Accepted Article

Journal of Frailty, Sarcopenia and Falls

Review Article

The effect of weight-bearing exercise on the mechanisms of bone health in young females: A systematic review

Tommy J. Cartledge, Joey Murphy, Charlie E. Foster, Byron Tibbitts

Centre for Exercise, Nutrition & Health Sciences, School for Policy Studies, University of Bristol, Bristol, United Kingdom

Abstract

Weight-bearing exercise (WBE) has been identified as an appropriate approach for increasing peak bone mass, however, there is a lack of specific physical activity recommendations in this area. Thus, the aim of this systematic review is to determine the optimal mode of WBE, specifically identifying the intensity, duration, frequency, and load, to elicit the optimal effect on bone mass in young females, aged 5-18. A literature search was conducted from the 28th of June to the 20th of July 2021 using PubMed/Medline, Web of Science and SPORTDiscus. The search produced 1405 results, of which 15 were deemed appropriate for inclusion. The majority of studies (n=12) found a significant positive effect for at least one bone measure through their respective WBE exposure (p<0.05). Bone mass accrual was found to be site-specific depending on WBE exposure type, load, and maturity status. Also, longitudinal effects on bone mass accrual were found exclusively in gymnastics participants, even if participation level decreased (i.e., retirement). The results of this study support the use of WBE to improve parameters of bone health. However, further research is needed as the optimal mode of WBE to elicit the optimal effect on bone mass is still unclear.

Keywords: Female, Healthy ageing, Paediatric, Peak bone mass, Physical activity

Introduction

Osteoporosis is a progressive skeletal disease characterised by a reduction of bone mineral density (BMD) and micro-architectural deterioration of bone tissue, consequently resulting in increased bone fragility[1]. The disease is considered a major health concern as approximately 3 million individuals in the UK are osteoporotic, estimated to cost £4.4 billion to treat, annually[2]. The prevalence of osteoporosis is disproportionately observed in older individuals (>50 years of age), particularly postmenopausal women, who account for approximately 75% of all osteoporotic cases[3]. A major consequence of osteoporosis is the increase in fragility fracture risk, with common sites including the vertebral spine, hip, and wrist[4]. Depending on the severity, a fragility fracture can be responsible for reduced quality of life, excess morbidity and even mortality[5]. In the UK, fragility fracture incidence has been estimated to double over the next 50 years, due to an ageing population[6]. Previous research shows that increasing bone mass at a younger age can help reduce fracture risk[7]. This indicates a need to implement appropriate and effective practices to reduce the high prevalence of osteoporotic cases in the female population and reduce fragility fracture risk.

Bone mass is observed to increase greatly during childhood and adolescence, peak during young adulthood, and decline throughout older age[8]. Thus, one main method to reduce fragility fracture risk is through the optimisation of peak bone mass (PBM) during childhood and adolescence (prepubertal [Tanner I] to pubertal [Tanner V])[9]. PBM can be characterised as the maximum volume of bone mass present at the end of bone maturation, which commonly occurs around 25 to 30 years of age[10,11]. It has been previously found that even a 10% increase in PBM can reduce the

The authors have no conflict of interest.

Corresponding author: *Tommy J. Cartledge, 15 Shaston Crescent, Dorchester, DT1 2EB, Dorset, 07714736021, UK*

E-mail: *tommyjc@btinternet.com*

Edited by: *Jagadish K. Chhetri*

Accepted *6 June 2022*

risk of hip fractures by approximately 30%[7]. As a result of the higher prevalence of osteoporosis and higher fragility fracture risk in women[12,13], this systematic review will solely focus on PBM development in children and adolescent girls aged 5-18 years. This age range has been chosen as 90% of PBM is reached by the age of 18 in females[10].

Physical activity (PA) is recognised as the primary approach for aiding skeletal development and PBM optimisation in children and adolescents[14,15]. Specifically, the most effective form of PA is weight-bearing exercise (WBE)[8,16]. WBE can be defined as any force-generating exercise that elicits a load to a skeletal region higher than that of daily living[8], for example, jogging, jumping, trampolining or tennis. Evidence is abundant in support of WBE for increased bone accrual. A prospective study[17] found prepubescent gymnasts had increased BMD by 30-85% compared to an inactive control group (CG). Additionally, a systematic review of 19 randomised control trials (RCT) identified WBE increased mean bone health parameters, over 6 months, by 0.9-4.9% in prepubescent individuals, 1.1-5.5% in early pubescent individuals and 0.3-1.9% in pubescent individuals[8]. A recent systematic review[16] also found drop-jumping interventions to have a beneficial effect on BMD, bone mineral content (BMC) and bone structure, without showing any adverse events (i.e., injury). The positive effects of WBE are also evident in longitudinal studies, for example, Luiz-de-Marco et al. (2020)[18] found adolescents who performed different forms of WBE (e.g., tennis, gymnastics, karate), over 12 months, had greater improvements in areal BMD of the lower body when compared to swimming, likely due to the buoyancy or weightless state when in water. The use of WBE for increased bone accrual is also supported by Wolff's Law[19], which states bone density changes will occur in response to repeated mechanical loading. The law proposes, bone will adapt accordingly to tolerate only the loads to which it is subjected. Hence, exposing bone to higher loads will increase bone strength and better equip them for such loads.

Although there is sufficient evidence to support the use of WBE for the increase in PBM in children and adolescents[8,16], a gap in the current knowledge base arises when determining the optimal intensity, volume and load of a given WBE, as well as the optimal form of WBE for the improvement of PBM. This is evident from the broadness of recommendations in PA guidelines for children and adolescents. For instance, the most recent UK Chief Medical Officers' (CMO) PA guidelines state that, for the improvement of muscle and bone strength, children should engage in a range of physical activities at different intensities. However, these guidelines do not suggest specific intensities, volumes or loads for specific types of PA[20]. The same can be said for the American College of Sports Medicine (ACSM) guidelines, though there is only mention of PA for the improvement of muscle strength, not bone strength[21].

Due to the lack of clarity in PA recommendations for the improvement of bone health and the high prevalence of osteoporosis in postmenopausal women, this systematic review aims to summarise the available literature to explore the optimal type, intensity, duration, frequency, and load of WBE to elicit the optimal effect on bone health (i.e., BMD, BMC and PBM) for children and adolescent girls aged 5 to 18 years of age. Follow-up questions for this review aim to include how do the effects of WBE: 1) differ at the same intensity, duration and load for different girls of the same age; 2) differ at different ages throughout childhood and adolescents for girls; 3) maintain over time for children and adolescent girls? In answering these questions, this review will bring new insight into the creation of evidence-based WBE programmes and PA guidelines for young girls aged 5 to 18.

Methodology

The current systematic review followed the methodological approach proposed by the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist[22] (see Table 1). Furthermore, ethics approval was granted by the CENHS Research Ethics Committee: approval number EAN 008-21.

Search Strategy

A search for relevant literature was conducted using three scientific databases: PubMed (including Medline), Web of Science and SPORTDiscus. The search was performed from the 28th of June to the 20th of July 2021 and included published literature from 2006 onwards, due to the latest systematic review examining studies up to 2005[8]. The following search terms and Boolean operators were used for all three databases: ('exercise' OR 'physical activity') AND ('children' OR 'adolescent') AND ('girl') AND ('bone density' OR 'bone mineral' OR 'bone mass'). The search was completed by the lead author and all titles were downloaded for screening.

Study Selection

After removing duplicates and articles deemed not to have a relevant title, 189 full article abstracts were screened using the following inclusion criteria: (1) used a RCT/cluster randomised trial (CRT) or non-RCT experimental design; (2) examined a WBE intervention for weight-bearing exercise/sport; (3) examined BMD or BMC as a primary measure via dual-energy X-ray absorptiometry (DEXA) scanning; (4) included at least one of the following bone sites: total body (TB), lumber spine (LS), femoral neck (FN), trochanter (TR) or forearm; and (5) examined healthy children and adolescent girls aged 5 to 18 years. Furthermore, articles were excluded if: (1) published in languages other than English; (2) used unpublished data; (3) only focused on bone metabolic markers; (4) examined boys only or published data not separated for gender; (5) no explanation of intervention programme; (6) participants

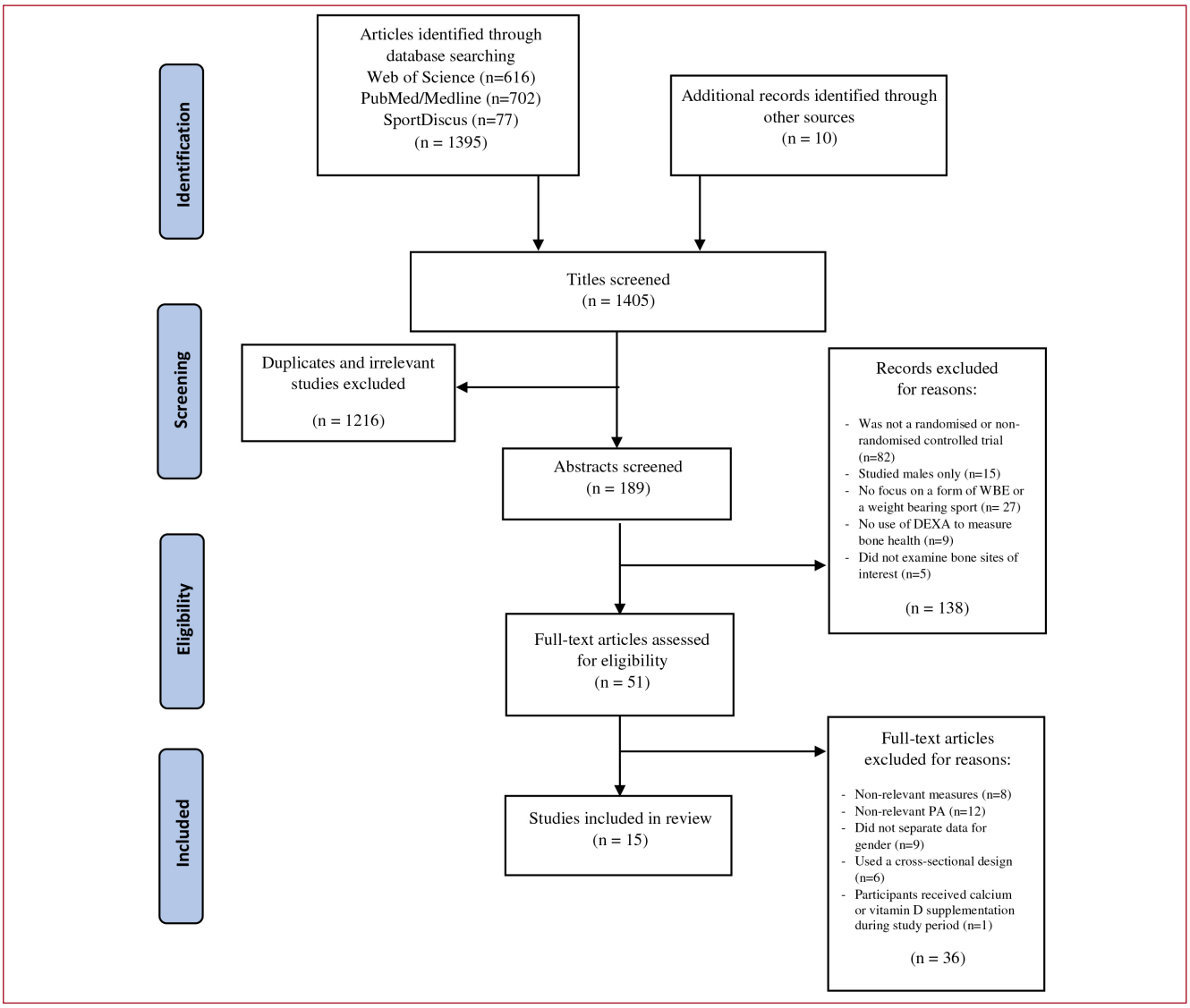

Figure 1. Detailed PRISMA flow diagram of the literature search process.

received calcium or vitamin D supplementation; (7) no indication of pubescent status; and (8) published in the year 2005 or earlier. Screening was performed by the lead author (TC) with two additional authors (CF and JM) independently screening 10% of the identified titles and abstracts. Inter-rater reliability was calculated between the three reviewers and a high level of agreement was found (98.3%). Further discussions were had to ensure clarity regarding discrepancies. This led to the first author being capable of reviewing the remaining article titles and abstracts. A 10% value was set as a result of a review by McDonagh et al (2013)[23], stating it to be an appropriate figure for screening at the title and abstract stage when conducting a dual review.

Data Extraction

An adapted version of the Cochrane data collection form for intervention reviews of RCTs and non-RCTs was used to gain a greater insight into the selected studies and guide data extraction[24]. This tool was used to extract and present the descriptive characteristics of each study, enabling the current systematic review to meet the Methodological Expectations of Cochrane Intervention Review's (MECIR) standards for collecting and reporting information from studies and analysing their results[24].

Outcome measures

The primary outcome variable was the change in BMD (g/cm^2) or BMC (g) across different bone sites, including TB,

LS, FN, GT and the forearm. For data grouping, it has been found that an increase in BMC may not necessarily indicate an increase in BMD and differences in levels of bone acquisition have been found between the two variables[25], hence BMC and BMD were examined separately. Furthermore, as differences in BMC and BMD can vary by location[26], bone sites were examined separately, also.

To determine the effects of moderator variables on BMD and BMC on the different bone sites, the following variables were extracted from selected studies: pubertal status, WBE mode, intervention strategy, WBE intensity, WBE duration, WBE frequency, and programme length. This review classified pubertal status as either prepubertal (Tanner stage I), early pubertal (Tanner stages II and III), or pubertal (Tanner stages IV and V)[9]. As pubertal status may change throughout the course of a given study, pubertal classification was defined as the pubertal status reported at baseline.

Quality Assessment

Reporting bias assessment

To measure the risk of bias in the included RCT studies, the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) was used[27]. This tool evaluates five domains of bias in an RCT, including: (1) random sequence generation; (2) deviations from the intended intervention, for both the effect of assignment and adherence; (3) missing outcome data; (4) measurement of outcome; and (5) selection of reported data.

The risk-of-bias in non-randomised studies of interventions (ROBINS-I) was used to report the risk of bias for the included non-RCT studies[28]. The ROBINS-I evaluates the risk of bias through 7 domains, including: (1) bias due to confounding; (2) selection of participants into the study; (3) classification of interventions; (4) deviations from intended interventions; (5) missing data; (6) measurement of outcomes; and (7) selection of reported result.

Quality of reporting analysis

To assess the methodological quality, the modified Jadad score was used as it assesses more details of methodological quality than solely randomization, blinding, and withdrawals/dropouts[29,30]. For 6 of the 8 items, an answer of 'yes' equals 1 point and an answer of 'no' equals 0 points; however, for 2 items an answer of 'no' equals minus 1 point, 'yes' equals one point and 'not described' equals 0 points. For interpretation, a score of 4-8 indicates good to excellent quality and a score of 0-3 indicates poor to low quality. It is important to note that the modified Jadad tool is designed for assessing RCTs. This tool is usable for assessing non-RCTs, however results in them being a lower level of quality.

Results

Overview of studies

The search produced 1395 items, which increased to 1405 through backward reference searching of topic-relevant primary literature, systematic reviews and meta-analyses. The total was reduced by 1216 items after duplicates were removed and titles were screened. A further 138 items were removed through abstract screening due to an incorrect study design, type of physical activity, study population or measure.

The remaining 51 items were read in full with a further 36 items excluded. In total, 15 studies[31-45] were included for analysis. Figure 1 provides further detail regarding the study selection process.

Study characteristics

Studies were numbered from 1-15 to assist reporting and grouped by the mode of WBE exposure. A detailed summary of key characteristics and findings from each study is outlined in Table 2. Of the included studies, five were conducted in Australia [2, 3, 5, 6, 10], four in the USA [7, 8, 13, 15], two in Brazil [9, 14], and one in Hong Kong [1], South Africa [4], France [11] and Finland [12]. Regarding study designs, five used a RCT design [2-6] and 10 used a longitudinal study design [1, 7-15]. The mean sample size was 88.4 participants. Four studies examined both adolescent girls and boys [2, 4, 5, 14], however, their data was split for gender. With respect to participants' pubertal status, eight studies examined females at all pubertal stages [1, 5, 7, 9, 11, 12, 14, 15], four studies examined females at prepubertal and early pubertal [2-4, 10] and three studies examined only one pubertal stage [6, 8, 13].

Regarding modes of WBE, all studies used a plyometric-based WBE intervention or sport, with the exception of studies 7 and 8, which both used a weight-bearing resistance-based exercise intervention. The mean intervention/study length was 20 months, with 10 studies being ≤12 months, three studies being between 13-36 months and two studies being >36 months.

Reporting bias outcome

Through completing a risk of bias assessment using the RoB 2 tool, four RCT studies averaged low reporting bias throughout the 5 domains (Table 3). Only two studies had one 'high risk' domain [2, 3], which was for 'missing outcome data'. For study 2, this was a result of a high attrition rate between follow-ups, as from baseline to initial follow-up, 25.3% of participants were lost and from initial follow-up to second follow-up, 45.7% of participants were lost. For study 3, the 'high risk' score was due to only 13% of included participants attending DEXA scanning meetings. Lastly, four studies [3-6] scored 'some risk' for the 'random sequence generation' domain as a result of no information being reported on the concealment of allocation for the randomisation process.

With the exception of studies 1 and 14, all included non-RCT studies either scored 'low risk' [7-9] or 'low-to-moderate risk' of reporting bias [10-13, 15] (Table 4). For both studies 1 and 14, the majority of domain scores

were either 'moderate risk' or 'serious risk' and so were classed as having a moderate-serious risk of reporting bias. It is important to note that the majority of studies scored 'serious risk' on confounding (n=7), mainly for not controlling for habitual food intake or habitual PA. An additional two domains, 'recruitment of participants' and 'classification of intervention', scored 'moderate risk' for the majority of studies (n=8 and n=7, respectively). This was due to the examination of athletes, which gives the potential for a selection bias and an unclear description of the intervention protocol, respectively.

Quality of reporting

Based on the criteria of the modified Jadad score, all included RCT studies scored 'good to excellent' methodological quality (Table 5). Study 2 was the only CRT included; however, due to the study using a statistical model which adjusted for variability both between clusters (schools) and within a cluster (students within the same school), it was deemed appropriate. As predicted, the included non-RCT studies had lower methodological quality compared to the RCT studies, with scores ranging from 2-3, which denote 'poor to low' methodological quality. The methodological quality of a non-RCT study will commonly be lower as the participants are not randomised, however, this was intensified by six non-RCT studies not outlining their exclusion/inclusion criteria or their measure of adverse effects. A common issue in all studies, except study 2, was the non-use of blinding, which could have caused a level of detection bias.

Study results

Jump-based exercise

Of the six studies that examined an in-school jump-based exercise intervention or jump-based activity, three found significant within-group changes in their measured bone outcomes for their respective intervention group (IG) or experimental group (ExG) and CG ([1, 4, 5] all $p<0.05$). The remaining three studies [2, 3, 6] did not measure within-group differences, therefore the changes within groups cannot be reported.

Only studies 1 and 4 found a significant between-group difference in their bone outcomes. Study 1 found participants in an ExG of regular rope skipping had significantly higher levels of calcanei BMD at baseline and follow-up (12-months) compared to a CG of no regular rope skipping, (0.45±0.08 g/cm^2 vs 0.41±0.09 g/cm^2, $p<0.05$ and 0.47±0.08 g/cm^2 vs 0.43±0.08 g/cm^2, $p<0.05$, respectively). However, the increases in calcanei BMD in both the ExG or CG from baseline to follow-up were not significant ($p=0.09$ and $p=0.12$, respectively). A significant increase was also observed for forearm BMD in both the ExG (0.26±0.10 g/cm^2 vs 0.22±0.10 g/cm^2, $p<0.05$) and CG (0.19±0.10 g/cm^2 vs 0.23±0.10 g/cm^2, $p<0.05$) between baseline and follow-up, with the ExG having a significantly higher forearm BMD at follow-up compared to CG (0.26±0.10 g/cm^2 vs 0.23±0.10 g/cm^2, $p<0.05$). A two-level regression analysis was also conducted by study 1, which identified that there was no time effect (B=0.005, $p=0.70$, 95% CI, -0.020 to 0.029), group effect (B=0.012, $p=.26$, 95% CI, -0.009 to 0.032) or group-by-time effect (B=0.003, $p=0.78$, 95% CI, -0.018 or 0.024) for forearm BMD. This suggests that forearm BMD did not significantly increase between baseline and follow-up, differ between groups at follow-up or in terms of changes in outcome, respectively. For calcanei BMD, the analysis found a significant group effect (B=0.023, $p<0.01$, 95% CI, 0.007 to 0.038), therefore participants in the ExG had significantly higher levels of calcanei BMD at follow-up when compared to the CG. Additionally, study 4 found a significant group-by-time effect for total hip (TH) BMC, in favour of the ExG compared to the CG ($p=0.04$, only p-value reported in text). Overall, there is mixed evidence for the beneficial effect of jump-based exercise on bone health, as three out of six studies did not find significant changes. However, of the studies that did find a positive effect on BMD and BMC, the benefits were mainly in the lower limbs, specifically the lower spine, hip and heal.

Resistance-based exercise

Studies 7 and 8 measured the effect of an in-school resistance-based exercise programme on BMD and/or BMC. Study 7 found participants in the ExG had a significantly larger percentage increase in LS BMD and BMC compared to the CG (5.5% and 4.1%, respectively; $p<0.05$), regardless of participation effort. For FN BMD and BMC, the increase observed in the ExG did not reach significance compared to the CG ($p>0.05$). When separating results for participation effort, it was found that high participation effort participants had a significantly greater increase in LS BMD and BMC (+8.2% and +5.7%, $p<0.01$, respectively) and FN BMD and BMC (+7.1% and +6.1%, $p<0.01$, respectively) when compared to the CG. For study 8, there was no significant change in all bone parameters over the course of the exercise programme between the ExG and CG ($p>0.05$). However, a trend toward a positive ExG effect was found for spine BMC ($p=0.05$, $f=0.15$, medium effect), LS BMC ($p=0.10$, $f=0.11$, small effect), proximal femur narrow neck (PFNN) width ($p=0.08$, $f=0.10$, small effect), TB BMC ($p=0.12$, $f=0.08$, small effect) and arm BMC ($p=0.07$, $f=0.12$, small effect). Furthermore, when separating results for Tanner stage (at baseline), the ExG had a significant increase for Tanner II in PFNN width ($p=0.01$) and Tanner III in LS BMD ($p=0.03$) compared to the CG. Collectively, this evidence suggests that resistance-based exercise has site-specific benefits to both BMD and BMC in adolescent girls, which depend on participation level and maturity stage.

Gymnastics

Four studies examined the effect of gymnastics participation on BMD and/or BMC [9-12]. Firstly, study 9 found participants partaking in artistic gymnastics

significantly increased femur BMD in multiple areas compared to a non-regular PA CG; specifically, the wards' triangle by 19% (1.12±0.16 g/cm^2 vs 0.94±0.11 g/cm^2, p<0.01), TR by 14% (0.91±0.12 g/cm^2 vs 0.80±0.12 g/cm^2, p=0.047) and the whole femur by 10% (1.12±0.11 g/cm^2 vs 1.01±0.12 g/cm^2, p=0.046). Spine and TB BMD were also measured; however, no significant differences were found between the ExG and CG (p=0.60 and p=0.12, respectively). Study 10 compared the effect of high or low non-elite artistic gymnastics participation to a non-gymnastics CG. Although the study found no significant group-by-time effect for all groups at all bone outcomes (p>0.05), a significant group effect was found for high gymnastics participation compared to the CG for arm BMC (165.4 g, 95% CI 150.4 to 180.4 vs 153.3 g, 95% CI 141.3 to 155.4, p<0.01). A significant time effect was also found for TB BMD and BMC and arm BMC for all groups (p<0.001). Study 11 found participants partaking in rhythmic gymnastics had significantly higher BMD at baseline and follow-up (12-months) compared to a non-regular PA CG at the following bone sites: TB (p<0.01 and p<0.05, respectively), TR (both p<0.001), FN (both p<0.001) and PR (both p<0.001). Moreover, study 11 found significant within-group increases in BMD at all measured bone outcomes for both rhythmic gymnastics and CG (all p<0.05). Lastly, study 12 examined the effect of competitive gymnasts (discipline not specified) and found FN BMC increased by 4.6% (95% CI 0.41 to 9.1, p<0.05) more for gymnasts compared to the CG during the first follow-up (36-months). During the second follow-up (84-months), FN BMC did not increase in either group, however gymnasts FN BMC was still significantly greater than the CG (p<0.01). A group-by-time effect was significant only at the FN (p=0.048). The mean LS BMC of the gymnasts was higher than the LS BMC of the CG at all time points throughout the study period, however the differences between the two groups at both follow-ups did not reach significance (p>0.05). Values of within-group differences were not measured and so the differences in LS and FN BMC within each group from baseline to final follow-up cannot be reported. Together, this evidence suggests that gymnastics participation, artistic or rhythmic, has beneficial effects on lower and upper limb BMD and BMC gains in adolescent girls compared to non-active individuals. Also, there is evidence that bone mass is increased further with greater gymnastics participation, as well as bone mass gains having a long-lasting effect throughout young adulthood, even if gymnastics participation is reduced (i.e., retirement).

Running

Study 12 also examined the effect of competitive running on FN and LS BMC. The results indicated that the mean increase in LS BMC throughout the study period was similar between the running group and CG, however running participants had a significantly higher LS BMC at follow-up one (57.3±12.2 g vs 51.4±8.4 g, p<0.01) and follow-up two (62.8±10.6 g vs 55.2±8.7 g, p<0.01). This pattern emerged for FN BMC also, at the first follow-up (4.4±0.8 g vs 4.0±0.6 g, p<0.001) and second follow-up (4.4±0.7 g vs 4.1±0.6 g, p<0.01). However, it can be seen there was little to no increase in FN BMC in either group between the first and second follow-up (within-group difference not measured). Study 13 examined the effect of endurance running in participants with either low (1 or 2 standard deviations or more below age and gender matched reference data) or normal bone mass. The study found that both the low and normal bone mass groups had a significant within-group increase in TB and LS BMC from baseline to follow up (all p<0.001). However, only the low bone mass group had a significant increase in TH BMC, with a mean increase of 0.3±0.4 g (p<0.05). Additionally, a significant between group difference was found for all bone sites at baseline and follow-up, in favour of the normal bone mass group (all p<0.01). Collectively, the results from studies 12 and 13 show running exercise can have positive benefits to BMC of adolescent girls, exclusively in the lower spine or hip regions. Moreover, low bone mass occurring at young childhood may be hard to increase to 'normal' or 'health' bone mass during adolescents and early adulthood.

Martial arts

Only study 14 examined the effect of martial arts on BMD, specifically karate/kung fu (grouped together due to similarities) and judo. The study found BMD of the TB and LS significantly increased between baseline and follow-up in both martial arts groups and the CG (p<0.05). However, an analysis of covariance indicated that neither martial arts group had a significant effect on TB or LS BMD changes (p=0.416 and p=0.230, respectively).

Tennis

Study 15 compared the effect of tennis participation on BMC between pre/early pubertal and pubertal girls. The study identified TB BMC to be significantly higher in the pubertal group compared to the pre/early group at baseline (2350.0±320.0 g vs 1750.0±350.0 g, p<0.001) and follow-up (2460.0±310.0 vs 2170.0±330.0, p<0.05). Furthermore, both groups had a significant increase in non-playing arm BMC at follow-up, with the pre/early group experiencing a much larger increase (+19.5%, p<0.001 vs +4.6%, p<0.001).

Discussion

Synthesis of overall findings

The main aim of this systematic review was to determine the optimal mode of WBE to elicit the optimal effect on bone mass in children and adolescent girls aged 5-18, with a focus on intensity, duration, frequency, and load.

The included studies comprised of six different modes of WBE exposure, namely: jump-based (n=6), resistance-based (n=2), gymnastics (n=4), running (n=2), martial arts (n=1) and tennis (n=1). All 15 studies found significant positive

effects to at least one measured bone outcome through their respective WBE exposure, with the exception of studies 2, 3 and 6, which did not find any significant benefits to bone accrual through the use of jump-based exercise. The remaining three jump-based exercise studies, and studies examining running participation, observed benefits exclusively in lower body BMD or BMC (e.g., lower spine, hip and heal). Studies examining the effect of resistance-based exercise found benefits to bone accrual were site-specific depending on the duration and frequency of exercise or maturity stage. Gymnastics participation demonstrated benefits to TB, upper limb and low limb bone accrual, with evidence suggesting that increased participation is linked to further increases in bone mass. Furthermore, evidence from studies 11 and 12 suggested there may be sustained effects of gymnastics participation through late adolescents and into young adulthood, even if participation is reduced or even stopped (i.e., retirement). There was weak evidence for the improvement of bone mass through martial arts participation, however, tennis participation had significant positive effects on participants' TB and arm BMC, with evidence suggesting bone accrual differences depending on maturity stage.

Interpretation of results

Summary of overall risk of bias and reporting quality

For both RCT and non-RCT studies, the majority of the risk of bias domains scored "low risk" (n=23 and n=40, respectively). Furthermore, all RCT studies scored "good to excellent" for reporting quality, with the non-RCT studies scoring "low", which is to be expected. These results infer confidence for their inclusion in this review.

Effects of WBE interventions or training protocols on bone accrual

The results from this review indicate that the mode of WBE exposure has a large effect on the location of bone accrual (i.e., bone site). For instance, jump-based and running exercises increased bone accrual exclusively in the lower body and hip region, respectively. Whereas gymnastics participation increased bone accrual in both upper and lower body bone sites. Therefore, this review supports the notion that bone response to mechanical loading is site-specific[19,46]. Prior cross-sectional and prospective studies have indicated that targeted mechanical loading, through using WBE, increases BMD at skeletal sites which are exposed to the highest levels of strain[47,48]. All bones have the potential to become osteoporotic[49] hence utilising a mode(s) of WBE that can increase PBM at the majority, if not all, bone sites or at bone sites that are most vulnerable to fragility fractures in later adulthood (e.g., LS, FN or wrist)[4], would be most beneficial for reducing the prevalence of osteoporosis and fractures in young girls. However, due to a lack of research in one particular WBE mode and the majority use of non-RCT designs, there is not enough evidence to conclude what the optimal WBE mode is to achieve this.

Evidence from this review may suggest the effects of WBE are load-dependent. also. Study 11 identified that gymnastics participation induced beneficial effects to bone mass at bone sites that are exposed to high and repetitive mechanical loads, such as the proximal femur. However, very low loading activity, swimming and a leisure activity control, only induced beneficial effects to bone mass at bone sites that are less prone to high mechanical loads, such as the LS and the radius. This supports evidence that only low levels of strain are needed to see beneficial bone adaption[50,51]. Studies 9 and 12 found beneficial effects to FN bone mass accrual in gymnastics athletes, but not for a non-active control and running athletes, respectively. It has been found that peak impact loads during gymnastics participation can reach 10 to 18 times bodyweight, hence the impact loads can be very high[52]. Despite this, neither study measured the exercise impact loads the gymnastics participants were exposed to throughout the study period, and so a definitive conclusion regarding the effect of load on bone accrual cannot be made. This limitation spans all included studies, except for study 6, which measured peak ground-reaction impact forces (PGRIF) in subsamples from each intervention group (n=3). This study found no benefits to bone mass accrual, compared to a control, after a 7-month single-leg drop landing intervention from 14 cm or 24 cm, which exerted PGRIF between 2.7–2.5–3.5 and 2.9–3.8 times bodyweight, respectively. Only one comparable prior study was identified[53], which found a positive bone mass accrual effect through single-leg drop-jumps that exerted a PGRIF of 4.25 times bodyweight, thus slightly larger than study 6. However, due to the lack of homogeneity between the included and prior studies, it cannot be concluded if the absence of effects in study 6 was a result of low impact loading or another determinant, such as pubertal status, which will be discussed.

Lastly, a positive dose-response relationship was found in studies 7 and 10, which identified increased participant effort/intensity in a resistance-based intervention and increased training volume, regarding duration and frequency, in gymnastics had a beneficial effect on bone mass accrual, respectively. However, contradicting evidence was found in study 13, which found a negative dose-response relationship between running mileage and bone mass increase for multiple participants. The study concluded that the negative dose-response relationship may be explained by the athletes with a high running mileage being in a chronic energy deficit, as athletes with a 'normal' BMD score at baseline (BMD z-score>1) trended towards having a positive relationship, with training volume and LS BMC increase. Thus, the results from study 13 suggest that the effect of WBE on bone mass may vary depending on behavioural or physiological factors. However, as this evidence has been identified in a very high energy expending activity, this finding may only be specific to running.

Effects of pubertal status on bone accrual

Studies 11, 13 and 15 all found a negative association with higher puberal status and change in bone mass. These findings are in support of Heinonen et al. (2000)[54], who found a 50min aerobics and drop-jumping intervention significantly increased BMC of the FN for premenarcheal, but not post-menarcheal, girls. An explanation for the drop in bone mass sensitivity to loading may be caused by the increase in circulating sex steroids during Tanner II-V (e.g., estradiol) and the decrease in growth hormone and insulin-like growth factor 1 (IGF-1) during Tanner III-IV[55,56].

Studies 7 and 8 identified WBE affects bone mass accrual at different bone sites depending on pubertal stage. Both studies identified that early pubertal participants (Tanner II) had significant increases in PFNN width, while later pubertal participants (Tanner III) had significant increases in LS BMD. These findings can, again, be explained by hormone changes throughout maturation. Previous studies have identified increased circulating estrogen inhibits periosteal apposition, which reduces growth in bone width, but enhances trabecular bone accrual[57,58]. Therefore, it can be expected that mechanical loading during an early pubertal stage (low circulating estrogen) will increase bone width, while mechanical loading during a later pubertal stage (high circulating estrogen) will increase BMD.

Although evidence suggests increased bone mass accrual occurs at an earlier pubertal stage, the same cannot be said for young girls that are pre-pubertal (Tanner I). Study 6 only examined individuals in Tanner stage I and found a drop-jump intervention had no significant effect on BMC or vBMD compared to non-active control. Also, study 8 found no intervention effect at LS BMC or FN BMD after controlling for height and TB mass, which could be explained by at least half of the included participants being Tanner stage I at baseline. These findings are supported by previous research, which has also concluded pre-pubertal girls (Tanner I) to be less responsive to bone loading compared to girls who are early-pubertal (Tanner II-II)[59,60].

Longitudinal effects of WBE

Studies 11 and 12, which examined gymnastics participation, found a longitudinal benefit to bone accrual as a result of WBE. These findings are supported by prior research, which identified a higher bone mass in retired gymnasts compared with controls, suggesting gymnastics participation has a favourable benefit to PBM in young adulthood[17,61,62]. There is evidence for an inverse relationship between the timing of maturation and PBM in young adulthood, with the onset and duration of maturation having a large effect of bone mass accrual[63-65]. However, the gymnastics participants in both studies 11 and 12 showed delayed maturation age, pubertal development and skeletal maturity, which are all very common in elite gymnasts, but still found a positive bone accrual effect[66,67]. An explanation for this could be that gymnasts are undergoing late catch-up growth compared to other sporting athletes, which exacerbates bone mass growth[68].

Although these findings are positive, only studies 11 and 12 identified longitudinal benefits due to WBE (gymnastics) participation out of the 15 analysed. Furthermore, study 11 had a short follow-up period of 12-months, which may not allow sufficient time to reflect the true longitudinal variation in bone mass. Hence, further longitudinal research is needed in both gymnastics and other sporting athletes.

Review strengths and limitations

To our knowledge, the current systematic review is the first to collate the most recent evidence to determine the optimal form of WBE to produce the optimal effect on bone mass accrual in girls aged 5 to 18. Therefore, the review provides summary evidence that builds on the available literature base and informs future research directions. Additionally, the reporting and transparency of this systematic review was improved by following the PRISMA checklist[22].

However, the limitations of this systematic review and the current research area should be recognised. First, the inclusion of studies or data examining girls only could be deemed a limitation, as boys still experience osteoporosis and bone health aliments at an older age, albeit at a lower prevalence, thus reducing the generalisability of the findings. A major limitation in this research area is the limited number of RCT studies (5 out of the included 15). As demonstrated, the included non-RCT studies had, on average, a higher reporting bias and a lower methodological quality compared to the included RCT studies. Another limitation identified in most studies was the lack of controlling for at least one important confounder, such as calcium or vitamin D intake and habitual PA, which are all considered important variables regarding bone mass accrual[69]. Furthermore, the lack of homogeneity within the literature regarding the WBE exposure protocol used, the bone sites measured, and the outcome measure used (i.e., BMD or BMC), hinders the ability to draw accurate comparisons between studies and therefore prevents accurate and appropriate conclusions from being made. A summary of specific limitations for each study can be seen in Table 6.

Although this review is unable to determine the optimal mode of WBE to improve bone mass accrual in girls aged 5 to 18, the findings still have implications for future research. For example, there is scope for research to compare different interventions or WBE exposure modes within the same study (i.e., increase intervention arms). By ensuring all participants in each intervention arm are equal at baseline, and by measuring the same bone sites with the same bone outcomes, comparison between WBE exposures (i.e., type, intensity, duration, frequency, and load) will be improved. Schools are a practical and cost-effective location to examine and introduce WBE interventions, however, of the included studies in this review, only jump-based exercise was examined in a school-based setting. Therefore, future

research should study multiple WBE exposure modes, such as resistance-based or gymnastics exercises, in a school-based setting. Many studies in this review experience poor compliance and high drop-out rates, thus future studies should examine the underlying reasons for low compliance to aid future development and implementation of effective interventions. Also, where possible, studies that have already conducted longitudinal follow-ups should perform future follow-ups when participants reach their PBM age (between 25 and 30 years). This will allow for an evaluation of the effect of the original intervention on PBM compared to those in the CG.

Overall, this review supports the understanding that WBE activities, although different for different modes and Tanner stage, improves PBM in young girls. As stated above, only a small increase in PBM is needed to see large improvements in bone health during older adulthood and reduce the likelihood of fractures and osteoporosis[7] Therefore, it is important to promote these WBE activities in future physical activity recommendations for children and adolescence, especially in the female population as the rates of fractures and osteoporotic cases are higher[12,13].

Conclusion

To conclude, this systematic review supports the knowledge that WBE during childhood and adolescence improves outcomes of bone mass, such as BMD and BMC, in girls aged 5-18 years. Furthermore, the findings support the notion that WBE during Tanner stages II and III may be the optimal time for bone mass accrual due to circulating hormonal factors. However, the exact type, intensity, duration, frequency, and load of WBE interventions to elicit the optimal gain in bone mass remains unclear. The longitudinal effects of WBE are also still unclear. Nonetheless, the review's findings give direction for future research, which should focus on improving the homogeneity within or between studies, so accurate comparisons can be made between different WBE protocols with respect to their effect on bone mass accrual.

Authors' contributions

Tommy Cartledge: searched and screened articles, completed the data extraction, and prepared the manuscript; Joey Murphy and Charlie EM Foster: completed screening of articles and assisted in the revision and preparation of the manuscript; Byron Tibbitts: provided supervision and support throughout the course of the review process.

References

1. Sambrook PN, Cameron ID, Chen JS, Cumming RG, Lord SR, March LM, et al. Influence of fall related factors and bone strength on fracture risk in the frail elderly. Osteoporos Int 2007;18(5):603-10.
2. National Institute for Health and Care Excellence. NICE impact falls and fragility fractures [Internet]. 2018 [cited 2021 June 25] Available from: https://www.nice.org.uk/media/default/about/what-we-do/into-practice/measuring-uptake/nice-impact-falls-and-fragility-fractures.pdf
3. Leib E, Winzenrieth R, Aubry-Rozier B, Hans D. Vertebral microarchitecture and fragility fracture in men: A TBS study. Bone 2014;62(1):51-5.
4. Marshall L, Lang T, Lambert L, Zmuda J, Ensrud K, Orwoll E. Dimensions and volumetric BMD of the proximal femur and their relation to age among older U.S. men. J Bone Miner Res 2006;21(8):1197-206.
5. Ioannidis G, Papaioannou A, Hopman W, Akhtar-Danesh N, Anastassiades T, Pickard L, et al. Relation between fractures and mortality: results from the Canadian Multicentre Osteoporosis Study. CMAJ 2009;181(5):265-71.
6. Svedbom A, Hernlund E, Ivergård M, Compston J, Cooper C, Stenmark J, et al. Osteoporosis in the European Union: a compendium of country-specific reports. Arch Osteoporos 2013;8(1):1-218.
7. Sandhu S. and Hampson G. The pathogenesis, diagnosis, investigation and management of osteoporosis. J Clin Pathol 2011;64(12):1042-50.
8. Heaney RP, Abrams S, Dawson-Hughes B, Looker A, Marcus R, Matkovic V, et al. Peak bone mass. Osteoporosis Int.2000;11(1):985-1009.
9. Hind K. and Burrows M. Weight-bearing exercise and bone mineral accrual in children and adolescents: A review of controlled trials. Bone 2007;40(1):14-27.
10. Bonjour JP, Theintz G, Law F, Slosman D, Rizzoli R. Peak bone mass. Osteoporosis Int 1994;Suppl.1:7-13.
11. Haapasalo HH. Physical activity and growing bone: development of peak bone mass with special reference to the effects of unilateral physical activity. Ann Chir Gynaecol 1998;87(3):250-252.
12. Sarafrazi N, Edwina AW, Shepherd, JA. Osteoporosis or low bone mass in older adults: United States, 2017-2018. NCHS 2021;405(1):1-8.
13. Sözen T, Özisik L, Basaran NC. An overview and management of osteoporosis. Eur J Rheumatol 2017;4(1):46-56.
14. Gunter K, Almstedt H, Janz, K. Physical Activity in Childhood May Be the Key to Optimizing Lifespan Skeletal Health. Exerc Sport Sci Rev 2012;40(1):13-21.
15. McDevitt H. and Ahmed S. Establishing good bone health in children. Paediatr Child Health 2010;20(2):83-87.
16. Gómez-Bruton A, Matute-Llorente Á, González-Agüero A, Casajús J, Vicente-Rodríguez G. Plyometric exercise and bone health in children and adolescents: a systematic review. World J Pediatr 2017;13(2):112-21.
17. Bass S, Pearce G, Bradney, M, Hendrich E, Delmas P, Harding A, et al. Exercise before puberty may confer residual benefits in bone density in adulthood: studies in active prepubertal and retired female gymnasts. J Bone Miner Res 1998;13(3):500-7.
18. Luiz-de-Marco R, Gobbo L, Castoldi R, Maillane-Vanegas S, da Silva Ventura Faustino-da-Silva Y, Exupério I, et al. Impact of changes in fat mass and lean soft tissue on bone mineral density accrual in adolescents engaged in different sports: ABCD Growth Study. Arch Osteoporos 2020;15(22):1-8.
19. Frost HM. Wolff's Law and bone's structural adaptations to mechanical usage: an overview for clinicians. Angle Orthod. 1994;64:175-88.
20. Department of Health and Social Care. Physical Activity Guidelines: UK Chief Medical Officers' Report [Internet]. Department of Health and Social Care; London, UK. 2019 [cited 2021 June 29]. Available from: https://www.gov.uk/government/publications/physical-activity-guidelines-uk-chief-medical-officers-report
21. American College of Sports Medicine. Physical Activity in Children and Adolescents [Internet]. 2022 [cited 2022 May 15]. Available from: https://www.acsm.org/docs/default-source/files-for-resource-

library/physical-activity-in-children-and-adolescents.pdf
22. Page MJ., McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;327(1):1-71.
23. McDonagh M, Peterson K, Raina P, Chang S. Shekelle P. Avoiding bias in selecting studies. Methods Guide for Effectiveness and Comparative Effectiveness Reviews [Internet] (Agency for Healthcare Research and Quality, 2013).
24. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. eds. Cochrane Handbook for Systematic Reviews of Interventions. version 6.2. London, Cochrane; 2021.
25. National Institute for Health Consensus Development Panel on Osteoporosis Prevention, Diagnosis, and Therapy. Osteoporosis Prevention, Diagnosis, and Therapy. JAMA. 2001;285(6):785-95.
26. Kannus P, Haapasalo H, Sievänen H, Oja P, Vuori I. The site-specific effects of long-term unilateral activity on bone mineral density and content. Bone 1994;15(3):279-84.
27. Higgins JPT, Sterne JAC, Savović J, Page MJ, Hróbjartsson A, Boutron I, et al. A revised tool for assessing risk of bias in randomized trials In: Chandler, J., McKenzie, J., Boutron, I. and Welch, V., eds. Cochrane Methods. Cochrane Database of Systematic Reviews; 2016. Issue 10 (Supplement 1).
28. Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ 2016;355(1):4919-50.
29. Oremus M, Wolfson C, Perrault A, Demers L, Momoli, F, Moride Y. Interrater reliability of the modified Jadad quality scale for systematic reviews of Alzheimer's disease drug trials. Dement Geriatr Cogn Disord 2001;12(3):232-6.
30. Liu Y, Li Z, Li H, Zhang Y, Wang P. Protective effect of surgery against early subtalar arthrodesis in displaced intra-articular calcaneal fractures. Medicine 2015;94(45):1984-88.
31. Ha A and Ng J. Rope skipping increases bone mineral density at calcanei of pubertal girls in Hong Kong: A quasi-experimental investigation. PLOS ONE 2017;12(12):1-9.
32. Daly R, Ducher G, Hill B, Telford R, Eser P, Naughton G, et al. Effects of a specialist-led, school physical education program on bone mass, structure, and strength in primary school children: a 4-year cluster randomized controlled trial. J Bone Miner Res 2016;31(2):289-98.
33. Nogueira R, Weeks B, Beck, B. An in-school exercise intervention to enhance bone and reduce fat in girls: The CAPO Kids trial. Bone 2014;68:92-99.
34. Meiring RM, Micklesfield LK, Avidon I, McVeigh JA. Osteogenic effects of a physical activity intervention in South African black children. J Musculoskelet Neuronal Interact 2014;14(3):276-85.
35. Weeks B, Young C, Beck B. Eight months of regular in-school jumping improves indices of bone strength in adolescent boys and girls: the POWER PE Study. J Bone Miner Res 2008;23(7):1002-11.
36. Wiebe P, Blimkie C, Farpour-Lambert N, Briody J, Marsh D, Kemp A, et al. Effects of single-leg drop-landing exercise from different heights on skeletal adaptations in prepubertal girls: a randomized controlled study. Pediatr Exerc Sci 2008;20(2):211-28.
37. Dowthwaite JN, Weiss DM, Thein-Nissenbaum J, Scerpella TA. A 2-yr, school-based resistance exercise pilot program increases bone accrual in adolescent girls. Transl J Am Coll Sports Med 2019;4(11):74-83.
38. Bernardoni B, Thein-Nissenbaum J, Fast J, Day M, Li Q, Wang S, et al. A school-based resistance intervention improves skeletal growth in adolescent females. Osteoporosis Int 2014;25(3):1025-32.
39. Exupério I, Agostinete R, Werneck A, Maillane-Vanegas S, Luiz-de-Marco R, Mesquita E, et al. Impact of artistic gymnastics on bone formation marker, density and geometry in female adolescents: ABCD-Growth Study. J Bone Metab 2019;26(2):75-82.
40. Burt LA, Ducher G, Naughton GA, Courteix D, Greene DA. Gymnastics participation is associated with skeletal benefits in the distal forearm: a 6-month study using peripheral Quantitative Computed Tomography. J. Musculoskelet Neuronal Interact 2013;13(4):395-404.
41. Maïmoun L, Coste O, Mura T, Philibert P, Galtier F, Mariano-Goulart D, et al. Specific bone mass acquisition in elite female athletes. J Clin Endocrinol Metab 2013;98(7):2844-53.
42. Pikkarainen E, Lehtonen-Veromaa M, Kautiainen H, Heinonen O, Viikari J, Möttönen T. Exercise-induced training effects on bone mineral content: a 7-year follow-up study with adolescent female gymnasts and runners. Scand J Med Sci Sports 2009;19(2):166-73.
43. Barrack M, Van Loan M, Rauh M, Nichols J. Body mass, training, menses, and bone in adolescent runners: a 3-year follow up. Med Sci Sports Exerc 2011;43(6):959-66.
44. Ito I, Kemper H, Agostinete R, Lynch K, Christofaro D, Ronque E, et al. Impact of martial arts (judo, karate, and kung fu) on bone mineral density gains in adolescents of both genders: 9-month follow-up. Pediatr Exerc Sci 2017;29(4):496-503.
45. Ducher G, Bass S, Saxon L, Daly R. Effects of repetitive loading on the growth-induced changes in bone mass and cortical bone geometry: A 12-month study in pre/peri- and postmenarcheal tennis players. J Bone Miner Res 2011;26(6):1321-29.
46. Ishikawa S, Kim Y, Kang M, Morgan, D. Effects of weight-bearing exercise on bone health in girls: a meta-analysis. Sports Med 2013;43(9):875-92.
47. Bass S, Saxon L, Daly R, Turner C, Robling A, Seeman E, et al. The effect of mechanical loading on the size and shape of bone in pre-, peri-, and post-pubertal girls: a study in tennis players. J Bone Miner Res 2002;17(12):2274-80.
48. Daly RM. The effect of exercise on bone mass and structural geometry during growth. Med Sci Sports Exerc 2007;51(1):33-49.
49. Rachner T, Khosla S, Hofbauer L. Osteoporosis: now and the future. Lancet 2011;377(9773):1276-87.
50. Courteix D, Lespessailles E, Peres S, Obert P, Germain P, Benhamou C. Effect of physical training on bone mineral density in prepubertal girls: A comparative study between impact-loading and non-impact-loading sports. Osteoporosis Int 1998;8(2):152-58.
51. Maïmoun L, Coste O, Philibert P, Briot K, Mura T, Galtier F, et al. Peripubertal female athletes in high-impact sports show improved bone mass acquisition and bone geometry. Metabolism 2013;62(8):1088-98.
52. McNitt-Gray J. Kinetics of the lower extremities during drop landings from three heights. J Biomech 1993;26(9):1037-46.
53. Fuchs R, Bauer J, Snow C. Jumping improves hip and lumbar spine bone mass in prepubescent children: a randomized controlled trial. J Bone Miner Res 2000;16(1):148-56.
54. Heinonen A, Sievänen H, Kannus P, Oja P, Pasanen M, Vuori I. High-impact exercise and bones of growing girls: a 9-month controlled trial. Osteoporosis Int 2000;11(12):1010-17.
55. MacKelvie K, Khan KM, McKay HA. Is there a critical period for bone response to weight-bearing exercise in children and adolescents? a systematic review. Br J Sports Med 2002;36(4):250-57.
56. Xu L, Wang Q, Wang Q, Lyytikäinen A, Mikkola T, Völgyi E, et al. Concerted actions of insulin-like growth factor 1, testosterone, and estradiol on peripubertal bone growth: a 7-year longitudinal study. J Bone Miner Res 2011;26(9):2204-11.
57. Seeman E. Periosteal bone formation: a neglected determinant of bone strength. N Engl J Med 2003;349(4):320-3.

58. Ausili E, Rigante D, Salvaggio E, Focarelli B, Rendeli C, Ansuini V, et al. Determinants of bone mineral density, bone mineral content, and body composition in a cohort of healthy children: influence of sex, age, puberty, and physical activity. Rheumatol 2012;32(9):2737-43.
59. MacKelvie K, McKay H, Khan K, Crocker P. A school-based exercise intervention augments bone mineral accrual in early pubertal girls. J Pediatr 2001;139(4):501-8.
60. Petit M, Mckay H, Mackelvie K, Heinonen A, Khan K, Beck T. A randomized school-based jumping intervention confers site and maturity-specific benefits on bone structural properties in girls: a hip structural analysis study. J Bone Miner Res 2002;17(3):363-72.
61. Erlandson M, Kontulainen S, Chilibeck P, Arnold C, Faulkner R, Baxter-Jones A. Higher premenarcheal bone mass in elite gymnasts is maintained into young adulthood after long-term retirement from sport: A 14-year follow-up. J Bone Miner Res 2011;7(1):104-10.
62. Eser P, Hill B, Ducher G, Bass S. Skeletal benefits after long-term retirement in former elite female gymnasts J Bone Miner Res 2009;24(12):1981-88.
63. Jackowski S, Erlandson M, Mirwald R, Faulkner R, Bailey D, Kontulainen S, et al. Effect of maturational timing on bone mineral content accrual from childhood to adulthood: evidence from 15years of longitudinal data. Bone 2011;48(5):1178-85.
64. Chevalley T, Bonjour J, Ferrari S, Rizzoli R. Deleterious effect of late menarche on distal tibia microstructure in healthy 20-year-old and premenopausal middle-aged women. J Bone Miner Res 2009;24(1):144-52.
65. Gilsanz V, Chalfant J, Kalkwarf H, Zemel B, Lappe J, Oberfield S, et al. Age at onset of puberty predicts bone mass in young adulthood. J Pediatr 2011;158(1):100-5.
66. Georgopoulos N, Markou K, Theodoropoulou A, Paraskevopoulou P, Varaki L, Kazantzi Z, et al. Growth and pubertal development in elite female rhythmic gymnasts. J Clin Endocrinol Metab 1999;84(12):4525-30.
67. Georgopoulos N, Markou K, Theodoropoulou A, Benardot D, Leglise M, Vagenakis, A. Growth retardation in artistic compared with rhythmic elite female gymnasts. J Clin Endocrinol Metab 2002;87(7):3169-73.
68. Georgopoulos N, Markou K, Theodoropoulou A, Vagenakis G, Benardot D, Leglise M, et al. Height velocity and skeletal maturation in elite female rhythmic gymnasts. J Clin Endocrinol Metab 2001;86(11):5159-64.
69. Golden N and Abrams S. Optimizing bone health in children and adolescents. Pediatrics 2014;134(4):1229-43.

Table 1. PRISMA checklist for systematic reviews and metanalyses.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	**p.1**
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	**p.1**
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	**p.1&2**
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	**p.2**
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	**p.2&3**
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	**p.2**
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	**p.2**
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	**p.2&3**
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	**p.3&4**
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	**p.3&4**
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	**p.4**
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	**p.4**
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	**N/A**
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	**N/A**
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	**N/A**
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	**N/A**
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	**N/A**
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	**N/A**
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	**N/A**
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	**N/A**

Table 1. *(Cont. from previous page).*

Section and Topic	Item #	Checklist item	Location where item is reported
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	p.4
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	p.4-6
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	p.4&5
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimates and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	p.6-9
	23b	Discuss any limitations of the evidence included in the review.	p.8
	23c	Discuss any limitations of the review processes used.	p.8&9
	23d	Discuss implications of the results for practice, policy, and future research.	p.8&9
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

Table 2. Summary of study characteristics.

No.	Author (country)	Study design and length (months)	Sample size (age at BL)	Methods	Measures	Key results
Jump-based exercise						
1.	Ha and Ng, 2017[28] (Hong Kong)	Longitudinal,12	Total; n=179 (Tanner I-V) ExG; n=66 (12.5 1.7) CG; n=110 (12.1 1.9)	**ExG;** Rope skipping for ≥60mins weekly. **CG;** No regular rope skipping, <60mins weekly.	Measures taken at BL and 12 months. **BMD (g/cm²)**, forearms and calcanei via DEXA. **Pubertal status,** self-reported scale. **Previous engagement** (rope skipping last 3 months), via self-reported questionnaire. **Habitual PA,** Chinese version of the PA Questionnaire-Children.	ExG significant ↑ in forearm BMD at follow-up compared to BL (0.26 0.10 vs 0.22 0.10, p<0.05). **ExG vs CG** No significant ↑ in calcanei BMD for ExG (p=0.09) or CG (p=0.12). ExG significantly higher forearm BMD at follow-up vs CG CG (0.26 0.10 vs 0.23 0.10, p<0.05). ExG significantly higher calcanei BMD at follow-up vs CG (0.47 0.08 vs 0.43 0.08, p<0.05). **Multilevel modelling analyses** ExG significantly higher volume of calcanei BMD at follow-up vs CG (B=0.02, 95% CI 0.01 to 0.04, p<0.01) but not forearm BMD (B=0.01, 95% CI -0.01 to 0.03, p=0.26).
2.	Daly et al., 2016*[29] (Australia)	Cluster RCT, 48 (Unit of randomisation= schools)	Total; n=362 (Tanner I-III) IG; n=192 (8.1 0.3) CG; n=170 (8.1 0.4)	**IG;** 100min/week of specialist-led PE (via specialist PE leaders). Plus 50mins/week in school PE lessons. (150mins/week total) **CG;** 150mins/week of in school PE lessons (via by the general classroom teachers).	Measures taken at BL, 24 months and 48 months. **BMC (g),** TB via DEXA. **Pubertal status,** self-reported scale and menarche status via questionnaire. **Habitual PA,** PA index to approximate average steps/day. **Total energy intake (kJ/d) and dietary calcium (mg/d),** 24-hour dietary record on a school day (assisted by parents and teachers)	**IG vs CG** Difference in TB BMC between groups after 24 months and 48 months were non-significant. 24 months; 169, 95% CI 150 to 188 vs 166, 95% CI 145 to 186, p>0.05 48 months; 476, 95% CI 455 to 496 vs 455, 95% CI 433 to 477, p>0.05
3.	Nogueira et al., 2014[30] (Australia)	RCT, 9	Total; n=138 (Tanner I-III) IG; n=71 (10.5 0.6) CG; n=67 (10.7 0.6)	**IG;** 10mins high intensity movement exercises (150 jumps, 50 kicks, 30-40 inverted position movements). 3xweekly (Tue-Thur) (except school holidays). Performed at max effort. Also took part in normal PE classes. **CG;** ~45mins, 2x weekly normal PE classes.	Measures taken at BL and 9 months. **BMD (g/cm²) and BMC (g),** TB, FN and LS via DEXA. **IBS (g²/cm⁴)** **Pubertal status,** peak height velocity. **Habitual PA,** BPAQ Food intake, ACAES	**IG vs CG** Greater ↑ in BMD and BMC at all measured bone sites for IG vs CG, differences all non-significant. **TB BMD** 7.5% vs 2.1% (p=0.08) **TB BMC** 13.7% vs 9.0% (p=0.44) **LS BMD** 11.0% vs 6.3% (p=0.32) **LS BMC** 21.6% vs 10.0% (p=0.25) **FN BMD** 10.0% vs 6.3% (p=0.59) **FN BMC** 12.3% vs 6.3% (p=0.309) **Δ** ↑ of LS IBS between IG and CG reached significance, 24.4% vs 12.0%, respectively (p<0.01).
4.	Meiring et al., 2014*[31] (South Africa)	Cluster RCT, 5	Total n=15 (Tanner I-III) ExG; n=8 (9.7 1.2) CG; n=7 (9.3 0.9)	**ExG;** WBE programme, 45mins, 2x weekly. Sprints, running + jumping, ladder hopping, single leg hopping and rope skipping. Also took part in normal PE classes. **CG;** ~35mins 1x weekly continued normal PE classes.	Measures taken at BL and 5 months. **BMC (g),** TB, LS, FN and TH via DEXA. **Pubertal status,** 5-scale Tanner stages (assisted by parents) (Tanner, 1962). **Previous PA engagement**, validated self-reported PA questionnaire (last 2 years before baseline; assisted by parents).	**ExG vs CG** Difference in ↑ between groups were all non-significant. TB BMC 822.6 195.5 vs 792.9 116.7 (p>0.05) LS BMC 24.3 6.2 vs 24.4 4.7 (p>0.05) FN BMC 3.0 0.5 vs 2.7 0.3 (p>0.05) TH BMC 18.7 5.5 vs 16.5 3.1 (p>0.05) Significant group-by-time effect for TH BMC (IG>CG) 18.7 5.5 vs 16.5 3.1, p=0.04. Significant time effects in both groups for TH, LS, and TB BMC (5 months>BL, all p<0.001):

Table 2. *(Cont. from previous page).*

No.	Author (country)	Study design and length (months)	Sample size (age at BL)	Methods	Measures	Key results
5.	Weeks, Young and Beck, 2008*[32] (Australia)	RCT, 8	Total n=53 (Tanner I-IV) IG; n=30 (13.7 0.4) CG; n=23 (13.7 0.5)	**IG;** 10min 2x weekly jumping activity in place of regular PE warm-up. Including; Jumps, hops, tuck-jumps, jump-squats, stride-jumps, star jumps, lunges, side-lunges and skipping. Frequency of exercise 1-3 Hz and a height of 0.2-0.4 m. Single instructor led. **CG;** 2x weekly usual PE warm-up.	Measures taken at BL and at a follow-up of 8 months. **BMD (g/cm²) and BMC (g),** TB, FN, TR and LS via DEXA. **IBS (g²/cm⁴)** **Pubertal status,** 1. 5-scale Tanner stages (assisted by parents) (Tanner, 1962). 2. Peak height velocity. Habitual PA, BPAQ. **Dietary calcium intake,** self-reported calcium-focused FFQ.	**IG vs CG** Both the IG and CG had a significant ↑ in TB BMC (+6.5%, $p<0.001$ and +4.0%, $p<0.01$, respectively) and LS BMC (+9.0%, $p<0.001$ and +10.7%, $p<0.001$). IG significant ↑ in LS BMD (+5.2%, $p<0.05$) and FN BMC (+13.9%, $p=0.05$) compared to BL measurements. CG non-significant ↑ in LS BMD (+1.5, $p>0.05$) and FN BMC (+4.9, $p>0.05$) compared to BL. Significant increase in LS IBS between baseline and follow-up in the IG (+13.9%, $p<0.001$) and CG (+10.9%, $p<0.001$). No significant between-group differences for any of the bone site ($p>0.05$).
6.	Wiebe et al., 2008[33] (Australia)	RCT, 7	Total n=42 (Tanner I only) IG HD (high drop); n=13 (7.9 1.1) IG LD (low drop); n=13 (7.8 0.9) CG; n=14 (7.9 0.8)	**IG;** 2 groups, 1. Single-leg drop-landing from 24cm (high drop, IGHD) 2. Single-leg drop-landing from 14cm (low drop, IGLD) 3xweekly sessions; 10 sets of 5 jumps (150 jumps weekly). **CG;** non-active control group.	Measures taken at BL and at a follow-up of 7 months. **BMC (g),** TB, FN and TR via DEXA. **vBMD (g/cm³),** FN and MFS **Pubertal status**, 5-scale Tanner stages (assisted by parents) (Tanner, 1962). **Habitual PA,** PYLTPAQ (Aaron et al., 1995). **Dietary calcium intake** via a self-reported 3-day diary (assisted by parents).	**IGHD vs IGLD vs CG** No significant differences were found in adjusted BMC or vBMD at any measured bone site for both IG and CG ($p>0.05$). Combing data for exercise groups found no significant differences in BMC or vBMD when compared to control ($p>0.05$).
				Resistance-based exercise		
7.	Dowthwaite et al., 2019[34] (USA)	Longitudinal, 24	Total; n=62 (Tanner I-IV) LO ExG; n=22 (11.4 0.3) HI ExG; n=19 (11.6 0.3) CG; n=21 (11.6 0.3)	**ExG;** 8-12mins (2-4 exercises), of resistance training, 2-3xweekly. Included; Lunges, push-ups, body rows, jumps/hops. Also took part in normal PE classes. 2 groups split for participant effort, either high (HI) or low (LO). **CG;** 45mins, 2-3 times weekly normal PE classes.	Measures taken at BL, 12 months and 24 months. **BMD (g/cm²) and BMC (g),** FN and LS (L1-L4) via DEXA. **Pubertal status** via a self-reported scale. **Habitual PA** via self-assed form (assisted by parents). **Exercise effort/intensity**, instructor-reported observation scale 1=low effort, 2=medium effort, 3=high effort.	**ExG vs CG** Difference in ↑ for LS BMD and BMC significantly greater for ExG vs CG (5.5% and 4.1%, respectively; $p<0.05$) Difference in ↑ for FN BMD and BMC greater in ExG vs CG but non-significant (3.4% and 3.8%, respectively; $p>0.05$) **HI ExG vs CG** HI IG significant ↑ in BMD and BMC at the LS (5.5% and 8.2%, respectively; $p<0.01$) and FN (6.1% and 7.1%, respectively; $p<0.01$) vs CG.
8.	Bernardoni et al., 2014[35] (USA)	Longitudinal, 7	Total; n=38 (Tanner II-III) ExG Tanner II; n=10 (11.5 0.3) ExG Tanner III; n=7 (11.8 0.3) CG Tanner II; n=11 (11.6 0.4) CG Tanner III; n=10 (11.8 0.3)	**ExG;** Targeted resistance exercise 8-12mins, 2x weekly. Included floor + resistance exercise using BW, hand weights and resistance bands. Also took part in normal PE classes. **CG;** 150mins/week of in school PE lessons. Both groups split for Tanner stage, either Tanner II or III.	Measures taken at BL and 7 months. **BMD (g/cm²) and BMC (g)** of the TB, LS, PF (including NN width [cm]) and arm via DEXA. **Pubertal status,** 1. Self-assessment using annotated line drawings (with parental assistance). 2. Gynecologic age (if post-menarche). **Habitual PA**, validated interviewer-administered questionnaire. **Exercise effort/intensity,** instructor-reported observation scale 1=low effort, 2=medium effort, 3=high effort.	**ExG vs CG** Non-significant differences between group at all bone sites ($p>0.05$). **Analysing Tanner stage together** +ve trend for ExG at the following bone sites; Spine BMC ($p=0.05$, $f=0.15$, medium effect) LS BMC ($p=0.10$, $f=0.11$, small effect) PF NN width ($p=0.08$, $f=0.10$, small effect) TB BMC ($p=0.12$, $f=0.08$, small effect) Arm BMC ($p=0.07$, $f=0.12$, small effect) **Adjusting for Tanner stage** A significant +ve effect for ExG Tanner II at PFNN width ($p=0.01$) and for ExG Tanner III at LS BMD ($p=0.03$).

Table 2. *(Cont. from previous page).*

No.	Author (country)	Study design and length (months)	Sample size (age at BL)	Methods	Measures	Key results
				Gymnastics		
9.	Exuperio et al., 2019[36] (Brazil)	Longitudinal, 12	Total; n=20 (Tanner I-V) ExG; n=10 (14.6 2.7) CG; n=10 (13.8 1.3)	**ExG;** Artistic gymnastic athletes, ~3 days/week and 207 42.6 mins/day training. **CG;** No regular PA (non-active) outside of school. Continued normal PE classes (45mins, 2x weekly).	Measures taken at BL and 12 months. **BMD (g/cm²),** TB and femur (FN, TR, Wards triangle, shaft and whole femur) via DEXA. **Pubertal status**, peak height velocity. **Weekly training load**, self-reported RPE calculations.	**ExG vs CG** Wards triangle, TR and whole femur BMD ↑ in ExG vs to CG, by 19% (1.12 0.16 vs 0.94 0.11, p<0.01), 14% (0.91 0.12 vs 0.80 0.12, p<0.05) and 10% (1.12 0.11 vs 1.01 0.12, p<0.05), respectively. Non-significant difference in TB BMD between groups (p=0.12).
10.	Burt et al., 2013[37] (Australia)	Longitudinal, 6	Total n=84 (Tanner I-II) ExG HG (High gymnastics); n=28 (9.1 1.3) ExG LG (Low gymnastics); n=28 (8.5 1.3) CG (No gymnastics); n=28 (8.5 1.3	3 groups based on artistic gymnastics participation; high-training (6-16 hours/week), low-training (1-5 hours/week) or non-gymnasts.	Measures taken at BL and at a follow-up of 6 months. **BMD (g/cm²) and BMC (g)**, TB and arm (humerus, ulna, radius, carpels and phalanges) via DEXA. **Pubertal status**, self-reported questionnaire (assisted by parents). **Training history**, parental-reported questionnaire.	**ExG vs CG** Non-significant differences between group at all bone sites (p>0.05). Significant time effects for TB BMD, TB BMC and arm BMC for all three groups (6 months>BL, all p<0.001). Significant group effect for arm BMC (IG HG>CG, p<0.05).
11.	Maïmoun et al., 2013[38] (France)	Longitudinal, 12	Total n=72 (Tanner I-V) ExG RG (rhythmic gymnasts); n=24 (13.9 1.7) ExG SW (swimmers); n=24 (14.4 1.5) CG; n=24 (14.3 1.8)	**ExG;** 2 groups, 1. rhythmic gymnastics (RG) 2. swimming (SW). >8 hours/week (23.0 2.7 for RG and 14.4 4.7 for SW) trainng. **CG;** Leisure activities for <3 hours/week.	Measures taken at BL and at a follow-up of 12 months. **BMD (g/cm²),** TB, FN, TR, LS (L2-L4), PF and arm radius via DEXA. **Pubertal onset in family** (menarche of mothers) via a self-reported standardised questionnaire (assisted by parents). **Training history**, self-reported questionnaire.	Significant ↑ observed at all bone sites in all 3 groups (p<0.05) **RG vs SW** Higher BMD at the TB, TR, FN, PF and LS in RG vs SW (all p<0.001). Same pattern at the 12-month follow-up (RG vs SW); TB 1.07 0.01 vs 0.99 0.01, p<0.001 TR 0.82 0.01 vs 0.69 0.01, p<0.001 FN 1.06 0.02 vs 0.75 0.02, p<0.001 PF 1.06 0.01 vs 0.88 0.01, p<0.01 **RG and SW vs CG** Non-significant difference at BL or follow-up between SW and CG. Higher BMD at the TB, TR, FN and PF in RG vs CG (all p<0.01). Same pattern was the same at the 12-month follow-up (RG vs CG); TB 1.07 0.01 vs 1.01 0.01, p<0.05 TR 0.82 0.01 vs 0.73 0.01, p<0.001 FN 1.06 0.02 vs 0.83 0.02, p<0.001 PF 1.06 0.01 vs 0.93 0.01, p<0.001
12.	Pikkarainen et al., 2009[39] (Finland)	Longitudinal, 84	Total n=142 (Tanner stage I-IV) ExG G (gymnastics); n=52 (13.0 1.7) ExG R (running); n=46 (13.0 1.9) CG; n=44 (13.0 1.7)	**ExG;** Comparison between 2 experimental groups, 1. Competing gymnastic athletes 2. Competing running athletes **CG**; Non-athletic control group.	Measures taken at BL and at follow-ups of 12, 24, 36 and 84 months. **BMC (g),** LS (L2-L4) and FN via DEXA. **Pubertal status**, 5-scale Tanner stages (assisted by parents) (Tanner, 1962). **PA history** via self-reported questionnaire.	**ExG G vs CG** 36 months; FN BMC 4.5 0.8 vs 4.0 0.6, p<0.001 84 months; FN BMC 4.5 0.7 vs 4.1 0.6, p<0.002 No significant differences for LS BMC between or within either ExG G or CG. **ExG R vs CG** 36 months; FN BMC 4.4 0.8 vs 4.0 0.6, p<0.001 LS BMC 57.3 12.2 vs 51.4 8.4, p<0.002 84 months; FN BMC 4.4 0.7 vs 4.1 0.6, p<0.002 LS BMC 62.8 10.6 vs 55.2 8.7, p<0.002

Table 2. *(Cont. from previous page).*

No.	Author (country)	Study design and length (months)	Sample size (age at BL)	Methods	Measures	Key results
Endurance running						
13.	Barrack et al., 2011[40] (USA)	Longitudinal, 36	Total n=39 (Tanner stage IV-V) ExG (runners) n=39 (15.9 0.2)	**ExG;** High school endurance runners. 2 groups split for bone mass, either 1. Low bone mass (LBM; BMD z-score ≤ -1) 2. Normal bone mass (NBM; BMD z-score > 1)	Measures taken at BL and at a follow-up of 36 months. **BMD (g/cm²) and BMC (g),** TB, LS (L1-L4), and TH via DEXA. **Pubertal status**, self-reported questionnaire. **Training and PA history**, athletic preparticipation medical history form (Van de Loo & Johnson, 1995).	**At BL,** Low bone mass n=15 Normal bone mass n=24 **LBM vs NBM** Both groups had a significant ↑ in TB and LS BMC (all p<0.001). Significant ↑ in TH BMC for LBM only (p<0.05) NBM had significantly higher TB, LS and TH BMC at BL and follow up, compared to low bone mass (all p<0.05). Follow up data (NBM vs LBM): TB; 2667.0 39.2 vs 2385.0 51.6, p<0.01 LS; 62.9 1.5 vs 52.2 2.0, p<0.01 TH; 32.7 0.8 vs 28.9 1.1, p<0.05
Martial arts						
14.	Ito et al., 2017*[41] (Brazil)	Longitudinal, 9	Total; n=35 (Tanner I-V) ExG (Karate and Kung Fu); n=10 (12.7 1.5) ExG (Judo); n=9 (12.9 1.4) CG; n=16 (12.7 2.5)	**ExG;** 2 groups, 1. Karate and Kung Fu athletes (3-5 days/week) 2. Judo athletes (3-6 days/week) **CG;** No sport engagement outside of school.	Measures taken at BL and 9 months. **BMD (g/cm²),** TB and LS via DEXA. **Pubertal status**, peak height velocity. **Training history**, self-reported questionnaire.	**ExG vs CG** TB and LS BMD ↑ significantly at follow-up in both ExG groups and CG (all p<0.05). Adjusted Δ BMD between BL and follow-up (all p<0.05). **TB:** CG (0.03, 95% CI 0.02 to 0.04) Karate/Kung Fu (0.04, 95% CI 0.03 to 0.06) Judo (0.04, 95% CI 0.02 to 0.06) **LS:** CG (0.08, 95% CI 0.06 to 0.10) Karate/Kung Fu (0.05, 95% CI 0.01 to 0.08) Judo (0.08, 95% CI 0.05 to 0.11) Non-significant differences between ExG groups at both bone sites (p>0.05).
Tennis						
15.	Ducher et al., 2011[42] (USA)	Longitudinal, 12	Total n=45 (Tanner stage I-V) ExG; Premenarcheal (pre/peri; Tanner I-III) n=13 (12.3 1.0) ExG; Postmenarcheal (post; Tanner IV-V) n=32 (14.5 1.4)	**ExG;** 2 groups, 1. Premenarcheal (pre/peri) tennis players 2. Postmenarcheal (post) tennis players. Both groups ≥2 hours/week.	Measures taken at BL and at a follow-up of 12 months. **BMC (g),** TB and humerus (both arms) via DEXA. **Pubertal status**, 5-scale Tanner stages (assisted by parents) (Tanner, 1962). **Training history** via self-reported questionnaire (assisted by parents).	**Pre/peri vs Post** Significantly higher TB BMC at BL and follow-up for post vs pre/peri (2350.0 320.0 vs 1750.0 350.0, p<0.001 and 2460.0 310.0 vs 2170.0 330.0, p<0.05, respectively). Significant ↑ in non-playing arm BMC for pre/peri and post (+19.5%, p<0.001 and +4.6, p<0.001, respectively).

***Abbreviations**: BMD= bone mineral density; vBMD= volumetric bone mineral density; BMC= bone mineral content; IBS; index of bone structural strength; ExG= experimental group; IG= intervention group; CG= control group; RCT= randomised control trial; BL= baseline; BPAQ= bone-specific physical activity questionnaire; PYLTPAQ= past year leisure-time physical activity questionnaire; FFQ= food frequency questionnaire; DEXA= dual-energy X-ray absorptiometry; WBE= weight bearing exercise; PA= physical activity; PE= physical education; CF= cardiovascular fitness; AC= aerobic capacity; FM= fat mass; LM= lean mass; %BF= percentage body fat; %FM= percentage fat mass; LST= lean tissue mass; WC= waist circumference; BP= blood pressure; RHR= resting heart rate; TB= total body; TH= total hip; TR= trochanter; LS= lumbar spine; FN= femoral neck; MFS= mid-femoral shaft; PF= proximal femur; NN width= narrow neck width; Δ= change in; ↑= increase; ↓= decrease; +ve= positive; 95% CI= 95% confidence interval; mins= minutes; mins/week= minutes per week; hours/week= hours per week.*
** studies that included girls and boys (data included for girls only).*

Table 3. Summary of risk of bias assessment for RCT studies using RoB 2 risk of bias criteria.

No.	Random sequence generation	Deviations from the intended intervention (effect of assignment and adherence, respectively)	Missing outcome data	Measurement of outcome	Selection of reported data	Total
2	LR	LR and SR	HR	LR	LR	LR: 4; SR: 1; HR:1
3	SR	LR and LR	HR	LR	LR	LR: 4; SR: 1; HR:1
4	SR	LR and LR	LR	LR	LR	LR: 5; SR: 1; HR:0
5	SR	LR and LR	LR	LR	LR	LR: 5; SR: 1; HR:0
6	SR	LR and LR	LR	LR	LR	LR: 5; SR: 1; HR:0
Total	LR: 1 SR: 4 HR; 0	LR: 9 SR: 1 HR; 0	LR: 3 SR: 0 HR; 2	LR: 5 SR: 0 HR; 0	LR: 5 SR: 0 HR; 0	LR: 23 SR: 5 HR; 2

LR= low risk, SR= some risk, HR= high risk

Table 4. Summary of risk of bias assessment for non-RCT studies using the ROBINS-I risk of bias criteria.

No.	Confounding	Recruitment of participants	Classification of intervention	Deviation from intended intervention	Missing data	Measurement of outcomes	Selection of the reported result	Total
1	SR	MR	MR	LR	MR	SR	LR	LR: 2; MR: 3; SR: 2; CR: 0; NI: 0
7	MR	LR	LR	LR	LR	LR	LR	LR: 6; MR: 1; SR: 0; CR: 0; NI: 0
8	SR	LR	LR	LR	LR	SR	LR	LR: 5; MR: 0; SR: 2; CR: 0; NI: 0
9	SR	MR	LR	LR	LR	LR	LR	LR: 5; MR: 1; SR: 1; CR: 0; NI: 0
10	MR	MR	MR	LR	LR	LR	LR	LR: 4; MR: 3; SR: 0; CR: 0; NI: 0
11	SR	MR	MR	LR	LR	LR	LR	LR: 4; MR: 2; SR: 1; CR: 0; NI: 0
12	MR	MR	MR	MR	LR	LR	LR	LR: 3; MR: 4; SR: 0; CR: 0; NI: 0
13	SR	MR	MR	LR	LR	LR	LR	LR: 4; MR: 2; SR: 1; CR: 0; NI: 0
14	SR	MR	MR	LR	SR	LR	LR	LR: 3; MR: 2; SR: 2; CR: 0; NI: 0
15	SR	MR	MR	LR	LR	LR	LR	LR: 4; MR: 2; SR: 1; CR: 0; NI:0
Total	LR: 0 MR: 3 SR: 7 CR: 0 NI: 0	LR: 2 MR: 8 SR: 0 CR: 0 NI: 0	LR: 3 MR: 7 SR: 0 CR: 0 NI: 0	LR: 9 MR: 1 SR: 0 CR: 0 NI: 0	LR: 8 MR: 1 SR: 1 CR: 0 NI: 0	LR: 8 MR: 0 SR: 2 CR: 0 NI: 0	LR: 10 MR: 0 SR: 0 CR: 0 NI: 0	LR: 40 MR: 20 SR: 10 CR: 0 NI: 0

LR= low risk, MR= moderate risk, HR= serious risk, CR= critical risk, NI= no information
= Overall moderate-to-serious risk of reporting bias

Table 5. Summary of modified Jadad score of included studies.

No.	Randomisation	Randomisation appropriate?	Blinding	Blinding appropriate?	Withdrawals and dropouts	Exclusion/ inclusion criteria	Adverse effects measure	Statistical analysis	Score
1	0	0	0	0	1	0	0	1	*2*
2	1	1	1	1	1	1	0	1	*7*
3	1	1	0	0	1	1	0	1	*5*
4	1	1	0	0	1	1	0	1	*5*
5	1	1	0	0	1	1	0	1	*5*
6	1	1	0	0	1	1	0	1	*5*
7	0	0	0	0	1	0	0	1	*2*
8	0	0	0	0	1	1	0	1	*3*
9	0	0	0	0	1	1	0	1	*3*
10	0	0	0	0	1	1	0	1	*3*
11	0	0	0	0	1	0	0	1	*2*
12	0	0	0	0	1	0	0	1	*2*
13	0	0	0	0	1	0	0	1	*2*
14	0	0	0	0	1	0	0	1	*2*
15	0	0	0	0	1	1	0	1	*3*

Scores of 0-3= poor to low quality and 4-8= good to excellent quality
= Good to excellent quality

Table 6. Summary of included study limitations.

No.	Author (country)	Limitations
1	Ha and Ng, 2017 (Hong Kong)	• Used a localised measure of BMD via DEXA rather than whole-body scans and so reduces comparability of results between other studies. • The time of measurement at follow-up differed between participants. • No clear description of exclusion/inclusion criteria or adverse effects measure. • Lack of relevant covariates measured: food intake (i.e., calcium), vitamin D and body composition (i.e., FM or LM). • Habitual PA relied on self-reported calendar-based data, rather than an objective measure (i.e., accelerometer). • ExG groups exercise routine was not clearly defined regarding, frequency, load and intensity.
2	Daly et al., 2016 (Australia)	• Unable to precisely measure/quantify the PE conducted by the classroom teachers, hence the CG may have received less PE than reported. • Lack of info on the participants sporting activities outside of school. • No measure of Vitamin D status. • High attrition rates between follow-ups. • Non-measure of adverse effects. • Use of a homogenous population (mainly white children from an affluent well-developed country), thus results might not be generalisable to other populations.
3	Nogueira et al., 2014 (Australia)	• Only 13% of the total sample had bone measures taken. • No measure of Vitamin D status. • Determinants of fitness, max vertical jump and VO2 max, were significantly different between the IG and CG at BL and so could have affected results • No info reported on the concealment of allocation for randomisation process. • Relied on self-reported data for habitual PA and food intake. • Non-measure of adverse effects.
4	Meiring et al., 2014* (South Africa)	• High attrition rate between BL and follow-up. • Use of a homogenous population (black pre-to-early pubertal children), thus results might not be generalisable to other populations. • Lack of relevant covariates measured: food intake (i.e., calcium) and vitamin D. • PA history relied on self-reported calendar-based data, rather than an objective measure (i.e., accelerometer). • No info reported on the concealment of allocation for randomisation process. • Non-measure of adverse effects.
5	Weeks, Young and Beck, 2008, (Australia)	• No info reported on the concealment of allocation for randomisation process. • No measure of Vitamin D status • Non-measure of adverse effects.

Table 6. *(Cont. from previous page).*

No.	Author (country)	Limitations
6	Wiebe et al., 2008 (Australia)	• No info reported on the concealment of allocation for randomisation process. • Non-measure of adverse effects. • No measure of Vitamin D status. • The two IG groups trained side-by-side, introducing a risk of performance bias.
7	Dowthwaite et al., 2019 (USA)	• Allocation of intervention based on school site (geographic basis only), hence groups may have been different at BL. • Only a third of potential students were enrolled in the study. • Different observers assessed intervention effect during year 1 and 2, introducing a risk of observer bias. • Statistical analysis did not allow for evaluation of time-varying covariates. • Habitual PA relied on self-reported calendar-based data, rather than an objective measure (i.e., accelerometer). • No clear description of exclusion/inclusion criteria or adverse effects measure.
8	Bernardoni et al., 2014 (USA)	• Small sample sizes for each group may have been underpowered to observe a significant difference for specific bone sites. • Use of a homogenous population, thus results might not be generalisable to other populations. • Difference in measurement interval time between BL and follow-up for the two groups. CG had a longer time between DEXA scans and so had more time for bone growth. • Non-measure of adverse effects. • Short study length.
9	Exuperio et al., 2019 (Brazil)	• Small sample size hindered the ability to adjust for multiple potential confounders and may be underpowered to observe a significant difference. • Lack of relevant covariates measured: food intake (i.e., calcium), vitamin D and habitual PA. • Non-measure of adverse effects.
10	Burt et al., 2013 (Australia)	• Trend for ExG LG to have an early follow-up and CG to have a late follow-up, hence the CG may have undergone additional bone accrual. • No measure of Vitamin D status. • No description of adverse effects measure. • Short study length.
11	Maïmoun et al., 2013b (France)	• ExG groups exercise routine was not clearly defined regarding, frequency, load and intensity. • No clear description of exclusion/inclusion criteria or adverse effects measure. • Lack of relevant covariates measured: food intake (i.e., calcium) and vitamin D.
12	Pikkarainen et al., 2009 (Finland)	• Participants in the two experimental groups were at different time points in their careers. Gymnasts were at the peak of their career at the beginning of the study and had been in their sport (i.e., exercising) for longer than the runners. • Throughout the study CG participants increased exercise volume and at the final follow-up many CG participants exercised more than the athletes in the experimental groups. • Information regarding PA level was measured retrospectively between 36 and 84 months and not in the follow-up measurement every 6 months. • No clear description of exclusion/inclusion criteria or adverse effects measure. • Relied on self-reported data for habitual PA and food intake.
13	Barrack et al., 2011 (USA)	• Participant's menstruation and training data between BL and follow-up recorded retrospectively, may be a source of data error. • No clear description of exclusion/inclusion criteria or adverse effects measure. • Lack of relevant covariates measured: vitamin D or calcium. • Non-use of a CG (i.e., no PA group) for comparison.
14	Ito et al., 2017 (Brazil)	• Lack of info regarding ExG training, only mins/week of training was reported in text. • Lack of relevant covariates measured: food intake (i.e., calcium), vitamin D and habitual PA. • Minimum sample size not reached at follow-up for one ExG. • The use of maturity offset to measure pubertal status, as it was developed in a European sample and in a different age range (11–15 years), however Tanner staging could not be used due to cultural barriers. • No clear description of exclusion/inclusion criteria or adverse effects measure. • Short study length.
15	Ducher et al., 2011 (USA)	• ExG groups exercise routine was not clearly defined regarding, frequency, load and intensity. • Lack of relevant covariates measured: food intake (i.e., calcium), vitamin D and habitual PA. • Non-measure of adverse effects. • Non-use of a CG (i.e., no PA group) for comparison.