The effects of exercise on bone mineral density in men: a systematic review and meta-analysis of randomised controlled trials

Blair R. Hamilton[1,2,3], Katherine A. Staines[3,4], George A. Kelley[5], Kristi S. Kelley[5]
Wendy M. Kohrt[6], Yannis Pitsiladis[2,3,7] and Fergus M. Guppy *[3,4]

Blair R. Hamilton, Email: b.r.hamilton@brighton.ac.uk
Katherine A. Staines, Email: k.staines@brighton.ac.uk
George A. Kelley, Email: gkelley@hsc.wvu.edu
Kristi S. Kelley, Email: kskelley@hsc.wvu.edu
Wendy M. Kohrt, Email: wendy.kohrt@cuanschutz.edu
Yannis Pitsiladis, Email: y.pitsiladis@brighton.ac.uk

Corresponding author:
Fergus M. Guppy *PhD*
School of Pharmacy and Biomolecular Sciences
Huxley Building
Lewes Road
University of Brighton
Brighton
BN2 4GJ
United Kingdom
Email: f.guppy@brighton.ac.uk
Tel: +44 (0)1273 641631

Twitter: @fergusguppy

1. The Gender Identity Clinic, Tavistock and Portman NHS Foundation Trust, London, UK
2. School of Sport and Service Management, University of Brighton, UK
3. Centre for Stress and Age-related Disease, University of Brighton, UK
4. School of Pharmacy and Biomolecular Sciences, University of Brighton, UK
5. Department of Epidemiology and Biostatistics, School of Public Health, West Virginia University, USA
6. Department of Medicine, University of Colorado Anschutz Medical Campus, and Eastern Colorado VA Geriatric, Research, Education, and Clinical Center, USA
7. Department of Movement, Human and Health Sciences, University of Rome “Foro Italico”, Rome, Italy

Word count: 4544

Key words: Exercise; Bone; Osteoporosis; Men; Meta-analysis

Please cite as: Hamilton *et. al*. (2021) The effects of exercise on bone mineral density in men: a systematic review and meta-analysis of randomised controlled trials. SportRχiv DOI:

All authors are aware that this has been submitted as a preprint

DOI: STORK.XXX.XXXX

ABSTRACT

Objective The aim of this systematic review and meta-analysis was to provide an updated analysis, including the use of more robust methods, on the effects of exercise on bone mineral density in men.

Methods: Randomised Control Trials of >24 weeks and published in English up to 01/05/20 were retrieved from 3 electronic databases, cross-referencing, and expert review. The primary outcome measures were changes in FN, LS, and lower limb BMD Standardised effect sizes were calculated from each study and pooled using the inverse heterogeneity model.

Results: A statistically significant benefit of exercise was observed on FN BMD (g = 0.21 [0.03, 0.40], Z = 2.23 p = 0.03), with no observed statistically significant benefit of exercise on LS BMD (g = 0.10 [-0.07, 0.26], Z = 1.15 p = 0.25).

Conclusion: This analysis provided additional evidence to recommend ground- and/or joint-reaction force exercises for improving or maintaining FN, but not LS BMD. Additional well-designed RCTs are unlikely to alter this evidence, although interventions that include activities that directly load the lumbar spine are needed to ensure this is not a potential method of improving LS BMD.

PROSPERO Registration number: CRD420201804

STRENGTHS AND LIMITATIONS

- This is a comprehensive systematic review and meta-analysis that updates the knowledge base from a previous review published in 2013.
- This meta-analysis uses the more robust inverse heterogeneity model that has not been previously used in this area.
- Strengthening of the evidence from the current analysis, it appears that exercise should now be *recommended* in this population.

DOI: STORK.XXX.XXXX

INTRODUCTION

Osteoporosis is a skeletal disorder characterized by low bone mineral density (BMD), compromised bone strength and architectural damage of the bone[1]. A reported 8.9 million osteoporotic fractures occur in Europe each year[1] and it is estimated to affect 200 million people worldwide[2]. While osteoporosis is four-fold more common in women than men, men have more osteoporosis-related complications[3, 4] with 30% experiencing a fracture[5]. Men who suffer a proximal femur fracture are also younger, less healthy and have a 2-3 fold higher mortality [6-8] and morbidity [9] than women who incur this fracture.

Exercise is a cost-effective, widely available intervention that has been reported to help maintain optimal BMD in men[10]; however, consideration of exercise modality is needed if the aim is to promote skeletal health. For example, swimming and cycling have been shown to have little or no effect on BMD[11, 12], with some studies finding that these forms of exercise are associated with low BMD at both the lumbar spine (LS) and proximal femur when compared to inactive controls[13-15]. Weight-bearing, high-impact and strength training exercises have been widely recognised to be a gold standard for long-term skeletal health, with a 60% reduction in hip-fracture risk in men who were physically active compared to inactive men[16]. A previous meta-analysis of randomised controlled trials (RCTs) by Kelley *et al.* in 2013[10] investigated the effects of exercise on BMD in men. The authors observed a moderate benefit of exercise on femoral neck (FN) BMD (g = 0.583 [0.031, 1.135]) but no benefit on LS BMD (g = 0.190 [-0.036, 0.416]). They concluded that there was insufficient evidence at that time to recommend ground- and/or joint-reaction force exercise for improving and/or maintaining FN and LS BMD in men and recommended that additional well-designed RCTs in men should be performed to formulate any final recommendations.

While the results reported by Kelley *et al.*[10] are noteworthy, they were limited to only three RCTs published up to August 2011 and lacked an assessment of BMD using quantitative computed tomography (QCT)[10]. However, since that time, additional RCTs have been

DOI: STORK.XXX.XXXX

published[17, 18] and more robust methods for the undertaking and interpretation of meta-analytic results have been developed[19-22]. Furthermore, to the best of the authors' knowledge, no systematic review of previous systematic reviews with meta-analysis or original systematic review with meta-analysis has been conducted on the effects of exercise on BMD in men since the original analysis. Finally, using previously developed guidelines for when to update a systematic review, it was decided that an updated review on this topic was needed [23]. Thus, given 1) the deleterious consequences of low BMD in men, 2) the potential benefits of certain types of exercise on BMD in men[10], 3) the lack of recent meta-analytic work in this area, 4) the use of more robust methods for conducting meta-analytic research[19-22] and 5) decision tree analysis of when to update a systematic review[23] we aimed to update the systematic review with meta-analysis by Kelley *et al*.[10], whereby we will examine the effects of exercise on BMD in men.

MATERIALS AND METHODS

Study Eligibility Criteria

As this meta-analysis aimed to update the Kelley *et al*.[10] meta-analysis, the same *a priori* inclusion criteria were employed (**Table 1**), with additional studies identified from 01/08/2011 forward. Studies not meeting the criteria outlined in **Error! Reference source not found.** were excluded from the analysis. Studies were limited to RCTs to ensure that confounders that are not understood were controlled, as well as to eliminate the overestimation that has been described in non-RCTs [24, 25]. Studies with multiple interventions (e.g., diet combined with exercise) were included only if there was an appropriate comparative control group (e.g., diet only). Resistance training studies were limited to those that included lower body activities.

This meta-analysis adhered to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines[26]. The protocol for this meta-analysis[27] was preregistered in PROSPERO (trial registration number: CRD42020180441 [28]).

DOI: STORK.XXX.XXXX

Table 1. Study eligibility criteria.

Criteria
1) Randomised trials with comparative control group
2) Published in English
3) Men 18 years of age and older
4) Participants not taking part in regular exercise prior to the study enrolment
5) Ground- and/or joint-reaction force exercise intervention of at least 24 weeks duration
6) Included in Kelley et *al.,* [10] OR published since 01 August 2011
7) Data available for changes in FN and/or LS BMD as assessed by dual-energy x-ray absorptiometry (DXA), dual-photon absorptiometry (DPA), or quantitative computed tomography (QCT)

Data Sources

Studies published up to 01/05/2020 were retrieved from three electronic sources (PubMed, Embase, SportDiscus). Keywords relevant to all searches included “exercise”, “bone”, and “randomised”. Based on PRISMA guidelines[26], an example of the search strategy can be found in **Supplementary material**[40]. The first author (BRH) conducted all electronic database searches. In addition to electronic database searches, cross-referencing from retrieved studies was also conducted.

Study Records and Selection

All studies were imported into EndNote (EndNote X9.3.1, Clarivate Analytics, USA) and duplicates removed electronically and manually by the first author (BRH). A copy of the reference database was then provided to the last author (FMG) for dual screening. Both authors (BRH and FMG) selected studies independent of each other. Multiple studies were handled by including only the most recently published articles. The screeners were not blinded to either the journal titles or to the study authors/affiliations. Reasons for exclusion were coded based on one or more of the following: 1) inappropriate population, 2) inappropriate intervention, 3) inappropriate comparison(s), 4) inappropriate outcome(s), 5) inappropriate study design or 6) other. On completion, the screeners met to discuss their selections and reconciled any discrepancies by consensus. If an agreement could not be achieved, the second author (KAS) provided a recommendation. The agreement rate, before the

DOI: STORK.XXX.XXXX

reconciliation of any discrepancies, was calculated using Cohen's κ statistic [29]. The precision of searches was also calculated as the number of studies included divided by the number of studies screened (less duplicates)[30]. We then calculated the number-needed-to-screen (NNS) by taking the reciprocal of the precision[30].

Data Abstraction

Prior to data abstraction, an electronic codebook developed by the authors of Kelley *et al.* [10] [GAK, KSK, WMK] was provided to the first and last authors (BRH/FMG). The extracted data were coded based on the following major categories; 1) study characteristics (e.g., author, journal, year, etc.), 2) participant characteristics (e.g., age, height, mass, etc.), 3) intervention details (e.g., type, length, frequency, etc.), and 4) outcome characteristics (e.g., sample sizes, baseline/post-exercise means and SDs, etc.).

The first (BRH) and last (FMG) authors extracted all data independent of one another before meeting to resolve any discrepancies by consensus. If an agreement could not be achieved, the second author (KAS) provided a recommendation. Prior to this, the overall agreement rate was assessed by Cohen's κ statistic[29].

Outcome Measures

A priori primary outcome measures were changes in FN and LS BMD measured by dual-energy x-ray absorptiometry (DXA), dual-photon absorptiometry (DPA), or quantitative computed tomography (QCT). Secondary, *a priori* outcomes included changes in body mass (BM), body mass index (BMI) in $kg.m^2$, lean mass (LM), and fat mass (FM). Obtaining missing data was attempted for all primary and secondary outcome measures if assessed by a study but the data provided proved inadequate to calculate an effect size. The last author (FMG) contacted the studies corresponding author three times via email with one-week between each communication. These communications were tracked (e.g., dates, responses, success rates, etc.) to establish the success rate of this process.

Risk of Bias Assessment

The risk of bias for each study was assessed using the recently revised Cochrane Risk of Bias instrument for RCTs (RoB 2)[31]. Using one or more signalling questions, the RoB 2 instrument assesses the risk of bias in five distinct domains: (1) bias arising from the randomisation process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in measurement of the outcome, and (5) bias in selection of the reported result. Based on signalling questions, each domain is assessed as either 'low risk', 'high risk', or 'some concerns'. Based on responses to each domain, the overall risk of bias for each study is then assessed as either 'low risk', 'high risk', or 'some concerns'. We chose to use this risk of bias instrument over the various study quality instruments, including those focused on exercise intervention studies[32, 33] given the difficulty of the latter in differentiating between the quality of reporting and the quality in the conduct of a study[31].

No studies were excluded from the analysis based on the risk of bias assessment[34]. The first (BRH) and last (FMG) authors undertook the risk of bias assessment independently of one another, before meeting to resolve any discrepancies by consensus. Where this could not be achieved, the second author (KAS) provided a recommendation.

Statistical Analysis

Calculation of effect sizes

The *a priori* primary outcomes for this meta-analysis were changes in FN, LS and lower limb BMD, calculated using the Hedges standardised mean difference effect size (ES), *g*, adjusted for small sample sizes[35]. The *g* for each group was calculated as the change score difference (absolute or relative) in the exercise group minus the change score difference for the control group, divided by the pooled standard deviation. If this information was not available, g was calculated using procedures described by Follmann *et al*.[36]. For studies reporting multiple post-intervention time points, *g* was calculated based on baseline and the

DOI: STORK.XXX.XXXX

final time point closest to the end of the intervention. Secondary, *a priori* outcomes for this meta-analysis were changes in BMI, FM, LM and BM, calculated using the original metric weighted mean difference (WMD).

Effect size pooling

Results were pooled using the recently developed inverse heterogeneity (IVhet) model[19], a model which is more robust than the traditional random-effects model employed by Kelley *et al*.[10]. Two-tailed z-alpha values <0.05 and non-overlapping 95% confidence intervals were considered statistically significant.

Heterogeneity and Inconsistency

For each pooled outcome, heterogeneity was assessed using Q[37], with an alpha level of <0.10 considered to represent statistically significant heterogeneity. Inconsistency was assessed using I^2, an extension of Q. For this meta-analysis, inconsistency was categorised as very low (<25%), low (25-50%), moderate (50-75%) or large (>75%)[37]. Absolute between-study heterogeneity was assessed using tau squared (τ^2). In addition, influence analysis was conducted by removing each study from our analysis once to examine the effect of that study on the overall findings. Given the expected small sample size, no subgroup or meta-regression analysis were planned a *priori* or conducted *post hoc*.

Meta-biases

Small-study effects (publication bias, etc.) were assessed qualitatively using the Doi plot and quantitatively using the Luis Furuya-Kanamori index (LFK index)[22, 38]. The Doi plot has been suggested to be more intuitive than the funnel plot and the LFK index more robust than the commonly used Egger' regression-intercept test[22, 38]. LFK values within ± 1, greater than ± 1 but within ± 2, and greater than ± 2 are considered to represent no, minor, and major asymmetry[22].

DOI: STORK.XXX.XXXX

Strength of evidence

The strength of findings for each outcome was assessed using the most recent version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) for meta-analysis tool [39]. Quality of evidence was assessed across the domains of risk of bias, consistency, directness, precision, and publication bias. Quality was judged as high (further research is very unlikely to change our confidence in the estimate of effect), moderate (further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate), low (further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate), or very low (very uncertain about the estimate of effect[39]

Software used for analysis

All data were analysed using Meta XL (version 5.3). All data are available as supplementary material[40].

PATIENT AND PUBLIC INVOLVEMENT

There was no direct patient or public involvement in this review.

RESULTS

Study Characteristics

A flow diagram that depicts the search process for study selection is shown in **Figure 1.** A full list of studies is available in **Supplementary Material A**. After initially identifying 140 citations, removal of 5 studies due to publication before 1st August 2011 and 36 duplicates both electronically and manually, 99 studies were screened. Of the 99 studies reviewed, 8 additional studies were included in this meta-analysis, taking the total number of studies to 11 and representing 30 groups (19 exercise, 11 control) and 883 participants (524 exercise, 359 control). The agreement rate between assessors for inclusion was 0.82, prior to the reconciliation of any discrepancies, with these all achieved by consensus between the first and last author. As with the original Kelley *et al.*[10] study, the number of groups exceeded

DOI: STORK.XXX.XXXX

the number of studies due to multiple studies including more than one intervention group[17, 41-45]. The major reasons for exclusion were inappropriate population (8.7%), inappropriate intervention (19.6%), inappropriate comparison (0%), inappropriate outcome(s) (28.3%), inappropriate study design (34.8%) and other (8.7%). The precision of searches was 8% while the number needed to screen was 12, with a general description of the studies included found in **Error! Reference source not found.**.

Of the additional studies included, all were published in peer-reviewed English-language journals, starting in 2014 and ending in 2020. The studies were conducted in a variety of countries; 4 in Australia, 3 in Denmark, and one each in Germany, South Korea, China and the United States. The maximum number of men for which final BMD assessment was available ranged from 6 to 109 in the exercise groups (mean ± SD= 30 ± 26, [median = 21]) and from 6 to 105 in the control groups (mean ± SD = 31 ± 29, [median =20]). None of the studies utilised a cross-over design and 9 provided sample size estimates[17, 41, 42, 44-49], while 2 did not[43, 50]. One study[44] was found to be a semi-randomized trial due to the participants in this study self-selecting to take part in the control group. However, a post hoc decision was made to include this semi-randomised trial as both exercise groups were randomly assigned. The agreement rate on data extraction was 1.0, with no discrepancies between the authors. One study[17] did not include all data within the manuscript and following three emails to the corresponding author there was no response.

Participant Characteristics

A description of the baseline characteristics of participants can be seen in **Error! Reference source not found.** and **Error! Reference source not found.**. Reported dropout rates ranged from 0 to 68% in the exercise groups (mean ± SD = 16 ± 20%, median= 10%) and 0 to 62% in the control groups (mean ± SD = 15 ± 17%, median= 8%).

DOI: STORK.XXX.XXXX

Table 2 Baseline characteristics of participants

Variable	Exercise					Control				
	Groups (#)	Participants (#)	Mean ± SD	Mdn	Range	Groups (#)	Participants (#)	Mean ± SD	Mdn	Range
Age (y)	17	558	66.1 ± 5.7	68	57-77	12	416	66.5 ± 5.4	67	58 - 80
BMI (kg/m^2)	3	65	26.9 ± 2.9	27.4	23.6 - 31.4	2	38	26.5 ± 2.6	26.7	23.9 - 29.0
Fat Mass	4	237	25.3 ± 1.9	25.4	22.8 - 27.5	3	172	28.5 ± 1.2	27.7	27.1 - 30.2
Lean Mass	6	296	52.6 ± 5.7	54.9	44.2 - 58.1	4	201	55.0 ± 5.6	57.6	45.4 - 59.3
BMD (g/cm^2)										
Femoral Neck	19	524	0.794 ± 0.215	0.862	0.212 -1.010	11	359	0.840 ± 0.198	0.919	0.236 -1.017
Lumbar Spine	12	440	1.115 ± 0.093	1.111	0.950 - 1.247	9	327	1.143 ± 0.116	1.140	0.960 - 1.310
BMD (g/cm^3)										
Lumbar Spine	1	19	0.176 ± 0.012	-	-	1	19	0.166 ± 0.028	-	-

Notes: #, Number of; SD, Standard Deviation; Mdn, Median; y, years; BMI, Body Mass Index; BMD, Bone Mineral Density; g/cm, grams per centimetre.

DOI: STORK.XXX.XXXX

Exercise Intervention Characteristics

A description of the exercise interventions can be seen in **Table 3**. The exercise interventions varied in length from 24 to 76 weeks (mean ± SD = 44 ± 16, median = 52 weeks). Compliance ranged from 46 to 96% (mean ± SD = 71 ± 13%, median = 70 %). Ten groups participated in supervised exercise, five participated in a combination of supervised and unsupervised activity while two took part in unsupervised exercise only. The location of the studies varied, with ten groups taking part in a facility-based activity, two in home-based activity, and five in a combination of home and facility-based activity.

BMD Assessment Characteristics

With the exception of one study[49] which used QCT to assess LS BMD, all of the studies assessing FN or LS BMD did so using DXA, Three studies used a Hologic DXA system (QDR 4500 Elite[41] or Discovery[17, 45]), while five studies used a Lunar DXA system (iDXA[43, 47], Prodigy[42, 46, 50] or Medix DXA[44] . Two studies did not report the DXA[46] or QCT[48] models utilised in their respective studies. Insufficient data were reported for the site-specific reliability of the instruments used to assess BMD at either the LS or FN.

Risk of Bias Assessment

Assessment using the Cochrane Risk of Bias Assessment Instrument (RoB2) is shown in **Figure 2** and **Supplementary Material B**. As can be seen, 55.3% of the studies were at an unclear or high risk of bias concerning: (1) deviations from intended interventions (22.2%), and (2) randomization process (33.1%). Given the inability to blind participants in exercise intervention trials, all studies were at a high risk of bias for the category 'Blinding of participants', except one study that stated participants/assessors were not blinded to the prime outcomes[47]. The overall risk of bias across all categories was approximately 50%.

DOI: STORK.XXX.XXXX

Table 3 General characteristics of the studies included

Study	Country	Participants	Exercise Intervention	BMD Assessment
Hong [41]	China	82 healthy men 65 to 74 y of age randomly assigned to a Tai Chi (n=26), resistance training (n=27), or control (n=29) group	3 d/wk.: Tai Chi: Yang style, 24 forms, 45 min; Resistance Training: 1 set, 30 reps, 7 exercises, TheraBand's used for resistance; for 12 months	DXA (Hologic QDR 4500 Elite) at the FN & LS
Kukuljan *et al*. [42]	Australia	176 healthy men 50 to 79 y of age randomly assigned to an exercise (n=46), exercise + milk (n=43), control (n=44) or milk (n=43) group	3 d/wk., 60–75 min/session, Resistance Training: 2–3 sets, 8–20 reps, 50–85% 1RM, 6–8 exercises plus 3 moderate-impact weight-bearing exercises (jumping & stepping) in between resistance exercises 3 sets of 10–20 reps, for 18 months	DXA (GE Lunar Prodigy) at the FN & LS
Zeilman III [46]	United States	16 sedentary men with irritable bowel syndrome 41 to 75 y of age randomly assigned to either an exercise (n=7) or control (n=9) group	3 d/wk., 50 min/session, stretching, flexibility calisthenics & walking with weighted vests and a pedometer, for 32 wks.	DXA (GE Lunar Prodigy) at the FN & LS
Bjerre *et al*. [48]	Denmark	214 men age 18 y or older with prostate cancer randomly assigned to either a football group (n=109) or usual care (n=105) group	2 d/wk., 60min/session, coach led recreational football, 20 mins warm up based on the Fédération Internationale de Football Association (FIFA) 11 + program, 20 mins drills, 20 mins match play for 24 wks.	DXA (Not Disclosed) at the FN & LS
Bolam *et al*. [17]	Australia	42 community dwelling men 50 to 74 y of age randomly assigned to an upper body resistance exercise and high-dose impact-loading (HI), upper body resistance exercise and moderate-dose impact-loading (MOD) (n=15), or control (n=14) group	4 d/wk.: 60min/session, 2 sessions supervised in clinic by an exercise physiologist and 2 unsupervised sessions at home. HI group: resistance training of biceps curl, triceps extension, latissimus pull down and chest press, 2 sets of 12 reps at 60% 1RM combined with cross jumping, drop jumping, bounding (range of reps of 40-80). MOD group: resistance training of biceps curl, triceps extension, latissimus pull down and chest press, 2 sets of 12 reps at 60% 1RM combined with cross jumping, drop jumping, bounding (range of reps 20-40) for 9 months	DXA (Hologic Discovery W) at the FN and LS.
Harding *et al.* [44]	Australia	42 community dwelling men over 45 y of age with osteopenia or osteoporosis were randomly assigned to a High-intensity progressive resistance and impact training program (n=34) or an Isometric axial compression exercise program (n=33). A sample of age-matched men, recruited and screened using identical criteria to the two exercise arms, but self-selecting to 'no intervention', formed a non-randomised parallel control group (n=26)	2 d/wk.: High-intensity progressive resistance and impact training program(HiRIT): barbell-based resistance training component consisted of five sets of five repetitions (≥80–85% of 1RM) of the three fundamental exercises, the deadlift, squat and overhead press; Supervised machine-based isometric axial compression exercise training (IAC): One self-initiated near-maximal five-second isometric contraction (≥80–85% of 1RM) was performed for the seated chest press, seated leg press, core and arm pull, and vertical lift on the bioDensity™ (Performance Health Systems, Chicago, USA) machine; for 8 months	DXA (Medix DR) at the FN
Helge *et al*. [43]	Denmark	26 healthy sedentary elderly men were randomly assigned to either a football group (n=9), a resistance group (n=9) or an inactive control group (n=8).	2d/ wk. Football : of small-sided play on a natural grass pitch: 3 × 15-min play interspersed with 2-min rest periods (week 0–12) and 4 × 15-mins play interspersed with 2-min rest periods (week 13–52); Resistance: 5 min of low intensity warm-up followed by resistance training including five exercises, each set of exercises was separated by 1.5-min rest, 5 min of core training was performed at the end of each training session; for 12 months	DXA (iDXA, Lunar) at both right and left FN
Kemmler *et al.* [49]	Germany	43 sedentary community-dwelling older men 73 to 91 y of age with osteopenia/osteoporosis and SMI-based sarcopenia were randomly assigned to a HIT-RT exercise group (n = 21) or a control group (n = 22)	2 d/wk., resistance exercises on machines, intensity of the exercise was consistently scheduled by prescribing a range of reps (i.e., 5–7 or 8–10 reps) and the corresponding degree of work to failure ("effort") (e.g., maximum effort minus 1–3 reps; defined as non-repetition maximum [nRM]; for 12 months	QCT (Not Disclosed) at LS

DOI: STORK.XXX.XXXX

Kim *et al.* [50]	South Korea	41 men receiving androgen deprivation therapy with prostate cancer were randomly assigned to either a home-based exercise intervention for preventing osteoporosis (HEPO) (n=23) or stretching exercise (n=18) control group.	HEPO: 150 mins/wk., a core program including weight-bearing exercise and resistance exercise combined with an optional program including stabilization/balance exercise and circuit resistive calisthenics	DXA (Lunar Prodigy) at FN and LS
Newton *et al.* [45]	Australia	154 men age 43-90 y with prostate cancer were randomly assigned to either impact loading + resistance training (ImpRes) (n=57), aerobic + resistance training (AerRes) (n= 50) or delayed aerobic exercise control group (n=47)	. ImpRes: 2d/wk. consisting of a series of bounding, skipping, drop jumping, hopping, and leaping activities that produced ground reaction forces of 3–5 times body weight, and was progressive in nature for 12 months supervised. AerRes: 2d/wk. consisting of 20–30 min of exercise using various modes including walking/jogging on a treadmill and cycling or rowing on stationary ergometers with intensity set at 60%–85% of maximal heart rate, 6 months supervised in clinic and 6 months homebased.	DXA (Hologic Discovery) at FN and LS
Uth *et al.* [47]	Denmark	57 men receiving androgen deprivation therapy for more than 6 months were randomly assigned to either a football group (n=29) or a standard care control group (n=28).	Football Group: 2d/wk., 2×15 min games weeks 1-4 and 3×15 min week 5-8. 3d/wk. 3×15 min in weeks 9-12. Week 13 onwards 2d/wk. 60min sessions	DXA (iDXA, Lunar)

Notes: BMD: Bone Mineral Density; y: years; d/wk.: days per week; 1RM: 1 repetition max; min: minutes; FN: femoral neck; LS: lumbar spine; DXA: dual-energy x-ray absorptiometry

DOI: STORK.XXX.XXXX

Changes in primary outcome measures

Changes in FN BMD

Changes in FN BMD can be seen in **Table 4** and **Figure 3**. Overall, a significant benefit of exercise was observed on FN BMD (g = 0.21 [0.03, 0.40], Z = 2.23 p = 0.03), with no asymmetry (LFK index= 0.43, **Supplementary Figure 1**) observed. In addition, no statistically significant heterogeneity was observed (Q = 20.66, p = 0.19) and inconsistency was considered to be very low (I^2 = 22.55%). Findings were similar when results were collapsed so that only one effect size represented each study. Influence analysis showed that removal of the exercise only group from one study [42] had the biggest influence, with a non-significant effect observed when this study was removed from the analysis (g = 0.18 [-0.02, 0.37], Z = 1.78, p = 0.08). The removal of the exercise and milk group from the same study [42] had a smaller non-significant effect when removed from the analysis (g = 0.16 ([-0.01, 0.32], Z = 1.78, p = 0.06). The removal of other studies from the analysis did not change the outcome (**Supplementary Table 1**). An evidence profile for changes in FN BMD is shown in online **Supplementary Table 3**. As can be seen, the outcome was considered critical and the overall strength considered high, with future additional studies unlikely to influence the overall direction of findings.

Changes in LS BMD

Changes in LS BMD can be seen in **Table 4** and **Figure 4**. There was no observed statistically significant benefit of exercise on LS BMD (g = 0.10 [-0.07, 0.26], Z = 1.15 p = 0.25), with minor asymmetry (LFK index= 1.07, **Supplementary Figure 2**) observed. In addition, no statistically significant heterogeneity was observed (Q = 11.25, p = 0.34) and inconsistency was considered very low (I^2 = 11.12%). Findings were similar when results were collapsed so that only one effect size represented each study. With the removal of each effect size from the model once the results remained non-significant. An evidence profile for changes in LS BMD is shown in online **Supplementary Table 4**. As can be seen, the outcome was considered

DOI: STORK.XXX.XXXX

critical and the overall strength of findings were considered high, with future additional studies unlikely to influence the overall direction of findings.

Table 4 Changes in primary and secondary outcomes

Variable	ES (#)	Participants (#)	Mean (95% CI)	Z (p)	Q (p)	I^2 (%)	LFK Index
Primary [a]							
Femoral Neck	17	801	0.21 (0.03, 0.40)	**2.23 (0.03)***	20.66 (0.19)	22.55	0.43 (none)
Lumbar Spine	11	736	0.10 (-0.07, 0.26)	1.15 (0.25)	11.25 (0.34)	11.12	1.07 (minor)
Secondary [b]							
BMI (kg/m^2)	3	103	-0.28 (-0.90, 0.35)	-0.83 (0.39)	0.19 (0.91)	0	-3.09 (major)
Body Mass (kg)	3	103	-0.31 (-0.99, 0.38)	-0.88 (0.38)	1.32 (0.52)	0	-5.39 (major)
Fat Mass (kg)	4	409	-0.27 (-0.75, 0.20)	-1.12 (0.26)	0.08 (0.99)	0	5.05 (major)
Lean Mass (kg)	4	497	-0.12 (-0.39, 0.15)	-0.85 (0.39)	1.10 (0.95)	0	-2.46 (major)

[a] Outcomes are reported as standardized effect size (g), [b] Outcomes reported as weighted mean difference (WMD); ES, effect size; #, number; participants (#), number of exercise and control participants nested within ES's and studies; Z(p), z-score and alpha value; Q(p), Cochran's Q statistic and alpha value; I^2 (%), I-squared.

* Statistically significant ($p<0.05$).

Changes in Secondary Outcomes

Changes in BMI

Changes in BMI can be seen in **Table 4** and **Figure 5**. There was no observed statistically significant benefit of exercise on BMI (WMD = -0.28 [-0.90, 0.35], Z = -0.83 p = 0.39), with major asymmetry (LFK index= -3.09), **Supplementary Figure 3**) observed. In addition, no statistically significant heterogeneity was observed (Q = 0.19, p = 0.91) and overall inconsistency was considered to be non-existent (I^2 = 0%). Findings were similar when results were collapsed so that only one effect size represented each study. With the removal of each effect size from the model once the results remained non-significant. An evidence profile for changes in BMI are shown in online **Supplementary Table 5**

DOI: STORK.XXX.XXXX

Changes in Body Mass

Changes in BM can be seen in **Table 4** and **Figure 6**. There was no observed statistically significant benefit of exercise on BM (WMD= -0.31 [-0.99, 0.38], Z = --0.88, p = 0.38), with major asymmetry (LFK index= -5.39, **Supplementary Figure 4**) observed. In addition, non-significant heterogeneity was observed (Q = 1.32, p = 0.52) and inconsistency was considered to be *non-existent* (I^2 = 0%). Findings were similar when results were collapsed so that only one effect size represented each study. With the removal of each effect size from the model once the results remained non-significant. An evidence profile for changes in BM are shown in online **Supplementary Table 6.**

Changes in Fat Mass

Changes in FM can be seen in **Table 4** and **Figure 7**. There was no observed statistically significant benefit of exercise on FN BMD (WMD = -0.27 [-0.75, 0.20], Z = -1.13 p = 0.26), with major asymmetry (LFK index= 5.05, **Supplementary Figure 5**) observed. In addition, non-significant heterogeneity was observed (Q =0.08, p = 0.99 and overall inconsistency was considered to be non-existent (I^2 = 0%). Findings were similar when results were collapsed so that only one effect size represented each study. With the removal of each effect size from the model once the results remained non-significant. An evidence profile for changes in FM are shown in online **Supplementary Table 7.**

Changes in Lean Mass

Changes in LM can be seen in **Table 4** and **Figure 8**. There was no observed statistically significant benefit of exercise on LM (WMD= -0.12 [-0.39, 0.15], Z = -0.85, p = 0.39), with major asymmetry (LFK index= -2.46, **Supplementary Figure 6**) observed. In addition, non-significant heterogeneity was observed (Q = 1.10, p = 0.95) and overall inconsistency was considered to be non-existent (I^2 = 0%). With the removal of each effect size from the model once the results remained non-significant. An evidence profile for changes in LM is shown in online **Supplementary Table 8.**

DOI: STORK.XXX.XXXX

DISCUSSION

The primary aim of the current systematic review and meta-analysis was to update the work by Kelley *et al.,*[10] examining the effects of exercise on BMD in men. The overall findings suggest that exercise is associated with a statistically significant benefit in FN but not LS BMD. The effect observed in FN BMD is smaller, yet still significant, than the previously reported effect by Kelley *et al.*[10]. However, due to the greater number of effect sizes included and the use of a more robust meta-analytical methodology, the observed effect in this study strengthens the evidence for the benefit of exercise on FN BMD in men. Based on GRADE, it was concluded that further research is unlikely to change the direction of the effect (**Supplementary Table 2**). These findings are different to the recently published meta-analysis[51], however Ashe *et. al.* did not use the more robust IVhet model, nor was their analysis preregistered as should be the gold standard for systematic review and meta-analysis[52]. Within their study, there was no statistically significant effect of exercise across any of the outcome measures associated with BMD, and unlike this analysis the authors suggest that the evidence would benefit from additional research, whereas our effect appears to be robust without any additional effect sizes added to the model.

While the current study strengthens the evidence in support of a positive effect of exercise on FN BMD, it does not strengthen the evidence for the same change in LS BMD. Kelley *et al.* describe a trend for a small statistically beneficial effect on LS BMD that was not observed in this study. The pooled effect size in the previous work was 50% higher than was observed in this updated analysis, and the greater number of studies included and the more robust methods in this analysis suggests that this effect is unlikely to be influenced by additional research. When we consider the interventions included in this systematic review and meta-analysis, the exercise(s) selected where predominantly lower limb activity which would be unlikely to cause adaptation to the lumbar spine (**Table 2**). This lack of observed effect in LS BMD suggests that the magnitude of strain being elicited by these interventions is insufficient to cause adaptation, even if these include activity in a standing position or ground reaction force.

DOI: STORK.XXX.XXXX

The findings for FN BMD were sensitive to influence analysis, where the effect remained significant on the removal of all effect sizes except for the Kukuljan *et al.*[42] exercise and milk sub-group, and the Kukuljan *et al.*[42] exercise only sub-group (**Supplementary Table 1**). When these groups were removed from the model separately, the overall effect only displayed a trend towards statistical significance, with the 95% confidence interval for the mean effect marginally crossing zero for both sub-groups. However, this is different from the observations of Kelley *et al.*[10] where Zeilman III[46] appeared to have the biggest influence on the FN BMD analysis. The increased influence of the Kukuljan *et al.*[42] exercise and milk subgroup was also observed in the LS BMD analysis of Kelley *et al.*[10]. The authors suggested that this may be due to the comparison with a milk only control group as opposed to the other studies which included comparisons with a non-intervention control group. However, due to the marginal nature of this change in outcome, combined with the more robust IVhet model and increased number of effect sizes included, the current analysis helps address the instability of the results described in the original analysis. For LS BMD, the calculated prediction interval of the original meta-analysis[10] included zero, suggesting no overall effect. Similarly, in the current analysis, the Kukuljan *et al.*[42] exercise and milk subgroup was also reported to have a large influence on the outcome. However, in this larger updated analysis, where the additional effect sizes represent an additional 616 participants, there was no influence of any of the included studies on the outcome, with no effect of the interventions on LS BMD observed.

In addition to the observed effects on BMD, there were no changes in any of the secondary outcomes (BMI, LM, FM, BM) analysed in this study, suggesting that the changes in BMD are occurring independently of any other changes in body composition. Kelley *et al.*[10] previously reported a trend for improved BMI, however, there were no additional effect sizes included in this updated analysis, suggesting that the current findings may be more robust given that the IVhet versus traditional random-effects model was used[19].

The major strength of this new analysis, in addition to the inclusion extra participants with studies published since 2011, is the use of the IVhet model. This more robust method of

DOI: STORK.XXX.XXXX

meta-analysis provides stronger evidence than the previous analysis and importantly shows that the trend towards significance observed with LS BMD is no longer seen with more effect sizes included. A limitation with this analysis, as with all meta-analysis is the likelihood of observing Simpsons Paradox[53], where the outcome is similar in sub-groups of a population but then reversed in the population as a whole. This is often due in meta-analysis of RCTs to imbalance between experimental and control conditions, especially where one control group acts for multiple intervention groups, where this leads to a control group approximately half the size of the intervention group. This occurred in minimal studies included within this analysis and the lack of change when influence analysis was conducted indicates that we are unlikely to be observing an ecological effect.

The previous study *suggested* that men at risk of osteoporosis should take part in exercise based on previous guidelines by the American College of Sports Medicine (ACSM[54]) and the Endocrine Society[55]. However, due to the strengthening of the evidence from the current analysis, it appears that exercise should now be *recommended* in this population based on more robust analyses. The recently updated physical activity guidelines from the WHO[56] and the ACSM recommends that 150 minutes of moderate-intensity aerobic activity and two muscle-strengthening sessions per week[54] would be sufficient for positive changes in FN BMD, provided that the intervention lasts at least 6 months and is continued.

CONCLUSIONS

This analysis provides additional up-to-date evidence to recommend ground- and/or joint-reaction force exercises for improving or maintaining FN but not LS BMD in men. Additional well-designed RCTs are unlikely to alter this evidence, although interventions that include activities that directly load the LS are needed to ensure that this is not a potential method of improving LS BMD.

CONTRIBUTIONS

DOI: STORK.XXX.XXXX

Conceptualization, FMG and KS; methodology, BRH, FMG and GK; writing--original draft preparation, BRH and FMG; writing--review and editing, ALL.

FUNDING INFORMATION
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors

ABBREVIATIONS
The following abbreviations are used in this manuscript:
BMD: Bone Mineral Density
DPA: Dual-energy Photon Absorptiometry
DXA: Dual-energy X-ray Absorptiometry
ES: effect size
FN: Femoral Neck
g: Hedges standardised mean difference effect size
GRADE: Grading of Recommendations Assessment, Development and Evaluation
IVhet: Inverse heterogeneity
LFK: Luis Furuya-Kanamori
LS: Lumbar Spine
NNS: Number-needed-to-screen
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
pQCT: peripheral Quantified Computer Tomography
RCT: Randomised Controlled Trial
SD: Standard Deviation

COMPETING INTERESTS
There are no competing interests for any author.

DATA SHARING
An Open Science Framework project entitled The effects of exercise on bone mineral density in men: a systematic review and meta-analysis of randomised controlled trials with all materials can be found here DOI 10.17605/OSF.IO/E6W3V **[40].**

REFERENCES
1. Unluhizarci K. Osteoporosis: Unawareness or Ignorance? Erciyes Medical Journal. 2019;41(1):1-3.

2. Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. European journal of rheumatology. 2017;4(1):46.

3. Center JR, Nguyen TV, Schneider D, Sambrook PN, Eisman JA. Mortality after all major types of osteoporotic fracture in men and women: an observational study. The Lancet. 1999;353(9156):878-82.

4. Feldstein A, Elmer PJ, Orwoll E, Herson M, Hillier T. Bone mineral density measurement and treatment for osteoporosis in older individuals with fractures: a gap in

evidence-based practice guideline implementation. Archives of internal medicine. 2003;163(18):2165-72.

5. Reginster J-Y, Burlet N. Osteoporosis: a still increasing prevalence. Bone. 2006;38(2):4-9.

6. Forsen L, Søgaard A, Meyer H, Edna T-H, Kopjar B. Survival after hip fracture: short- and long-term excess mortality according to age and gender. Osteoporosis international. 1999;10(1):73-8.

7. Haentjens P, Magaziner J, Colón-Emeric CS, Vanderschueren D, Milisen K, Velkeniers B, et al. Meta-analysis: excess mortality after hip fracture among older women and men. Annals of internal medicine. 2010;152(6):380-90.

8. Holt G, Smith R, Duncan K, Hutchison J, Gregori A. Gender differences in epidemiology and outcome after hip fracture: evidence from the Scottish Hip Fracture Audit. The Journal of bone and joint surgery British volume. 2008;90(4):480-3.

9. Sterling RS. Gender and race/ethnicity differences in hip fracture incidence, morbidity, mortality, and function. Clinical Orthopaedics and Related Research®. 2011;469(7):1913-8.

10. Kelley GA, Kelley KS, Kohrt WM. Exercise and bone mineral density in men: a meta-analysis of randomized controlled trials. Bone. 2013;53(1):103-11.

11. Khan K, McKay H, Kannus P, Wark J, Bailey D, Bennell K. Physical activity and bone health: Human Kinetics; 2001.

12. Barry DW, Kohrt WM. BMD decreases over the course of a year in competitive male cyclists. Journal of Bone and Mineral Research. 2008;23(4):484-91.

13. Fredericson M, Ngo J, Cobb K. Effects of ball sports on future risk of stress fracture in runners. Clinical Journal of Sport Medicine. 2005;15(3):136-41.

14. Fredericson M, Chew K, Ngo J, Cleek T, Kiratli J, Cobb K. Regional bone mineral density in male athletes: a comparison of soccer players, runners and controls. British journal of sports medicine. 2007;41(10):664-8.

15. Nagle KB, Brooks MA. A systematic review of bone health in cyclists. Sports health. 2011;3(3):235-43.

DOI: STORK.XXX.XXXX

16. Kujala UM, Kaprio J, Kannus P, Sarna S, Koskenvuo M. Physical activity and osteoporotic hip fracture risk in men. Archives of Internal Medicine. 2000;160(5):705-8.

17. Bolam KA, Skinner TL, Jenkins DG, Galvao DA, Taaffe DR. The Osteogenic Effect of Impact-Loading and Resistance Exercise on Bone Mineral Density in Middle-Aged and Older Men: A Pilot Study. Gerontology. 2015;62(1):22-32.

18. Allison SJ, Folland JP, Rennie WJ, Summers GD, Brooke-Wavell K. High impact exercise increased femoral neck bone mineral density in older men: a randomised unilateral intervention. Bone. 2013 Apr;53(2):321-8.

19. Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. Contemporary clinical trials. 2015;45:130-8.

20. Furuya-Kanamori L, Thalib L, Barendregt JJ. Meta-analysis in evidence-based healthcare: a paradigm shift away from random effects is overdue. International Journal of Evidence-Based Healthcare. 2017;15(4):152-60.

21. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. Annals of internal medicine. 2017;167(4):268-74.

22. Furuya-Kanamori L, Barendregt JJ, Doi SA. A new improved graphical and quantitative method for detecting bias in meta-analysis. International journal of evidence-based healthcare. 2018;16(4):195-203.

23. Garner P, Hopewell S, Chandler J, MacLehose H, Akl EA, Beyene J, et al. When and how to update systematic reviews: consensus and checklist. bmj. 2016;354:i3507.

24. Sacks H, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. The American journal of medicine. 1982;72(2):233-40.

25. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273(5):408-12.

DOI: STORK.XXX.XXXX

26. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med. 2009;6(7):e1000097.

27. Hamilton BR, Staines KA, Kelley GA, Kelley KS, Kohrt WM, Pitsiladis Y, et al. The effects of exercise on Bone Mineral Density in Men: a protocol for a systematic review and meta-analysis of randomised controlled trials. SportRxiv; 2020.

28. Hamilton BR, Staines K, Kelley GA, Kelley KS, Kohrt WM, Pitsiladis YP, et al. The effects of exercise on bone mineral density in men: a systematic review and meta-analysis of randomised controlled trials. 2020 [cited; Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020180441

29. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychology. Bulletin, 70, 213â. 1968;220.

30. Lee E, Dobbins M, DeCorby K, McRae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. BMC medical research methodology. 2012;12(1):51.

31. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. bmj. 2019;366.

32. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. Physical therapy. 2003;83(8):713-21.

33. Smart NA, Waldron M, Ismail H, Giallauria F, Vigorito C, Cornelissen V, et al. Validation of a new tool for the assessment of study quality and reporting in exercise training studies: TESTEX. International journal of evidence-based healthcare. 2015;13(1):9-18.

34. Ahn S, Becker BJ. Incorporating quality scores in meta-analysis. Journal of Educational and Behavioral Statistics. 2011;36(5):555-85.

35. Hedges LV, Olkin I. Statistical methods for meta-analysis: Academic press; 2014.

36. Follmann D, Elliott P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. Journal of clinical epidemiology. 1992;45(7):769-73.

37. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003;327(7414):557-60.

38. Furuya-Kanamori L, Xu C, Lin L, Doan T, Chu H, Thalib L, et al. P value–driven methods were underpowered to detect publication bias: analysis of Cochrane review meta-analyses. Journal of clinical epidemiology. 2020;118:86-92.

39. Schünemann H, Brożek J, Guyatt G, Oxman A. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from guidelinedevelopment org/handbook. 2013.

40. Hamilton BR, Staines K, Kelley GA, Kelley KS, Kohrt WM, Pitsiladis YP, et al. The effects of exercise on bone mineral density in men: a systematic review and meta-analysis of randomised controlled trials. 2021 [cited 2021 March 6th]; Available from: https://osf.io/e6w3v/

41. Hong WL. Tai Chi and resistance training exercise: would these really improve the health of the elderly?: The Chinese University of Hong Kong (Hong Kong); 2004.

42. Kukuljan S, Nowson CA, Sanders KM, Nicholson GC, Seibel MJ, Salmon J, et al. Independent and combined effects of calcium-vitamin D3 and exercise on bone structure and strength in older men: an 18-month factorial design randomized controlled trial. The Journal of Clinical Endocrinology & Metabolism. 2011;96(4):955-63.

43. Helge EW, Andersen TR, Schmidt JF, Jorgensen NR, Hornstrup T, Krustrup P, et al. Recreational football improves bone mineral density and bone turnover marker profile in elderly men. Scand J Med Sci Sports. 2014 Aug;24 Suppl 1:98-104.

44. Harding AT, Weeks BK, Lambert C, Watson SL, Weis LJ, Beck BR. Effects of supervised high-intensity resistance and impact training or machine-based isometric training on regional bone geometry and strength in middle-aged and older men with low bone mass: The LIFTMOR-M semi-randomised controlled trial. Bone. 2020 Apr 11;136:115362.

45. Newton RU, Galvao DA, Spry N, Joseph D, Chambers SK, Gardiner RA, et al. Exercise Mode Specificity for Preserving Spine and Hip Bone Mineral Density in Prostate Cancer Patients. Med Sci Sports Exerc. 2019 Apr;51(4):607-14.

DOI: STORK.XXX.XXXX

46. Zeilman III CJ. Inflammatory bowel disease, osteoporosis, exercise, and bone mineral density. University of Florida. 2007;56.

47. Uth J, Hornstrup T, Christensen JF, Christensen KB, Jorgensen NR, Schmidt JF, et al. Efficacy of recreational football on bone health, body composition, and physical functioning in men with prostate cancer undergoing androgen deprivation therapy: 32-week follow-up of the FC prostate randomised controlled trial. Osteoporos Int. 2016 Apr;27(4):1507-18.

48. Bjerre ED, Jørgensen AB, Petersen TH, Eriksen AR, Midtgaard J, Krustrup P, et al. Football Compared with Usual Care in Men with Prostate Cancer (FC Prostate Community Trial): A Pragmatic Multicentre Randomized Controlled Trial. Sports Medicine. 2019;49(1):145-58.

49. Kemmler W, Kohl M, Frohlich M, Jakob F, Engelke K, von Stengel S, et al. Effects of High-Intensity Resistance Training on Osteopenia and Sarcopenia Parameters in Older Men with Osteosarcopenia-One-Year Results of the Randomized Controlled Franconian Osteopenia and Sarcopenia Trial (FrOST). J Bone Miner Res. 2020 Apr 9.

50. Kim SH, Seong DH, Yoon SM, Choi YD, Choi E, Song Y, et al. The Effect on Bone Outcomes of Home-based Exercise Intervention for Prostate Cancer Survivors Receiving Androgen Deprivation Therapy: A Pilot Randomized Controlled Trial. Cancer Nurs. 2018 Sep/Oct;41(5):379-88.

51. Ashe MC, Santos IKd, Edward NY, Burnett LA, Barnes R, Fleig L, et al. Physical Activity and Bone Health in Men: A Systematic Review and Meta-Analysis. Journal of Bone Metabolism. 2021;28(1):27-39.

52. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. Systematic reviews. 2012;1(1):1-9.

53. Rücker G, Schumacher M. Simpson's paradox visualized: the example of the rosiglitazone meta-analysis. BMC Medical Research Methodology. 2008;8(1):1-8.

DOI: STORK.XXX.XXXX

54. Bloomfield SA, Little K, Nelson M, Yingling V. American College of Sports Medicine position stand: physical activity and bone health. Med Sci Sports Exerc. 2004;195(9131/04):3611.

55. Watts NB, Adler RA, Bilezikian JP, Drake MT, Eastell R, Orwoll ES, et al. Osteoporosis in men: an Endocrine Society clinical practice guideline. The Journal of Clinical Endocrinology & Metabolism. 2012;97(6):1802-22.

56. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. British journal of sports medicine. 2020;54(24):1451-62.

FIGURE LEGENDS

Figure 1. PRISMA Flow Chart for Study Selection

Figure 2. Risk of Bias Assessment using the RoB 2 assessment tool.

Figure 3 Forest plot for changes in FN BMD. Forest plot for point estimate standardized effect size changes (g) in FN BMD. The black squares represent the standardized mean difference (g) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall standardized mean difference (g) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

Figure 4 Forest plot for changes in LS BMD. Forest plot for point estimate standardized effect size changes (g) in FN BMD. The black squares represent the standardized mean difference (g) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall standardized mean difference (g) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

Figure 5 Forest plot for changes in BMI. Forest plot for point weighted mean difference (WMD) in BMI. The black squares represent the weighted mean difference (WMD) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall weighted mean difference (WMD) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

Figure 6 Forest plot for changes in Body Mass. Forest plot for point weighted mean difference (WMD) in Body Mass. The black squares represent the weighted mean difference (WMD) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall weighted mean difference

(WMD) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

Figure 7 Forest plot for changes in Fat Mass. Forest plot for point weighted mean difference (WMD) in Fat Mass. The black squares represent the weighted mean difference (WMD) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall weighted mean difference (WMD) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

Figure 8. Forest plot for changes in Lean Mass. Forest plot for point weighted mean difference (WMD) in Fat Mass. The black squares represent the weighted mean difference (WMD) while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall weighted mean difference (WMD) while the left and right extremes of the diamond represent the corresponding 95% confidence intervals.

DOI: STORK.XXX.XXXX