmler et al. 2010; Lord et al. 1996; von Stengel et al. 2011a), two studies combined strength, weight-bearing and balance exercises (Marques et al. 2011a; Park et al. 2008) and two studies used aerobic activities (including walking) as the primary intervention (Brooke-Wavell et al. 1997; Marques et al. 2011b). Other two studies used weight-bearing activities (such as jumping or stepping up and down a block) as the primary training modality (Korpelainen et al. 2006; Lau et al. 1992). One intervention exclusively consisted of Tai Chi sessions (Woo et al. 2007), and another one included only whole body vibration training (von Stengel et al. 2011b). Seven interventions were entirely composed of strength or resistance exercise training (Chuin et al. 2009; Marques et al. 2011b; Nichols et al. 1995; Rhodes et al. 2000; Taaffe et al. 1999; Vincent and Braith 2002; Woo et al. 2007), and one (von Stengel et al. 2011a) incorporated whole body vibration combined with conventional training (including strength,

Table 1 Characteristics of included studies

Study	Country	Subjects ^a	Exercise intervention ^b	Supplementation; supervision	Dropout; compliance	BMD outcomes and device	Comparison group ^c
Brooke-Wavell et al. 1997	UK	78 subjects assigned into an EG ($n=38$, mean age=64.9 y) and a CG ($n=40$, mean age=64.2 y); 100% female	52 wk of training that consisted of self-monitored walking 3.5 times per week for 14.8 min/day for the first 12 wk, followed by 20.4 min/day of walking, 4.8 days/wk, for the subsequent 40 wk	No additional suppl; supervision: no	6%; not reported	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lumar)	Non-exercise control (9 women take up swimming (2 days/wk, 20 min/session) ^c
Chuin et al. 2009	Canada	18 subjects assigned into an EG ($n=11$, mean age=65.4 y) and a CG ($n=7$, mean age=67.4 y); 100% female	24 wk of RT performed 3 days/wk for 60 min/session. Exercise sessions consisted of 15 min warm-up and 3 sets of 8 rep at 80% of 1 RM (45 min), focused on the large and small muscle groups of the upper and lower body	No additional suppl; supervision: yes	0%; >91.7%	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lumar)	Non-exercise placebo
England et al. 2005	Sweden	40 subjects assigned into an EG ($n=21$, mean age=72.8 y) and a CG ($n=19$, mean age=73.2 y); 100% female	52 wk of training performed 2 days/wk for 50 min/session. Exercise sessions consisted of 10 min warm-up, 10 min aerobic, 12 min strength, 5 min balance and coordination, 11 min cool-down/stretching/relaxation	No additional suppl; supervision: yes	18.8%; 67%	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lumar)	Non-exercise control
Jessup et al. 2003	USA	18 subjects assigned into an EG ($n=9$, mean age=69.1 y) and a CG ($n=9$, mean age=69.4 y); 100% female	32 wk of training performed 3 days/wk for 60–90 min/session. Exercise sessions consisted of 5 min warm-up, strength exercises; load-bearing walking and stair climbing for 35–40 min and balance exercises while wearing weighted vests (carry up to 10% of the participant's body weight), 5 min cool-down	1,000/day of calcium and 400 IU/day of vitamin D; supervision: yes	10%; not reported	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Norland)	Non-exercise control
Kemmler et al. 2010	Germany	227 subjects assigned into an EG ($n=115$, mean age=68.9 y) and a CG ($n=112$, mean age=69.2 y); 100% female	18 months of training performed 2 days/wk for 60 min/session and 2 days/wk of home sessions (20 min each). Exercise sessions consisted of warm-up/aerobic dance (20 min at 70%–80% maximum HR), balance training (5 min), functional gymnastics, strength exercise for the upper body and unilateral dynamic weight-bearing leg exercises; home exercises emphasized strength and flexibility exercises	Calcium (1,500 mg/day) and vitamin D 500 IU/day; supervision: yes (group sessions)	7.7%; 76.3% (group sessions) and 42.2% (home sessions)	Lumbar spine (L1–L4) and femoral neck assessed by DXA (Hologic)	Low-intensity wellness program (designed not to cause physical adaptations) ^c
Korpelainen et al. 2006	Finland	160 subjects assigned into an EG ($n=84$, mean age=72.9 y) and a CG ($n=76$, mean age=72.8 y);	30 months of training that included a 60-min supervise group session (1 day/wk) for a 6-month period each year and home sessions (20 min each) usually performed	No additional suppl; supervision: yes (group sessions)	16.9%; 75% (group sessions)	Femoral neck assessed by DXA (Lumar)	Non-exercise control

Lau et al. 1992	China	100% female 23 subjects assigned into an EG ($n=11$, mean age=79 y) and a CG ($n=12$, mean age=75 y); 100% female	3 days/wk. Group sessions included 15 min warm-up and 45 min devoted to jumping and balance exercises including walking, knee bends, leg lifts, heel rises and drops, dancing, stamping, stair climbing and stepping up and down from benches; home exercises were similar to those in the supervised sessions 40 wk of training performed 4 days/wk. Exercise sessions consisted of step up and down a block (9 inches/23 cm) 100 times and 15 min of upper trunk exercises while standing 42 wk of training performed 2 days/wk for 6 min/session. Exercise sessions consisted of 5 min warm-up, 35 min aerobic, balance, coordination and strengthening exercises; 15 min stretching; 5–10 min cool-down period 32 wk of training performed 2 days/wk for 60 min/session. Exercise sessions consisted of 10 min warm-up; 15 min weight-bearing activities, 10 min muscular endurance, 10 min balance training, 10 min agility training, 5 min stretching	No additional suppl.; supervision: yes	16.7%; not reported	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Norland)	Non-exercise control
Lord et al. 1996	Australia	138 subjects assigned into an EG ($n=68$, mean age=71.7 y) and a CG ($n=70$, mean age=71.5 y); 100% female	42 wk of training performed 2 days/wk for 6 min/session. Exercise sessions consisted of 5 min warm-up, 35 min aerobic, balance, coordination and strengthening exercises; 15 min stretching; 5–10 min cool-down period 32 wk of training performed 2 days/wk for 60 min/session. Exercise sessions consisted of 10 min warm-up; 15 min weight-bearing activities, 10 min muscular endurance, 10 min balance training, 10 min agility training, 5 min stretching	No additional suppl.; supervision: yes	22.9%; 72.9%	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lunar)	Non-exercise control
Marques et al. 2011a	Portugal	60 subjects assigned into an EG ($n=30$, mean age=70.1 y) and a CG ($n=30$, mean age=68.2 y); 100% female	32 wk of training performed 2 days/wk for 60 min/session. Exercise sessions consisted of 10 min warm-up; 15 min weight-bearing activities, 10 min muscular endurance, 10 min balance training, 10 min agility training, 5 min stretching	No additional suppl.; supervision: yes	18.3%; 68.3%	Lumbar spine (L1–L4) and femoral neck assessed by DXA (Hologic)	Non-exercise control
Marques et al. 2011b	Portugal 2011	71 subjects assigned into an REG ($n=23$, mean age=67.3 y), an AEG ($n=24$, mean age 70.3 y) and a CG ($n=24$, mean age=67.9 y); 100% female	32 wk of training performed 3 days/wk for 60 min/session. RE sessions consisted of 8–10 min warm-up, 30–40 min specific resistance training which included 2 sets of 6–8 rep at 75–80% of 1 RM, focused on quadriceps, hamstrings, gluteus, trunk, arms and abdominal wall muscle groups; 5–10 min cool-down period. AE sessions 10–15 min warm-up, 35–40 min of dynamic aerobic activities (65–85% heart rate reserve), strength exercises (first 6 weeks), 10 min cool-down period	No additional suppl.; supervision: yes	23.9%; 78.4% (RE) and 77.7%(AE)	Lumbar spine (L1–L4) and femoral neck assessed by DXA (Hologic)	Non-exercise control
Nichols et al. 1995	USA	28 subjects assigned into an EG ($n=14$, mean age=67.8 y) and a CG ($n=14$, mean age=65.2 y); 100% female	24 wk of isotonic weight training performed 3 days/wk. Exercise sessions consisted of 1 set of 10–12 rep at 50% of 1 RM and progressed to 3 sets at 80% of 1 RM	Calcium intake 800 mg/day; supervision: yes	17.6%; 86.8%	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lunar)	Non-exercise control
Park et al. 2008	Japan	50 subjects assigned into an EG ($n=25$, mean age=68.3 y)	48 wk of training performed 3 days/wk for 60 min/session. Exercise sessions consisted of	No additional suppl.; supervision: yes	10%; not reported	Lumbar spine (L2–L4) and femoral neck	Non-exercise control

Table 1 (continued)

Study	Country	Subjects ^a	Exercise intervention ^b	Supplementation; supervision	Dropout; compliance	BMD outcomes and device	Comparison group ^c
Rhodes et al. 2000	Canada	and a CG ($n=25$; mean age=68.4 y); 100% female 44 subjects assigned into an EG ($n=20$, mean age=68.8 y) and a CG ($n=18$, mean age=68.2 y); 100% female	stretching, strength, weight-bearing (at an intensity above 65%–5% of the maximal HR), balance and posture correction training 1 year of RT performed 3/wk for 60 min/session. Exercise sessions consisted of 20 min warm-up and 3 sets of 8 rep at 75% of 1 RM performed in circuit, focused on the large muscle groups of the upper and lower body	No additional suppl.; supervision: yes	13.6%, ≈85%	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Lunar)	Non-exercise control
von Stengel et al. 2011b	Germany	96 subjects assigned into a VVTG ($n=34$, mean age=68.1 y), an RVTG ($n=29$, mean age=67.9 y) and a CG ($n=33$, mean age=67.6 y); 100% female	1 year of whole body vibration training performed 3 days/wk for 15 min/session. Exercise sessions consisted of 7 one-legged or two-legged dynamic leg strengthening exercises, performed on the plates in standing position; VVT G vibrated at a frequency of 35 Hz and the RVTG at 12.5 Hz	Calcium (1,200 mg/day) and vitamin D (800 IU/day); supervision: no ^c	11.1%; 73% (VVTG) and 68% (RVTG)	Lumbar spine (L1–L4) and femoral neck assessed by DXA (Hologic)	Low-intensity wellness program (designed to avoid impact on our primary endpoints) ^d
Taaffe et al. 1999	USA	46 subjects assigned into an EG 1 ($n=12$, mean age=69.4 y), EG 3 ($n=11$, mean age=71.0 y) and a CG ($n=12$, mean age=68.9 y); 36% female	24 wk of RT performed 2 days/wk (EG1) and 3 days/wk (EG2). Exercise sessions consisted of warm-up, 3 sets of 8 rep at 80% of 1 RM, focused on the large muscle groups of the upper and lower body, and cool-down (stretching) period	No additional suppl.; supervision: yes	10.3%, 99% and 97%	Lumbar spine (L2–L4) assessed by DXA (Hologic)	Non-exercise control
Villareal et al. 2004	USA	112 subjects assigned into an EG ($n=65$, mean age=83 y) and a CG ($n=47$, mean age=83 y); 53.6% female	36 wk of training performed 3 days/wk for 90–120 min/session. Exercise sessions consisted of successive phases of physical therapy, resistance (progressed to 3 sets, 8–12 rep at 85–90% of 1 RM) and endurance exercises (progressed to 4×5 min at 85–90% of peak HR)	Calcium and vitamin D to adjust intake to 1,200 mg/d and 800 IU/d; supervision: yes	26.9%, 73.3% (exercise) and 96.7% (control)	Lumbar spine (L2–L4) and femoral neck assessed by DXA (Hologic)	Home-exercise group (focused primarily on flexibility) ^f
Vincent and Braith 2002	USA	62 subjects assigned into an LEXG ($n=24$, mean	24 wk of RT performed 3 days/wk for 30 min/session. Exercise sessions included warm-up, 13	No additional suppl.; supervision: yes	26.2%, >85% (LEX and HEX)	Lumbar spine (L2–L4) and femoral neck	Non-exercise control

von Stengel et al. 2011a	Germany	<p>age=67.6 y), HEXG ($n=22$, mean age=66.6 y) and a CG ($n=16$, mean age=71 y); not reported</p> <p>141 subjects assigned into an EG ($n=47$, mean age=68.6 y), EG plus vibration group (EVG, $n=46$, mean age=68.8 y) and a CG ($n=48$, mean age=68.1 y); 100% female</p>	<p>exercises on resistance machines and cool-down. LEX sessions consisted 1 set of 13 rep at 50% of 1 RM and HEX sessions consisted of 1 set of 8 rep at 75% of 1 RM</p> <p>72 wk of training performed 2 days/wk for 60 min/session and 2 days/wk of home sessions (15–20 min each). Controlled exercise sessions consisted of dancing aerobics, balance training, functional gymnastics and dynamic leg-strength training on vibration plates (without vibration). EVG performed an identical exercise regimen with vibration (25–35 Hz) during the leg-strengthening sequence. Home training session included some strength and stretching exercises of the joint session</p>	<p>Calcium and vitamin D to adjust intake to 1,500 mg/d and 400 IE/day; supervision: yes (joint sessions)</p>	<p>10.6%, 75%/43% (joint sessions/home sessions EG) and 80%/45% (joint sessions/home sessions EVG)</p>	<p>Lumbar spine (L1–L4) assessed by DXA (Hologic)</p> <p>Low-intensity wellness program^a</p>	<p>assessed by DXA (Lunar)</p>
Woo et al. 2007	China	<p>176 subjects assigned into a Tai Chi group ($n=28$ W, mean age=69.7 y/30 M, mean age=68.2 y), a REG ($n=29$ W, mean age=69.6 y/30 M, mean age=68.7 y) and a CG ($n=30$ W, mean age=69.6 y/29 M, mean age=68.1 y); 50% female</p>	<p>52 wk of training performed 3 days/wk. Tai Chi sessions used the Yang style with 24 forms. RE sessions included 6 exercises repeated 30 times using theraband of medium strength</p>	<p>No additional suppl; supervision: no mention</p>	<p>2.2%, 81% (tai chi) and 76.6% (RE)</p>	<p>Lumbar spine (L2–L4) and femoral neck assessed by DXA (Hologic)</p> <p>Non-exercise control</p>	

EG exercise group, CG control group, RT resistance training, rep repetitions, RM one repetition maximum, LEXG low-intensity exercise group, HEXG high-intensity exercise group, REG resistance exercise group, AEG aerobic exercise group, IITG vertical vibration training group, RIT rotational vibration training group, W women, M men, suppl supplementation, dday(s), wk week(s), y year(s)

^a Number of participants included in the analysis (final sample size)

^b Exercise protocol, frequency, intensity and duration

^c Only the first 3 sessions, and every 6 weeks an instructor controlled if the exercises were still being executed properly

⁴ Intervention that would not be expected to improve our primary outcome measures

endurance, balance and functional exercises). The remaining two interventions used a combination of resistance exercise with endurance (Villareal et al. 2004) or with weight-bearing exercises (Jessup et al. 2003). Resistance exercise intensity ranged between 50 and 90% of 1 repetition maximum (1-RM). Aerobic intensity ranged from 65 to 90% of sub-maximum heart rate.

Most studies were center-based: they were delivered in places such as a community center, a training facility or a gymnasium. In three studies (Kemmler et al. 2010; Korpelainen et al. 2006; von Stengel et al. 2011a), participants received interventions both in a group setting at a center and as individuals in their own homes. One intervention consisted of self-monitored walking (Brooke-Wavell et al. 1997). Details regarding supervision of the exercise sessions, apart from home sessions which were not supervised, were not reported in one trial (Woo et al. 2007), and one study reported that all walking sessions were unsupervised (Brooke-Wavell et al. 1997). Compliance with the prescribed exercise (group) interventions measured as a percentage of attended sessions ranged from 67% to 99%, and the median attendance was 77%. Four studies did not report attendance rate (Jessup et al. 2003; Lau et al. 1992; Park et al. 2008). In general interventions were delivered three to four times per week, and exercise sessions lasted at least 60 min. Home sessions tended to be shorter (lasting around 20 min). Session duration was not specified in five of the included interventions (Nichols et al. 1995; Taaffe et al. 1999; Woo et al. 2007). The length from baseline to post-intervention test ranged from 6 (24 weeks) to 30 months, whereas 13 interventions (Brooke-Wavell et al. 1997; Chuin et al. 2009; Kemmler et al. 2010; Korpelainen et al. 2006; Lau et al. 1992; Lord et al. 1996; Park et al. 2008; Rhodes et al. 2000; von Stengel et al. 2011a, b; Woo et al. 2007) lasted 40 weeks or more. Five interventions from five studies (Jessup et al. 2003; Kemmler et al. 2010; Villareal et al. 2004; von Stengel et al. 2011a, b) increased daily calcium and vitamin D intake levels of all participants (to adjust calcium intake up to 1000–1500 mg and vitamin D up to 400–800 IU) by means of supplementation during the intervention. One study (Nichols et al. 1995) only increased daily calcium intake level to 800 mg via dairy products or calcium supplement.

BMD outcomes were measured before and after intervention in each individual study by the same

radiographic techniques (DXA), but measurements were taken using DXA devices from three different manufacturers (Norland, Cooper Surgical, Trumbull, CN, USA; Lunar, GE Medical Systems, Madison, WI, USA and Hologic Bedford, MA, USA). Four trials reported ongoing BMD assessment at more than two time-points (Korpelainen et al. 2006; Nichols et al. 1995; Villareal et al. 2004; Woo et al. 2007). BMD at lumbar spine was assessed in 17 trials (Brooke-Wavell et al. 1997; Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Lau et al. 1992; Lord et al. 1996; Marques et al. 2011a; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; Taaffe et al. 1999; Villareal et al. 2004; Vincent and Braith 2002; von Stengel et al. 2011a, b; Woo et al. 2007), and femoral neck BMD was also assessed in 16 trials (Brooke-Wavell et al. 1997; Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Korpelainen et al. 2006; Lau et al. 1992; Lord et al. 1996; Marques et al. 2011a, b; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; von Stengel et al. 2010b; Villareal et al. 2004; Vincent and Braith 2002); thus 14 trials included both bone site measures (Brooke-Wavell et al. 1997; Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Kemmler et al. 2010; Lau et al. 1992; Lord et al. 1996; Marques et al. 2011a; Nichols et al. 1995; Park et al. 2008; Rhodes et al. 2000; von Stengel et al. 2011b; Villareal et al. 2004; Vincent and Braith 2002). Three trials did not focus on femoral neck (Taaffe et al. 1999; von Stengel et al. 2011a; Woo et al. 2007) and 2 trials on lumbar spine site (Korpelainen et al. 2006; Marques et al. 2011b).

Mean final values in BMD along with SDs were available for nine studies (Chuin et al. 2009; Englund et al. 2005; Jessup et al. 2003; Lord et al. 1996; Marques et al. 2011a, b; Park et al. 2008; Rhodes et al. 2000; Vincent and Braith 2002). Absolute change values in BMD at follow-up along with SDs were available for four studies (Brooke-Wavell et al. 1997; Kemmler et al. 2010; von Stengel et al. 2011a, b). Final means were extracted from mean percent change for two studies (Lau et al. 1992; Woo et al. 2007). SDs were imputed using the pre-training SD for one study (Woo et al. 2007), while the SDs of another study (Lau et al. 1992) were obtained from CI. Two studies (Nichols et al. 1995; Taaffe et al. 1999) reported final means and SE, and one study (Korpelainen et al. 2006) reported final means and

CI; thus SDs were obtained from SE and CI, respectively.

Excluded studies

Ninety-four articles were excluded because of no relevant skeletal region of interest or unit of measurement ($n=16$), age below 60 years ($n=40$), no RCT ($n=7$), duplicate study population ($n=2$), no BMD data reported or available ($n=18$), hormonal therapy included ($n=4$), no isolated exercise treatment ($n=3$), no extractable data ($n=3$) and other measurement technique ($n=1$).

Risk of bias in included studies

Risk of bias is summarized in Fig. 2, which illustrates the methodological quality summary. Judgments ('yes', 'no', 'unclear') for each domain across all the included studies are shown, demonstrating an overall low across-studies potential risk of bias. Thus, most trials had low risk of bias across the six domains. Risk of bias for the domain 'other sources of bias' is 'high' only for the question-based entry 'small sample size', and the entry 'exercise duration' has a moderate risk of bias as 10 trials were sustained for less than 12 months.

Effects of interventions

Data on lumbar spine BMD were available for 22 study group comparisons of exercise interventions vs. control from 17 of the included trials. A total of 769 exercise intervention participants and 571 controls were included. Exercise interventions resulted in a small increase in BMD at this site of 0.011 g/cm^2 (WMD, 95% CI 0.003 to 0.020; $p=0.007$). Heterogeneity was statistically significant in this subset of studies ($Q=44.0$, $df=21$, $p=0.002$); thus studies do not share a common effect size. Our estimate for the SD of the true effect sizes (T) was 0.011. Therefore, considering that our summary effect was 0.011, we expected that some 95% of the true effects will fall in the approximate range of -0.011 to 0.033. In addition, the degree of inconsistency was moderate ($I^2=52\%$, 95% CI 22–71%), which indicates that a considerable amount of the variability across studies was due to heterogeneity rather than chance.

The 16 trials assessing femoral neck BMD provided 19 study group comparisons, including 651 subjects in

the intervention group and 541 subjects in the control group. The combined WMD in BMD was 0.016 g/cm^2 (95% CI 0.005 to 0.027; $p=0.004$). Heterogeneity was statistically significant in this analysis ($Q=49.4$, $df=18$, $p<0.001$). The SD of the summary effect (T) was 0.015; thus most effects will fall in the range of -0.013 to 0.045. Moreover a moderate proportion of the observed variance was real ($I^2=64\%$, 95% CI 40–78%). Figures 3 and 4 show the results from meta-analysis of all included RCTs. Table 2 lists results from meta-analyses and all sensitivity analyses.

Sensitivity analysis including only final values generally did not show any divergence from the heterogeneity evident in the primary analyses including final and change values at lumbar spine and femoral neck ($I^2\approx 55\%$). The WMD in BMD at lumbar spine was 0.016 g/cm^2 (95% CI -0.004 to 0.036; $p=0.111$), and an increase in BMD of 0.020 g/cm^2 (95% CI 0.003 to 0.037; $p=0.023$) was observed at femoral neck. Under the random-effects model, the WMD in BMD at both sites were nearly identical for all trials and trials including only females (0.011 vs. 0.012 for lumbar spine and 0.016 vs. 0.014), and heterogeneity in both sites ($I^2\approx 65\%$) was not divergent from the analysis with all trials. Similarly, sensitivity analyses for the effects of high training frequency (≥ 3 days/week) did not show divergent results from the primary meta-analysis, with similar moderate heterogeneity in both analyses ($I^2\approx 60\%$). No significant effects in BMD were evident at lumbar spine when subgroup analysis excluded trials including calcium and/or vitamin D supplementation. Although a more pronounced overall effect was observed for femoral neck (WMD= 0.022 g/cm^2), the analysis for this subgroup of trials was heterogeneous ($I^2=69\%$). Treatment effects at both lumbar spine and femoral neck were significant and more robust when the analysis only included trials with a total sample of more than 50 participants, but I^2 value for femoral neck analysis increased to 79%.

The subgroup analysis of trials evaluating protocols of combined loading protocols that incorporated impact activity with resistance exercises was consistent in showing positive effects of this type of exercise on BMD at the lumbar spine ($I^2=0\%$). The femoral neck analysis for this type of exercise was heterogeneous ($I^2=62\%$), and the observed effect was non-significant. Trivial and non-significant treatment effects at both lumbar spine and femoral neck were

Fig. 2 Risk of bias summary

	Adequate sequence generation?	Allocation concealment?	Blinding?	Incomplete outcome data addressed?	Free of selective reporting?	Free of other bias (small sample size)?	Free of other bias (reduced training frequency)?	Free of other bias (reduced exercise duration)?
Brooke-Wavell et al. 1997	?	?	+	+	+	-	+	+
Chuin et al. 2009	?	?	+	+	+	-	+	-
Englund et al. 2005	?	?	+	+	+	-	-	+
Jessup et al. 2003	+	?	+	+	-	-	+	-
Kemmler et al. 2010	+	?	+	+	+	+	+	+
Korpelainen et al. 2006	+	+	+	+	+	+	+	+
Lau et al. 1992	+	?	+	+	?	-	+	-
Lord et al. 1996	+	?	+	+	+	+	-	+
Marques et al. 2011a	+	+	+	+	+	-	-	-
Marques et al. 2011b	+	+	+	+	+	-	+	-
Nichols et al. 1995	?	?	+	?	+	-	+	-
Park et al. 2008	+	?	+	+	+	-	+	-
Rhodes et al. 2000	?	?	+	+	+	-	+	+
Taaffe et al. 1999	?	?	+	+	+	-	?	-
Villareal et al. 2004	?	?	+	+	+	+	+	-
Vincent and Braith 2002	+	?	+	-	+	-	+	-
von Stengel et al. 2011a	+	?	+	+	+	+	+	+
von Stengel et al. 2011b	+	?	+	+	+	-	+	+
Woo et al. 2007	+	?	+	?	+	+	+	+

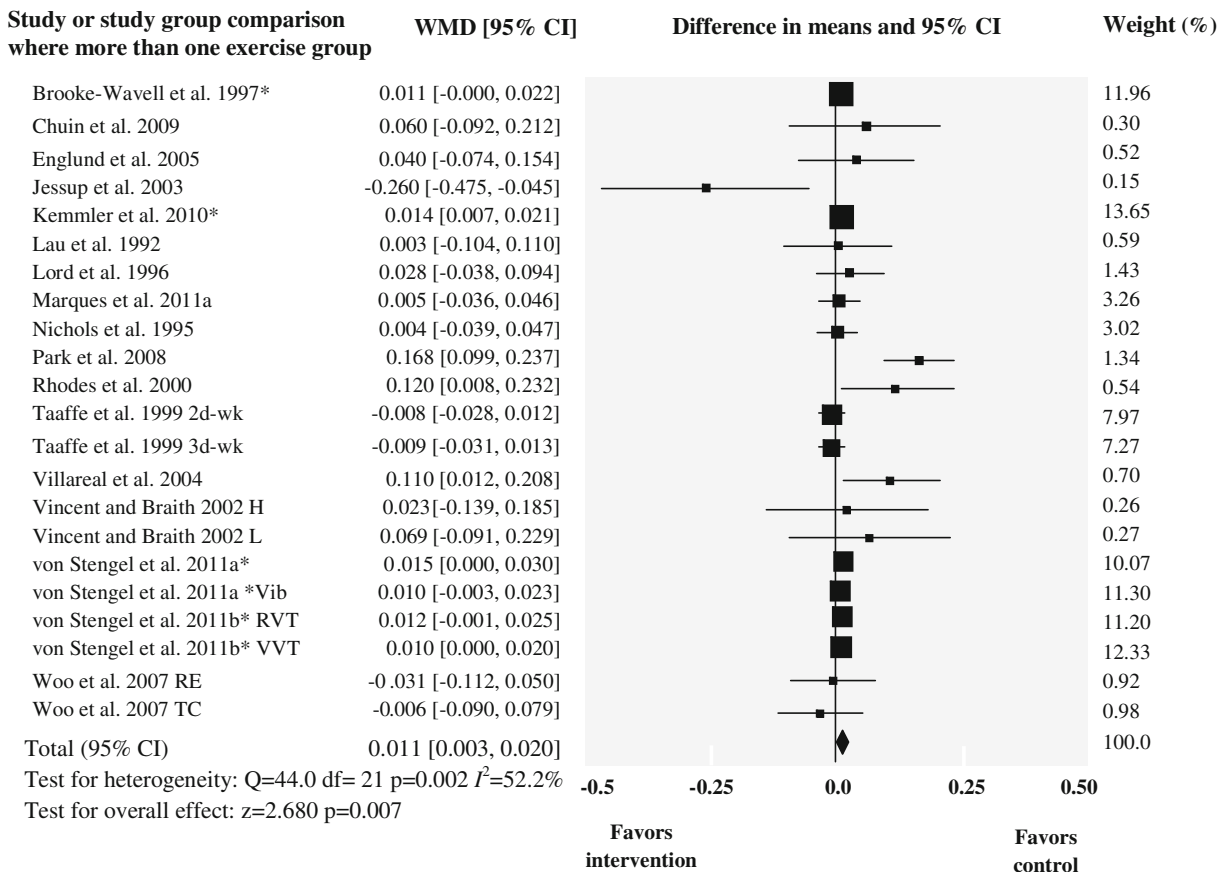


Fig. 3 Forest plot for lumbar spine BMD. *Diamonds* represent overall weighted mean difference (WMD, grams per square centimeter) calculated by random-effect model with 95% CI.

also confirmed ($I^2=0\%$) in the subgroup analyses of protocols evaluating low-impact exercises. Sensitivity analyses for the effects of odd-impact protocols that incorporate group exercise classes were heterogeneous in having a positive effect at the lumbar spine ($I^2=87.5\%$) and femoral neck ($I^2=83.5\%$). An increase in BMD of 0.039 g/cm^2 (95% CI 0.002 to 0.075; $p=0.038$) and 0.036 g/cm^2 (95% CI 0.012 to 0.061; $p=0.004$) was observed at lumbar spine and femoral neck, respectively. Finally, the subgroup analyses of trials evaluating protocols of resistance exercise training were consistent in showing no positive and non-significant effects on BMD at lumbar spine ($I^2=11.8\%$) and a positive although also non-significant at femoral neck ($I^2=22.3\%$). Notably, using a random-effects model Q between groups was 3.993 ($df=3$, $p=0.262$) in lumbar spine subgroup analysis and Q between groups was 4.322 ($df=3$, $p=0.229$) in femoral neck subgroup analysis, which corroborate that treatment effect does

*Absolute change values. *RVT* rotational vibration training, *VVT* vertical vibration training, *d-wk* days-week, *H* high, *L* low, *Vib* vibration, *RE* resistance exercise, *TC* Tai Chi

not differ among protocols with variations in the skeletal loading characteristics for both bone sites.

Funnel plots for lumbar spine and femoral neck meta-analyses are presented in Figs. 5 and 6, respectively. Visual inspection of these plots seems to indicate no evidence of asymmetry of trials. In the numerical tests, the results of Egger's test showed no evidence of publication bias for lumbar spine (1-tailed $p=0.248$). Moreover, trim-and-fill method (random-effects model) suggested that two studies were missing to the left side of the mean effect (black circles in Fig. 5); thus the imputed point estimate would reduce the magnitude of the pooled effect size to 0.009 g/cm^2 (difference in means, 95% CI -0.002 to 0.019). Results for femoral neck also showed no evidence of publication bias. Similarly, Egger's test result was not significant (1-tailed $p=0.197$), while trim-and-fill method suggested that one study was missing to the right side of the mean effect (black circles in Fig. 6), which

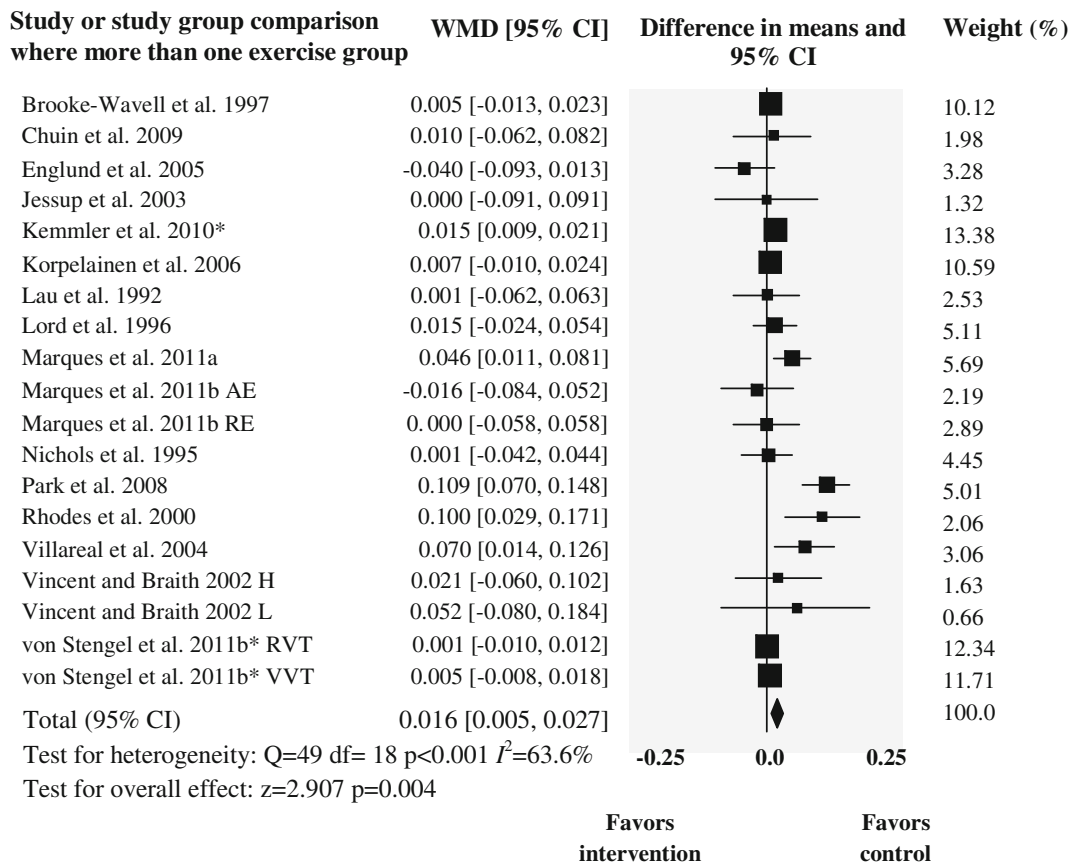


Fig. 4 Forest plot for femoral neck BMD. *Diamonds* represent overall weighted mean difference (WMD, grams per square centimeter) calculated by random-effect model with 95% CI.

*Absolute change values. *RE* resistance exercise, *AE* aerobic exercise, *RVT* rotational vibration training, *VVT* vertical vibration training, *H* high, *L* low

increases the magnitude of the pooled point estimate to 0.019 g/cm² (difference in means, 95% CI 0.007 to 0.030).

Discussion

Summary of main findings

After many decades of research investigating the association between bone biology, aging and outcomes of exercise interventions, this systematic review and meta-analysis of 19 studies supports the view that exercise of mixed loading impact is associated with significant increases in BMD mean values of 0.011 g/cm² (95% CI 0.003 to 0.020 g/cm², $p=0.007$) for the lumbar spine and 0.016 g/cm² (95% CI 0.005 to 0.027 g/cm², $p=0.004$) for femoral neck in older adults.

Despite these encouraging results, not all impact exercise protocols appear effective in reducing bone loss (results presented in Table 2). Combined loading studies of impact activity mixed with high-magnitude joint reaction force loading through resistance training were effective at lumbar spine (WMD 0.016 g/cm², 95% CI 0.002 to 0.036; $p=0.028$), and no inconsistency existed among these trials. Odd-impact protocols were also effective in increasing BMD at lumbar spine (WMD 0.039 g/cm², 95% CI 0.002 to 0.075; $p=0.038$) and femoral neck (WMD 0.036 g/cm², 95% CI 0.012 to 0.061; $p=0.004$), and changes were larger, although heterogeneity was evident ($I^2=87.5\%$ and $I^2=83.5\%$, respectively). The observed high I^2 may be related with the small sample sizes of two studies (Jessup et al. 2003; Lau et al. 1992) which also limited the statistical power to detect significant differences. Moreover, those studies did not report the attendance rate, which may also affect the post-training results. We found

Table 2 Summary of meta-analyses and sensitivity analyses (random-effects) by region of interest

Analysis	No. study group comparisons	Sample size: EG/CG	Heterogeneity (Q , p -value)	Inconsistency (I^2)	WMD, 95% CI	Test of null (Z value, p -value)
All included trials						
Lumbar spine	22	769/571	$Q=44.0$, $p=0.002$	$I^2=52.2$	0.011 g/cm ² , 0.003 to 0.020	$Z=2.630$, $p=0.007$
Femoral neck	19	651/541	$Q=49.4$, $p<0.001$	$I^2=63.6$	0.016 g/cm ² , 0.005 to 0.027	$Z=2.907$, $p=0.004$
Final means						
Lumbar spine	22	769/571	$Q=42.5$, $p=0.004$	$I^2=50.6\%$	0.016 g/cm ² , -0.004 to 0.036	$Z=1.593$, $p=0.111$
Femoral neck	19	651/541	$Q=43.2$, $p=0.001$	$I^2=58.3\%$	0.020 g/cm ² , 0.003 to 0.037	$Z=2.276$, $p=0.023$
Only females						
Lumbar spine	17	575/467	$Q=44.3$, $p<0.001$	$I^2=63.9\%$	0.012 g/cm ² , 0.002 to 0.022	$Z=2.259$, $p=0.024$
Femoral neck	16	540/478	$Q=44.7$, $p<0.001$	$I^2=66.5\%$	0.014 g/cm ² , 0.003 to 0.025	$Z=2.485$, $p=0.013$
Sessions ≥ 3 dwk						
Lumbar spine	18	638/446	$Q=39.8$, $p=0.001$	$I^2=57.3\%$	0.013 g/cm ² , 0.004 to 0.022	$Z=2.795$, $p=0.005$
Femoral neck	16	532/422	$Q=42.1$, $p<0.001$	$I^2=64.4\%$	0.016 g/cm ² , 0.005 to 0.028	$Z=2.762$, $p=0.006$
No supplements						
Lumbar spine	14	410/308	$Q=32.3$, $p=0.002$	$I^2=59.7\%$	0.018 g/cm ² , -0.003 to 0.039	$Z=1.671$, $p=0.095$
Femoral neck	13	385/326	$Q=38.2$, $p<0.001$	$I^2=68.6\%$	0.022 g/cm ² , 0.001 to 0.044	$Z=2.063$, $p=0.039$
Total sample size ≥ 50						
Lumbar spine	13	631/463	$Q=26.3$, $p=0.010$	$I^2=54.3\%$	0.015/cm ² , 0.006 to 0.025	$Z=3.116$, $p=0.002$
Femoral neck	8	443/405	$Q=33.0$, $p<0.001$	$I^2=78.8\%$	0.024 g/cm ² , 0.009 to 0.039	$Z=3.164$, $p=0.002$
Combined loading						
Lumbar spine	6	362/296	$Q=4.5$, $p=0.482$	$I^2=0\%$	0.016 g/cm ² , 0.002 to 0.030	$Z=2.203$, $p=0.028$
Femoral neck	4	269/248	$Q=7.9$, $p=0.048$	$I^2=62.0\%$	0.014 g/cm ² , -0.011 to 0.040	$Z=1.109$, $p=0.268$
Low-impact						
Lumbar spine	5	218/132	$Q=1.2$, $p=0.876$	$I^2=0\%$	0.009 g/cm ² , -0.020 to 0.024	$Z=1.283$, $p=0.200$
Femoral neck	2	125/85	$Q=0.6$, $p=0.904$	$I^2=0\%$	0.002 g/cm ² , -0.048 to 0.045	$Z=0.182$, $p=0.856$
Odd-impact						
Lumbar spine	4	75/76	$Q=24.0$, $p<0.001$	$I^2=87.5\%$	0.039 g/cm ² , 0.002 to 0.075	$Z=2.075$, $p=0.038$
Femoral neck	4	143/141	$Q=24.3$, $p<0.001$	$I^2=83.5\%$	0.036 g/cm ² , 0.012 to 0.061	$Z=2.883$, $p=0.004$
RE						
Lumbar spine	7	114/67	$Q=6.8$, $p=0.340$	$I^2=11.8\%$	-0.002 g/cm ² , -0.019 to 0.015	$Z=-0.217$, $p=0.828$
Femoral neck	6	114/67	$Q=6.4$, $p=0.266$	$I^2=22.3\%$	0.023 g/cm ² , -0.009 to 0.054	$Z=1.414$, $p=0.157$

EG exercise group, CG control group, RE resistance exercise, dwk days/week, WMD weight mean difference

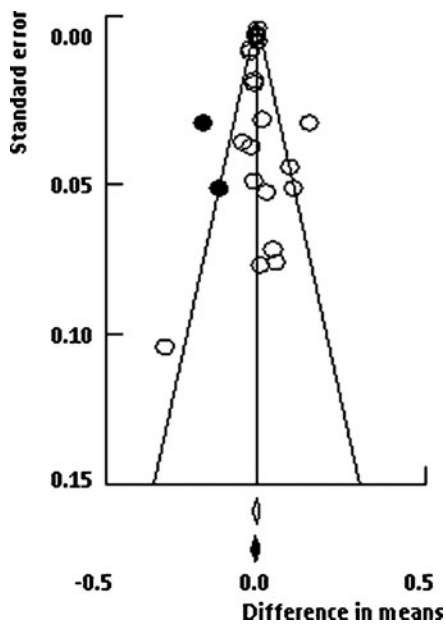


Fig. 5 Funnel plot test exploring publication bias (random-effects model). *Black circles* represent the studies imputed when the trim-and-fill method was applied

consistency among results for low-impact and RE studies on lumbar spine and femoral neck, although non-significant BMD changes were evident amongst these types of protocols at any site and amongst the RCTs that provided a combined loading impact exercise at femoral neck. These findings were less robust and had low power than those for the overall analysis as there

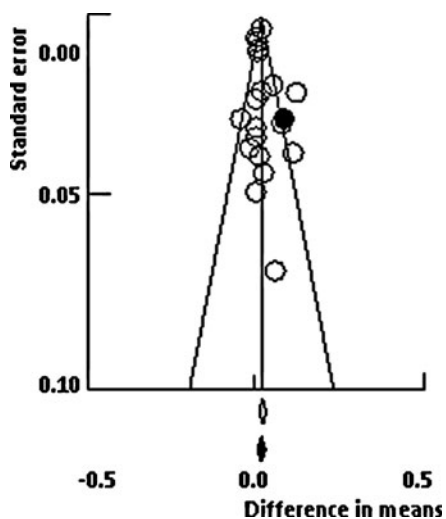


Fig. 6 Funnel plot test exploring publication bias (random-effects model). *Black circle* represents the study imputed when the trim-and-fill method was applied

were fewer studies, which implies that results need to be interpreted carefully. Taken together, subgroup analysis corroborates that treatment effect does not differ among protocols, which lends support to a summary effect (meta-analysis) combining different exercise training with variations in the skeletal loading characteristics for each bone site. However, considering the subgroup analysis our findings indicate that odd-impact loading has the higher potential for preserving BMD at the lumbar spine and femoral neck in older women.

Applicability and quality of the evidence

Our structured systematic searches resulted in 22 RCT study group comparisons evaluating lumbar spine BMD and 19 RCT study group comparisons evaluating femoral neck BMD resulting in an overall sample size of 1340 and 1192 participants, respectively. Though the data from the pooled summaries alone seem to support benefit with exercise protocols, they should be viewed with caution because of the moderate heterogeneity amongst studies for lumbar spine ($I^2=52\%$) and femoral neck ($I^2=64\%$). As systematic reviews bring together studies that are methodologically diverse, heterogeneity in their results is to be expected (Higgins et al. 2002). For example, heterogeneity is likely to arise through diversity in participants' characteristics, doses, lengths of follow-up and study quality. We explored the extent to which heterogeneity affects the conclusions of the meta-analysis through sensitive analyses. Similar magnitude of treatment effects was found in the female-subjects studies, in studies with high training frequency, in studies with no use of supplements and in large-sample size studies. The level of heterogeneity was preserved on all these sensitive analyses. Furthermore, there was considerable variability in the type and dose of exercise prescribed amongst the different intervention trials, all of which may account for the marked variability in the skeletal response to training. Sensitivity analyses were not undertaken on compliance and dropout rates, as values did not diverge extensively.

Other key methodological limitations of the studies, namely aspects of concealment of allocation, should also be considered, as they may limit internal validity. Indeed, only one RCT contained a statement as to whether concealment of allocation had occurred or not, although consolidate standards for reporting of RCTs (CONSORT) are now available to researchers

(Moher et al. 2001), and most of the included trials were undertaken at a time with possible access to these standards. Nevertheless, the lack of reporting these aspects of study quality in our included trials should not be interpreted as that they were not undertaken.

For our meta-analyses, we included RCTs assessing the effect of exercise alone which contributed to the external validity of results, as the additive effects of hormone therapy combined with the exercise may influence the positive findings observed on previous meta-analysis (Martyn-St James and Carroll 2009) and reviews (Lanyon 1996; Kohrt et al. 2004).

The primary outcome for this review was BMD, which is a surrogate marker for fractures (Kanis et al. 2008). The elevated incidence of fractures in older adults is viewed with high concern and may simultaneously act as the driving force for the development and improvement of preventive strategies such as physical activity. However, studies that address the effect of exercise on incidence of fractures are lacking.

Potential biases in the review process

One of the strengths of this review is the comprehensive search strategy that identified a large number of studies from 10 countries. The 19 included trials generally had sound methods and had a low risk of bias, with the main methodological weaknesses being the small sample size (<100 subjects) in 13 trials and reduced duration of the exercise intervention in 10 trials (sustained for less than 12 months). Variation in clinical end-points definition, monitoring and reporting can also be important sources of error. Thus, all included studies measured BMD before and immediately after the exercise intervention, and all assessed the outcome with DXA technology, precluding the effect of variability in inter-instrument reliability. Despite this homogeneity, pencil and fan beam technology was used. Although a wide variety of technologies are available for the assessment of osteoporosis, there is a general view that DXA is the “gold standard” and validly measures a real BMD along with other attracting advantages such as high speed, precision, low radiation exposure and availability of reference data (Watts 2004).

Although the methods in most of the trials failed to blind participants, personnel and outcome assessors, this should not represent a limitation, resulting in biased estimates of treatment effect, because of the

objective nature of the outcome measurement. Indeed, the magnitude of bias associated with inadequate blinding of participants is likely to be greater for more subjective outcomes (Wood et al. 2008; Higgins and Green 2009).

Because the validity of any meta-analytical review can potentially be compromised by heterogeneity in patient characteristics, we predefined inclusion and exclusion criteria that ensured reasonable likeness between subjects (such as age, health and functional status, diseases and use of pharmacological or nutritional therapy affecting bone).

A random-effects model was used based on our understanding that studies do not share a common effect size (the true effect size varies from one study to the next), and not on the outcome of the test of homogeneity as previously observed (Kelley et al. 2001). This model enables to generalize to a range of scenarios, which is in fact the goal of a meta-analysis. Moreover, to illustrate our findings we used mean differences that are clinically relevant and easy to interpret, as the pooled estimate is expressed in the same unit of the measure technique (grams per square centimeter).

Reported compliance with the exercise interventions was high amongst the trials included in the present meta-analysis. No adverse effects associated with the exercise interventions were reported in any trial, and there were a low number of unsupervised exercise trials.

Regarding publication bias, examination of funnel plots revealed symmetry of study effect sizes for both lumbar spine and femoral neck BMD.

Nevertheless, few limitations should be emphasized. Evidence from intervention trials indicates that BMD and geometry adaptations to loading vary by age, skeletal site and sex. However, due to scant data in older men, we could not synthesize the exercise effects for this population. Moreover, most patients studied were women and Caucasian (although men and other ethnic groups were included), and extrapolation to elderly males and other populations should be made with caution.

Agreements and disagreements with other studies or reviews

The association between exercise and bone health has been widely debated (Kohrt et al. 2004), but the parallel influence of the aging process and related constraints

impose other demanding challenges, namely defining an exercise program well tolerated by older adults and that efficiently stimulate bone remodeling.

To our knowledge, the present study is the first systematic review to include a comprehensive analysis of RCTs assessing the effect of exercise interventions with different impact load characteristics on the lumbar spine and femoral neck sites amongst older adults. An earlier meta-analysis of Martyn-St James and Carroll (2009) also observed that structured exercise protocols of combined loading and exercise programs of low-impact have the potential for preserving BMD at the lumbar spine and femoral neck in postmenopausal women. Moreover, the overall relative change in lumbar spine BMD estimated amongst the included RCTs was small (0.011 g/cm^2) but consistent with other exercise reviews in postmenopausal women (Wallace and Cumming 2000; Martyn-St James and Carroll 2006, 2009). Our findings were also consistent with other reviews also reporting a significant effect of exercise on femoral neck BMD (Martyn-St James and Carroll 2008; Wallace and Cumming 2000; Martyn-St James and Carroll 2006, 2009). However, our results were slightly larger, which is surprising given the absence of high-impact studies, the higher age of participants and the fact that only RCTs were included. It has been recognized that trials employing random allocation methods prevent selection and confounding biases (Akobeng 2005) but will yield more conservative results compared with non-random allocation methods (Moher et al. 2001). When previous reviews restricted the overall analyses to only RCTs study groups, results became non-significant (Martyn-St James and Carroll 2009) or modest compared with controlled trials (Wolff et al. 1999).

Conclusions

Implications for practice

The purpose of this systematic review and meta-analysis was to report more information on exercise training by looking at the body of evidence, for the goal of prescribing optimal exercise regimes as therapy for aging bone loss. The current ACSM position statement on exercise and physical activity for older adults (Chodzko-Zajko et al. 2009) concludes that aerobic exercise training may be effective

in counteracting age-related declines in BMD and that high-intensity resistance exercise training preserves or improves BMD. However, those conclusions were based mostly on previous meta-analyses of exercise effects in premenopausal and postmenopausal women (mean age of the postmenopausal subject groups is close to 55–60 years) due to lack of previous meta-analysis in older adults (age >60 years). Furthermore, although the majority of evidence indicates that exercises generating high-intensity loading forces are more effective in increasing BMD (Nikander et al. 2005), none of the RCTs in our review included high-impact loading protocols such as vertical jumps, rope jumping or running at >9 km/h. As older adults compared to young samples (such as premenopausal and postmenopausal women) are more physically prone to injuries due to physiological and functional decline (Tanaka and Seals 2003), high-impact loading protocols incorporating the evidence based in animal models, which supports the notion that greater strain magnitudes provide the most effective stimuli for bone formation (Bailey and Brooke-Wavell 2008), may not easily and effectively be incorporated in exercise prescriptions.

From a clinical point of view, a decrease of 1 SD in femoral neck BMD was associated with an increase in risk ratio for hip fracture by 2.94 in older men and by 2.88 in older women at the age of 65 years (Johnell et al. 2005). Therefore, it seems probable that the increase in BMD reported here might represent a relevant decrease in relative risk for hip fracture.

The largest effect sizes at both lumbar spine and femoral neck were observed in protocols that combined loading studies of impact activity mixed with high-magnitude joint reaction force loading through resistance training (Englund et al. 2005; Kemmler et al. 2010; Lord et al. 1996; Villareal et al. 2004). Accordingly, this specific exercise type may in fact provide a loading stimulus that is both adequate in its strain magnitude and rates and unusual in its loading pattern distributions at these sites. However, because of the differing combinations of skeletal loading activities evaluated in the trials included in these analyses (high inconsistency between trials) and the reduced number of included studies, the recommendation of this type of impact components should be interpreted with caution. Current recommendations regarding optimum exercise for preserving bone mineral density in older adults should advise that

low-impact activities seem to be ineffective in increasing BMD, apart from other physiological and psychological benefits that may be derived from participation in this type of exercise activities.

Implications for research

Future studies of the prevention of bone loss or fracture should not consider low-intensity impact training but, rather, should focus on the optimal combination of impact loading activities (mainly odd-impact loading exercises) and possibly on the improvement of muscle strength and balance impairment by using additional exercises focused on those functional components. Muscle strength and balance outcomes should be reported in addition to bone health end-points. Studies addressing the association between the increments on BMD with exercise loading and the relative risk reduction of fractures risk are required. Studies may also focus on the interaction between exercise and other risk factors for hip fracture (e.g. weight loss, body composition and pharmacological treatments). There is obviously a need for RCTs rather than other types of study, for better descriptions of studies, including the exercise protocol description, specifically the skeletal loading characteristics of the protocols, and trials should consistently apply the standards that are available for reporting of RCTs (CONSORT). Concern about bone health became especially relevant with postmenopausal bone loss; however, given the prevalence of low BMD in men ≥ 50 years of age (Kiebzak et al. 2002) and the scarce data on exercise effects, it may be especially important to focus on this population. Moreover, there is a lack of studies targeting osteoporotic subjects (irrespective to age), which are required in order to establish the therapeutic effect of exercise in those fragile individuals.

Fifty-eight percent of studies in this review failed to cite ethnicity, and because it is considered a critical variable, this should be reported in future exercise studies. In addition, future research needs to assess and report data on calcium and vitamin D intake because reduced supplies of calcium are associated with a reduced bone mass and osteoporosis, whereas vitamin D deficiency (commonly found in the elderly) leads to a decreased mineralization of bone (Gennari 2001), and there is subsequent confounding potential in relation to exercise-induced changes in BMD.

Considering that the time taken for completion of the bone remodeling cycle (bone resorption, formation and mineralization) is around 3–4 months (Frost 1986) and the varying precision of imaging techniques to assess bone changes in different time intervals, exercise interventions with longer duration (more than 1 year) are clearly advised.

Strength and stiffness are biomechanical parameters typically used to characterize the integrity of bone, which depends on a number of interrelated factors, including bone density, size and shape (Griffith and Genant 2008). Although advances in noninvasive bone imaging techniques, such as peripheral quantitative computed tomography (pQCT) and magnetic resonance imaging (MRI), have been made, the potential of exercise for improving bone strength remains controversial. Thus, there is a need for further well-designed (long-term and adequate sample size) RCTs that properly address this topic.

Acknowledgments This research was funded by the Portuguese Foundation of Science and Technology, grant FCOMP-01-0124-FEDER-009587 - PTDC/DES/102094/2008, and individual grants SFRH/BD/36319/2007 and SFRH/BSAB/1025/2010.

References

- Akobeng AK (2005) Understanding randomised controlled trials. *Arch Dis Child* 90(8):840–844. doi:10.1136/adc.2004.058222
- Bailey CA, Brooke-Wavell K (2008) Exercise for optimising peak bone mass in women. *Proc Nutr Soc* 67(1):9–18
- Berard A, Bravo G, Gauthier P (1997) Meta-analysis of the effectiveness of physical activity for the prevention of bone loss in postmenopausal women. *Osteoporos Int* 7(4):331–337
- Berlin JA (1997) Does blinding of readers affect the results of meta-analyses? *Lancet* 350:185
- Borenstein M, Hedges LV, Higgins JP, Rothstein HR (2009) *Introduction to meta-analysis*. Wiley, Chichester
- Brooke-Wavell K, Jones PR, Hardman AE (1997) Brisk walking reduces calcaneal bone loss in post-menopausal women. *Clin Sci (Lond)* 92(1):75–80
- Chodzko-Zajko WJ, Proctor DN, Fiatarone Singh MA, Minson CT, Nigg CR, Salem GJ, Skinner JS (2009) American College of Sports Medicine position stand. Exercise and physical activity for older adults. *Med Sci Sports Exerc* 41(7):1510–1530. doi:10.1249/MSS.0b013e3181a0c95c
- Chuin A, Labonté M, Tessier D, Khalil A, Bobeuf F, Doyon CY, Rieth N, Dionne IJ (2009) Effect of antioxidants combined to resistance training on BMD in elderly women: a pilot study. *Osteoporos Int* 20(7):1253–1258
- Cooper C, Cole ZA, Holroyd CR, Earl SC, Harvey NC, Dennison EM, Melton LJ, Cummings SR, Kanis JA (2011) Secular

- trends in the incidence of hip and other osteoporotic fractures. *Osteoporos Int*. doi:10.1007/s00198-011-1601-6
- Cummings SR, Melton LJ (2002) Epidemiology and outcomes of osteoporotic fractures. *Lancet* 359(9319):1761–1767. doi:10.1016/S0140-6736(02)08657-9
- Dhanwal DK, Dennison EM, Harvey NC, Cooper C (2011) Epidemiology of hip fracture: worldwide geographic variation. *Indian J Orthop* 45(1):15–22. doi:10.4103/0019-5413.73656
- Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56(2):455–463
- Egger M, Davey Smith G, Schneider M, Minder C (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315(7109):629–634
- Englund U, Littbrand H, Sondell A, Pettersson U, Bucht G (2005) A 1-year combined weight-bearing training program is beneficial for bone mineral density and neuromuscular function in older women. *Osteoporos Int* 16(9):1117–1123
- Frost HM (1986) Intermediary organization of the skeleton. CRC, Boca Raton
- Gennari C (2001) Calcium and vitamin D nutrition and bone disease of the elderly. *Public Health Nutr* 4(2B):547–559
- Griffith JF, Genant HK (2008) Bone mass and architecture determination: state of the art. *Best Pract Res Clin Endocrinol Metab* 22(5):737–764. doi:10.1016/j.beem.2008.07.003
- Gullberg B, Johnell O, Kanis JA (1997) World-wide projections for hip fracture. *Osteoporos Int* 7(5):407–413
- Higgins JPT, Green S (eds) (2009) *Cochrane handbook for systematic reviews of interventions* version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from <http://www.cochrane-handbook.org>.
- Higgins J, Thompson S, Deeks J, Altman D (2002) Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 7(1):51–61
- Hind K, Burrows M (2007) Weight-bearing exercise and bone mineral accrual in children and adolescents: a review of controlled trials. *Bone* 40(1):14–27. doi:10.1016/j.bone.2006.07.006
- Hsieh YF, Robling AG, Ambrosius WT, Burr DB, Turner CH (2001) Mechanical loading of diaphyseal bone in vivo: the strain threshold for an osteogenic response varies with location. *J Bone Miner Res* 16(12):2291–2297. doi:10.1359/jbmr.2001.16.12.2291
- Jessup JV, Horne C, Vishen RK, Wheeler D (2003) Effects of exercise on bone density, balance, and self-efficacy in older women. *Biol Res Nurs* 4(3):171
- Johnell O, Kanis J (2005) Epidemiology of osteoporotic fractures. *Osteoporos Int* 16(Suppl 2):S3–S7. doi:10.1007/s00198-004-1702-6
- Johnell O, Kanis JA, Oden A, Johansson H, De Laet C, Delmas P, Eisman JA, Fujiwara S, Kroger H, Mellstrom D, Meunier PJ, Melton LJ 3rd, O'Neill T, Pols H, Reeve J, Silman A, Tenenhouse A (2005) Predictive value of BMD for hip and other fractures. *J Bone Miner Res* 20(7):1185–1194. doi:10.1359/JBMR.050304
- Kanis JA, McCloskey EV, Johansson H, Oden A, Melton LJ 3rd, Khaltayev N (2008) A reference standard for the description of osteoporosis. *Bone* 42(3):467–475. doi:10.1016/j.bone.2007.11.001
- Kelley GA, Kelley KS, Tran ZV (2001) Resistance training and bone mineral density in women: a meta-analysis of controlled trials. *Am J Phys Med Rehabil* 80(1):65–77
- Kemmler W, Von Stengel S, Engelke K, Häberle L, Kalender WA (2010) Exercise effects on bone mineral density, falls, coronary risk factors, and health care costs in older women: the randomized controlled senior fitness and prevention (SEFIP) study. *Arch Intern Med* 170(2):179–185
- Khosla S, Riggs BL (2005) Pathophysiology of age-related bone loss and osteoporosis. *Endocrinol Metab Clin North Am* 34(4):1015–1030. doi:10.1016/j.ecl.2005.07.009
- Kiebzak GM, Beinart GA, Perser K, Ambrose CG, Siff SJ, Heggeness MH (2002) Undertreatment of osteoporosis in men with hip fracture. *Arch Intern Med* 162(19):2217–2222
- Kohrt WM, Bloomfield SA, Little KD, Nelson ME, Yingling VR (2004) American College of Sports Medicine position stand: physical activity and bone health. *Med Sci Sports Exerc* 36(11):1985–1996
- Korpelainen R, Keinanen-Kiukaanniemi S, Heikkinen J, Vaananen K, Korpelainen J (2006) Effect of impact exercise on bone mineral density in elderly women with low BMD: a population-based randomized controlled 30-month intervention. *Osteoporos Int* 17:109–118
- Lanyon LE (1996) Using functional loading to influence bone mass and architecture: objectives, mechanisms, and relationship with estrogen of the mechanically adaptive process in bone. *Bone* 18(1 Suppl):37S–43S
- Lanyon L, Skerry T (2001) Postmenopausal osteoporosis as a failure of bone's adaptation to functional loading: a hypothesis. *J Bone Miner Res* 16(11):1937–1947. doi:10.1359/jbmr.2001.16.11.1937
- Lau EM, Woo J, Leung PC, Swaminathan R, Leung D (1992) The effects of calcium supplementation and exercise on bone density in elderly Chinese women. *Osteoporos Int* 2(4):168–173
- Lord SR, Ward JA, Williams P, Zivanovic E (1996) The effects of a community exercise program on fracture risk factors in older women. *Osteoporos* 6(5):361–367
- MacKellvie KJ, Khan KM, McKay HA (2002) Is there a critical period for bone response to weight-bearing exercise in children and adolescents? A systematic review. *Br J Sports Med* 36(4):250–257
- Marques EA, Mota J, Machado L, Sousa F, Coelho M, Moreira P, Carvalho J (2011a) Multicomponent training program with weight-bearing exercises elicits favorable bone density, muscle strength, and balance adaptations in older women. *Calcif Tissue Int* 88(2):117–129. doi:10.1007/s00223-010-9437-1
- Marques EA, Wanderley F, Machado L, Sousa F, Viana JL, Moreira-Goncalves D, Moreira P, Mota J, Carvalho J (2011b) Effects of resistance and aerobic exercise on physical function, bone mineral density, OPG and RANKL in older women. *Exp Gerontol* 46(7):524–532. doi:10.1016/j.exger.2011.02.005
- Martyn-St James M, Carroll S (2006) High-intensity resistance training and postmenopausal bone loss: a meta-analysis. *Osteoporos Int* 17(8):1225–1240
- Martyn-St James M, Carroll S (2008) Meta-analysis of walking for preservation of bone mineral density in postmenopausal women. *Bone* 43(3):521–531. doi:10.1016/j.bone.2008.05.012

- Martyn-St James M, Carroll S (2009) A meta-analysis of impact exercise on postmenopausal bone loss: the case for mixed loading exercise programmes. *Br J Sports Med* 43(12):898–908. doi:10.1136/bjism.2008.052704
- Moher D, Schulz KF, Altman D (2001) The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA* 285(15):1987–1991
- Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, Macera CA, Castaneda-Sceppa C (2007) Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Circulation* 116(9):1094–1105
- Nichols JF, Nelson KP, Peterson KK, Sartoris DJ (1995) Bone mineral density responses to high-intensity strength training in active older women. *J Aging Phys Act* 3(1):26–38
- Nikander R, Sievanen H, Heinonen A, Kannus P (2005) Femoral neck structure in adult female athletes subjected to different loading modalities. *J Bone Miner Res* 20(3):520–528. doi:10.1359/JBMR.041119
- Park H, Kim KJ, Komatsu T, Park SK, Mutoh Y (2008) Effect of combined exercise training on bone, body balance, and gait ability: a randomized controlled study in community-dwelling elderly women. *J Bone Miner Metabol* 26(3):254–259
- Rhodes EC, Martin AD, Taunton JE, Donnelly M, Warren J, Elliot J (2000) Effects of one year of resistance training on the relation between muscular strength and bone density in elderly women. *Br J Sports Med* 34(1):18–22
- Schwab P, Scalapino K (2011) Exercise for bone health: rationale and prescription. *Curr Opin Rheumatol* 23(2):137–141. doi:10.1097/BOR.0b013e3283434501
- Sterne JA, Gavaghan D, Egger M (2000) Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 53(11):1119–1129. doi:10.1016/S0895-4356(00)00242-0
- Taaffe DR, Duret C, Wheeler S, Marcus R (1999) Once-weekly resistance exercise improves muscle strength and neuromuscular performance in older adults. *J Am Geriatr Soc* 47(10):1208–1214
- Tanaka H, Seals DR (2003) Invited review: dynamic exercise performance in Masters athletes: insight into the effects of primary human aging on physiological functional capacity. *J Appl Physiol* 95(5):2152–2162. doi:10.1152/jappphysiol.00320.2003
- Vainionpää A, Korpelainen R, Vihriala E, Rinta-Paavola A, Leppaluoto J, Jamsa T (2006) Intensity of exercise is associated with bone density change in premenopausal women. *Osteoporos Int* 17(3):455–463
- Villareal DT, Steger-May K, Schechtman KB, Yarasheski KE, Brown M, Sinacore DR, Binder EF (2004) Effects of exercise training on bone mineral density in frail older women and men: a randomised controlled trial. *Age Ageing* 33(3):309–312
- Vincent KR, Braith RW (2002) Resistance exercise and bone turnover in elderly men and women. *Med Sci Sports Exerc* 34(1):17–23
- von Stengel S, Kemmler W, Engelke K, Kalender WA (2011a) Effects of whole body vibration on bone mineral density and falls: results of the randomized controlled ELVIS study with postmenopausal women. *Osteoporos Int* 22:317–325
- von Stengel S, Kemmler W, Bebenek M, Engelke K, Kalender WA (2011b) Effects of whole body vibration training on different devices on bone mineral density. *Med Sci Sports Exerc* 43(6):1071–1079. doi:10.1249/MSS.0b013e318202f3d3
- Wallace BA, Cumming RG (2000) Systematic review of randomized trials of the effect of exercise on bone mass in pre- and postmenopausal women. *Calcif Tissue Int* 67(1):10–18
- Watts NB (2004) Fundamentals and pitfalls of bone densitometry using dual-energy X-ray absorptiometry (DXA). *Osteoporos Int* 15(11):847–854. doi:10.1007/s00198-004-1681-7
- Wolff I, van Croonenborg JJ, Kemper HC, Kostense PJ, Twisk JW (1999) The effect of exercise training programs on bone mass: a meta-analysis of published controlled trials in pre- and postmenopausal women. *Osteoporos Int* 9(1):1–12
- Woo J, Hong A, Lau E, Lynn H (2007) A randomised controlled trial of Tai Chi and resistance exercise on bone health, muscle strength and balance in community-living elderly people. *Age Ageing* 36:262–268
- Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 336(7644):601–605. doi:10.1136/bmj.39465.451748.AD