



Published in final edited form as:

Transl J Am Coll Sports Med. 2019 June ; 4(11): 74–83.

A 2-yr, School-Based Resistance Exercise Pilot Program Increases Bone Accrual in Adolescent Girls

Jodi N. Dowthwaite, PhD¹, Deena M. Weiss, MD, MPH², Jill Thein-Nissenbaum, PT, DSc³, Tamara A. Scerpella, MD^{1,4}

¹Department of Orthopedic Surgery, SUNY Upstate Medical University, Syracuse, N.Y.

²University of Wisconsin School of Medicine and Public Health (UWSMPH), Madison, WI

³Department of Family Medicine and Community Health, UWSMPH, Madison, WI

⁴Department of Orthopedics and Rehabilitation, UWSMPH, Madison, WI

Abstract

Purpose: The current analysis evaluates cumulative benefits after year two (Y2) of a school-based resistance training intervention.

Methods: Adolescent girls were enrolled and measured at the beginning of 6th grade (baseline, BL) and again at 1st follow-up (FU1: Y1 end) and 2nd follow-up (FU2: Y2 end). School gym classes met alternate school days. Site 1 had standard gym classes (CON). Site 2 gym classes included 8-12 minutes of resistance training (INT); INT girls were categorized based on observed participation effort and time (LO, HI). Measurements included: 1) height and weight; 2) questionnaires to assess extracurricular exercise and diet (calcium, vitamin D); 3) dual-energy X-ray absorptiometry (DXA, Lunar Prodigy). Whole body less head (SUB) scans yielded bone mineral content (BMC) and body composition. Lumbar spine (L1-L4) and femoral neck (FN) scans yielded BMC and areal bone mineral density (BMD); radius scans yielded ultradistal and 1/3 BMD. ANCOVA compared group means for percent gains from BL to FU2, accounting for biological maturity, BL height, height change, inter-scan interval, organized activity, calcium and vitamin D.

Results: In 62 girls (21 CON, 41 INT), intention to treat analyses detected INT vs. CON advantages for L1-L4 BMC and BMD (4.1%, 5.6%; $p < 0.05$). HI effort participants ($n = 19$) demonstrated advantages for BMC and BMD at L1-L4 and FN (5.7% to 8.2%, $p < 0.01$) vs. CON.

Conclusions: Over two school years, this resistance intervention yielded lumbar spine advantages; enthusiastic participation (HI) yielded lumbar spine and femoral neck advantages. Further work is warranted to evaluate benefit persistence after intervention cessation.

Keywords

exercise; DXA; osteoporosis prevention; pediatric; female; bone growth

Introduction

For the past two decades, investigators have attempted to define osteogenic youth exercise, with the goal of optimizing skeletal gains during growth (1–5). Skeletal gains acquired during childhood and adolescence likely persist to improve adult bone health (6–12). Certain activities, such as artistic gymnastics and soccer have been associated with greater bone mass, geometry and/or density at various sites (6, 13–21). Many highly osteogenic activities impart high impact and/or “odd” impact loading (18, 19)(18, 19, 19), such that tissue strains exceed habitual levels and/or are out of the plane of “normal” movements. Accordingly, attempts to identify broadly generalizable, easily implemented osteogenic exercise regimens have focused on programs utilizing impact as a main form of exercise. Interventions testing school-based jumping programs in pre- and peri-pubertal boys and girls have been associated with small to moderate advantages in bone acquisition over 7-20 month periods (1–5).

Our team has previously reported benefits in bone acquisition in 6th grade girls exposed to a school-based resistance training intervention, relative to age matched controls (22). We tested this program because of its potential generalizability and ease of administration in a group setting, using simple low-cost equipment including resistance bands, hand-held weights and medicine balls. In addition, we were attracted to the potential for the exercise program to affect all skeletal sites, not just the weight bearing sites typically improved by jumping programs. However, after the initial 8-month intervention, significant benefits were identified only in a subsample of high effort participants who were Tanner breast stage 2 or 3 at baseline (22). We noted significantly greater gain in lumbar spine BMD in Tanner 3 girls (4%) and femoral narrow neck width (19%) in Tanner 2 girls compared to maturity-matched controls (22). We hypothesized that continued exposure to this resistance training protocol would yield significant skeletal advantages at clinically important sites, including lumbar spine, proximal femur and distal radius. Therefore, we undertook the present analysis to evaluate skeletal benefits following a second, consecutive year of this school-based exercise intervention.

Methods

Study protocols were approved by our University’s Institutional Review Board, in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). We enrolled sixth-grade girls from a local, Midwestern, suburban school district with two geographically separated middle schools (22). Students were recruited during 6th grade orientation at each school in late summer. The PI and other members of the research team sat at a table during each orientation session and provided verbal and written information regarding the study. Potential participants provided their contact information and were given a copy of the consent document. They were formally enrolled when they attended a baseline measurement session. As study participants were minors at enrollment, they provided informed assent, accompanied by parental informed consent.

Study measurements were obtained at the beginning of sixth grade (baseline, BL), approximately eight months later at the completion of the school year (follow-up 1, FU1),

and again one year later at the completion of the seventh grade school year (follow-up 2, FU2). Subjects were assigned to the intervention or control group based upon geographically determined attendance at one of two middle schools in that district; formal randomization beyond chance geographic distribution was not carried out (1st school= control (CON); 2nd school= intervention (INT)). The two schools draw from a school district with the following racial/ethnic composition: 85% white, 4% black, 4% Hispanic, 5% Asian, 2% mixed (23).

Intervention

At baseline, all participants were transitioning from elementary to middle school and began attending one of the two participating middle schools for the first time as sixth graders. At both schools, required gym classes met two or three times per week (alternating school days) and included traditional gym class activities like basketball, volleyball, and dodgeball. At the intervention school, the focal resistance training intervention was also incorporated at the beginning of most gym class sessions, as part of the standard regimen for all students at this school (regardless of study enrollment). Accordingly, the only extra time commitments and protocols for study participants vs. non-participants were baseline and follow-up measurement sessions. The resistance training program utilized exercise bands, hand held weights, medicine balls and body weight to achieve progressive overload in all major muscle groups (see Video, SDC 1, Resistance Training Intervention, 1 minute 40 seconds, 3.96MB). In general, exercises focused on whole-body, multi-planar, functional movements. Specific content varied day-to-day and included lunges, body rows, push-ups, pull-ups, planks and other combined arm and leg resistance exercises. Occasionally, jumps/hops were included for less than 1-2 minutes.

The intervention activity sessions lasted 8-12 minutes and included 2-4 exercises at each of 4 stations (1 minute per exercise, 40 seconds on, 20 seconds off, for each exercise). Pre-recorded bells provided timing at 40 and 20 seconds each minute, played in conjunction with high-energy music. Progressive overload was achieved by the introduction of higher intensity modifications of exercises as the school year progressed and by the incorporation of increasingly stiffer elastic bands and heavier hand-held weights. Increases were made at student discretion under the guidance of a supervising physical education instructor and were individualized in order to provide a safe, but challenging workout for each student. Each session was preceded by a 4-6 minute floor warm up, including static and dynamic activities to engage core, arm and leg muscles. The four physical education instructors at the intervention school were well versed in this program, as each had been using it as part of the middle school curriculum for the previous 2-3 years.

The intervention activity began approximately one month after physical education classes commenced in the fall, and concluded with the onset of warm weather, allowing outdoor activities in the spring. Initial familiarization with techniques and equipment required 3-4 weeks additional 'ramp-up' time. Therefore, the intervention spanned a period of 6 months, from November 1 to May 1, minus school vacation weeks, providing 24 weeks of intervention activity exposure per school year. Because gym classes at the elementary school did not incorporate resistance band exercises, and because it is unusual for young girls to

perform these exercises on their own, the vast majority of participants reported that they had not engaged in formal resistance training at any point prior to baseline measurements.

Participants

Female study participants were recruited from an upper-Midwestern suburban school district. The total number of female students entering the 6th grade was 112 girls at the intervention school and 97 girls at the non-intervention school; we enrolled as many students from each site as were willing to participate. Ethnic/racial composition of the included sample was: 88.7% white, non-Hispanic; 4.8% mixed race; 4.8% Asian; and 1.6% Hispanic. Participants ranged in age from 11.0-12.1 years at BL and 12.6-13.7 years at FU2.

Physical Maturity Assessment

In girls, the most rapid period of bone accrual occurs in the 4 years roughly centered at menarche (24). Our rationale for maturity assessment is based on this biological foundation, as well as our prior work confirming the value of menarche as a maturational milestone for centering longitudinal studies of bone accrual (7, 25). Thus, in this study, menarche date was recorded to inform assignment of maturity group based on relative bone accrual progression. Self-assessed Tanner stage is unreliable for precise pubertal staging at higher maturity levels, but for pubertal onset in girls, self-assessment and clinical assessment exhibit strong agreement (26). Accordingly, we used self-assessed Tanner stage as the basis for distinguishing between pre-pubertal and pubertal status (an indicator of estrogen exposure), and we used menarche as a discrete marker reflecting peak bone mineral accrual velocity. At baseline and both follow-up sessions, all participants provided self-assessed Tanner breast and pubic staging, based on line drawings with short descriptions. In addition, menarche date was queried and recorded at each measurement session until menarche was achieved. On this basis, we divided participants into three biological maturity groups, as follows: PRE= pre-menarche at BL and FU2; PERI= pre-menarche at BL, post-menarche at FU2; POST= post-menarche at BL and FU2.

Intervention Compliance

To minimize the risk that poor compliance or attendance would confound results, our study design included, a priori, a plan to quantify intervention participation time and to record participation effort. A single member of our research team observed all intervention physical education classes over the course of the first year (Year 1, from baseline to FU1), documenting daily effort and participation minutes for each study participant (22). Although participation in the intervention was mandatory at the intervention school, not every student in each physical education class was enrolled in our study. Therefore, close observation of study participants was feasible, as each physical education class contained only 2-9 of our study participants.

Effort was qualitatively assessed for every intervention activity session, categorizing effort for each participant as low (=1), medium (=2) or high (=3) based on observation. Participant effort was rated as “low” for those who: a) stopped exercising when the teacher was not looking; b) decreased exercise time by moving slowly between stations; and/or c) exercised with low vigor. In contrast, participant effort was rated as “high” for those who: a) exercised

regardless of teacher observation; b) maximized exercise time for each station; and c) participated with vigor. Participant effort was rated as “medium” if they fell somewhere between the two extremes or demonstrated inconsistent effort during an intervention session. For each session, individual numeric effort scores were multiplied by the number of participation minutes (effort*minutes). All effort*minutes session scores were summed to create subject-specific totals. On this basis, effort and failure to participate for any reason were incorporated into the summed effort score (e.g. absence, illness, injury, unprepared status).

Subject-specific effort*minutes totals were used to determine a dichotomous grouping variable based upon total effort during Year 1: low effort (LO, current n=22) mean= 774 effort*minutes (range, 381-1080); high effort (HI, current n=19) mean= 1561 effort*minutes (range, 1155-1982). During Year 2 (7 months directly preceding FU2), intermittent class observations were performed over the course of the school year to evaluate effort agreement with Year 1, and physical education instructors were consulted to quantify intervention minutes. Based on this information, at FU2, all but one subject, who missed significant time due to illness, was classified in the same effort group as for FU1; therefore, effort groupings were maintained from FU1 to FU2 for statistical analyses.

Densitometry and Other Measurements

At BL, FU1, and FU2, total body and regional DXA scans were performed by a single International Society for Clinical Densitometry (ISCD) certified technologist using a GE Healthcare Lunar iDXA densitometer (Madison, WI). Using enCore software (version 13.31), total body scans assessed fat mass for whole body (WB) and non-bone lean mass (nbFFM) for whole body less head (SUB). Percent body fat was calculated for WB (fat mass/total mass). Regional scans provided bone mineral content (BMC) for SUB, lumbar spine (L1-L4) and femoral neck (FN); areal bone mineral density (BMD) was evaluated for L1-L4, FN, ultradistal radius (radUD) and 1/3 radius (rad1/3) regions of interest.

To calculate age-specific CVs, duplicate same-day scans are required; this practice is uncommon in growing children and adolescents, as it increases risks from greater radiation exposure to rapidly growing tissues. Accordingly, to evaluate quality control, we relied on coefficients of variation (CVs) for DXA variables calculated using duplicate scans of 30 postmenopausal females: femoral neck BMD and total-body BMC CVs were <1%; lumbar spine BMD and FN BMC CVs were <2.0%; radius BMD CVs were <4% (22). CVs for lean and fat mass were <1%, assessed in young adult females, aged 18-23 years (whole body scans only) (27).

Anthropometrics and other data were collected contemporaneous with DXA scans, as follows. Height was measured, in stocking feet, via wall-mounted stadiometer (Holtain Model 602VR, Harpenden), to the nearest 0.1 cm, and weight was measured in light clothing via electronic force plate (Leonardo Mechanograph® GRFP STD, Novatec), to the nearest 0.01 kg. Body mass index (BMI, kg/m²) was calculated from resultant mass and height. Forearm length was measured with a ruler to the nearest 0.1 cm, from the tip of the olecranon to the tip of the ulnar styloid. A validated, pediatric, semi-quantitative food frequency questionnaire was completed to assess food consumption habits (28), yielding

intakes of calcium and vitamin D, including supplementation. For the present analysis, mean intakes were calculated from the results of questionnaires administered at FU1 and FU2, representing habitual intakes between BL and FU2. With assistance from parents, subjects used a calendar-based form to report hours per week of participation in non-aquatic organized activity, excluding gym class time, over the same intervals (BL to FU1, FU1 to FU2), allowing calculation of mean background organized activity dose (h/wk) from BL to FU2 (22). In prior research, 12 month activity means from a similar population correlated strongly with coaches' logs ($r>0.97$, $p<0.0001$) (14).

Statistical Analyses

A priori power calculations were based on cross-sectional gymnast vs. non-gymnast comparisons for a group of girls of similar age to the cohort studied here, at two effect size levels, derived from a longitudinal, observational study (7, 29). We hypothesized that the intervention would yield benefits comparable to 50-75% of the effect sizes observed in gymnasts versus non-gymnasts. At 50% of observed gymnast advantages, 10-31 girls per group would be required to achieve at least 80% power to detect significant differences between intervention participants and controls. At 75% of observed gymnast advantages, cell sizes of 10-14 would provide at least 80% power to detect significant differences.

SPSS v. 23 was used to perform statistical analyses (IBM, Armonk, NY), employing $\alpha=0.05$. Normality of distributions was evaluated using Kolmogorov-Smirnov tests. Where appropriate, dependent variables were ln-transformed for analysis, or Kruskal-Wallis tests were used to evaluate group differences in descriptive statistics. Otherwise ANOVA or RM ANOVA were used to assess group differences, as appropriate. For ANOVA and RM ANOVA, Levene's test was used to evaluate equality of error variances. Intervention status was coded for comparisons as follows: CONTROL (CON); INTERVENTION (INT); LOW EFFORT (LO) and HIGH EFFORT (HI). Intention to treat analyses compared results for CON vs. INT; to evaluate intervention success with high compliance, we compared CON vs. HI. To evaluate potential bias in maturational status, chi square analysis was used to evaluate associations between maturity status and intervention status (INT vs. CON; CON vs. LO vs. HI), as well as race/ethnicity and intervention status. We used Pearson correlations to assess the potential relationship between inter-scan interval (date difference, BL to FU2, years) and dependent variables.

To minimize potential for Type 1 error through multiple testing, dependent variables were limited to the following: sub-head BMC (SUBBMC), radUDBMD, rad1/3BMD, FNBMC, FNBMD, L1-L4BMC and L1-L4BMD. Both BMC and BMD were evaluated for L1-L4 and FN to allow comparison of our results to prior research (see discussion). Included covariates were: maturity status, inter-measurement background organized physical activity exposure (excluding gym class), mean inter-scan calcium intake, natural logarithm of mean vitamin D intake, inter-scan interval, baseline dependent variable (BLDV), baseline height (BLHeight) and inter-scan change in height (FU2 minus BL).

Results

Of the original 68 participants enrolled and measured at baseline (BL), 62 were measured at FU2 and form the basis of this analysis. Descriptive statistics are presented in Table 1 for the total sample (n=62), and separately, for the CON (n=21), LO (n=22) and HI (n=19) groups. (The 6 study participants who did not return for FU2 included 2 CON and 4 INT.) As the two schools have slightly different race/ethnicity profiles for their student bodies, we tested for differences in CON vs. INT profiles; among individuals returning for FU2, racial and ethnic groups were distributed evenly among control vs. intervention and control vs. LO vs. HI groupings (chi square $p>0.91$).

Of the 62 participants included in the present analyses, most had not achieved menarche at BL in the fall of their 6th grade year; by FU2, most were post-menarcheal. In terms of pubertal onset, 56/62 had entered puberty at baseline (self-reported Tanner breast or pubic stage II or higher); few girls were pre-pubertal at BL [1 CON (6%), 3 LO (14%) and 2 HI (10%)]. As described in the methods section, to reflect maturational progression over the full study period, participants were classified in one of 3 groups: pre-menarcheal at BL and FU2 (PRE, n=21), pre-menarcheal at BL and post-menarcheal at FU2 (PERI, n=32) or post-menarcheal at BL and FU2 (POST, n=9).

There was no significant difference in menarche status by group (INT vs. CON, chi square $p>0.09$; CON vs. LO vs. HI: chi square $p>0.30$). However, there was a trend for more CON to be classed as PRE or POST compared to INT [CON: PRE= 43%, PERI= 33%, POST= 24%; INT: PRE= 29%, PERI= 61%, POST= 10%]; the majority of INT were classed as PERI. On this basis, we included maturity status as an independent variable in regression analyses and ANCOVA.

Table 2 shows raw data for dependent variables at BL and FU2; no GROUP differences were detected for BL dependent variables using ANOVA ($p>0.14$). Likewise, from BL to FU2, no significant GROUP differences were detected for mean inter-scan values for physical activity level (Kruskal-Wallis $p>0.53$), calcium intake or ln vitamin D intake (ANOVA $p>0.37$).

There was a significant difference among intervention groups for inter-scan interval (Kruskal-Wallis, $p<0.001$), with both LO and HI having longer inter-scan intervals than CON (LO=19.3+0.7mos, HI=19.2+0.8mos, CON=18.5+0.7mos). However, this interval disparity represented only ~4% of the total inter-scan interval (~3 weeks). In addition, from BL to FU2, no significant GROUP (LO, HI, CON) differences (RM ANOVA: $p>0.23$) or GROUP*TIME interactions ($p>0.12$) were detected for height, weight, BMI or percent body fat, or for age at either time point (Kruskal-Wallis $p>0.08$), suggesting that body growth was similar across the inter-scan interval for all groups, and was not likely to have confounded the significant group differences identified in bone accrual. Nonetheless, we evaluated this potential source of bias by including inter-scan interval and change in height as independent variables in ANCOVA; inter-scan interval exerted no significant influence as a covariate in any of the models ($p>0.20$).

Intervention participation was quantified for each individual during Year 1. Mean participation time over the 24 weeks of the intervention was 22.3 minutes/week. Similarly, when only the HI group was considered, mean participation time was 22.1 minutes/week. Thus, the division of participants into HI and LO effort groups (effort minutes=effort*minutes) was mainly a function of observed effort, as mean exposure time was not greater in HI than LO.

Intention to treat ANCOVA (INT vs. CON) indicated that intervention exposure conferred site-specific advantages in gain of both BMC and BMD at L1-L4, regardless of intervention effort or participation time (effort*minutes), ($p < 0.05$, 5.6% and 4.1%, respectively) [Figure 1]. No covariates exerted significant individual influence (all $p > 0.22$).

To quantify differences in percentage gains between BL and FU2 as a function of high intervention compliance, we used ANCOVA to compare adjusted group means for HI vs. CON (LO excluded). Significant, site-specific intervention advantages were observed for L1-L4BMC, L1-L4BMD, FNBMC and FNBMD (respectively, $p = 0.006$, 0.005 , 0.007 , 0.006) [Table 3]. These results indicated relative advantages for HI vs. CON in percent gain, ranging from 3.9% to 7.6% for these same outcomes ($p < 0.008$) [Figure 2]. Intervention exposure appeared to confer no advantage in radius properties at either site, with a trend for intervention disadvantage in ultradistal BMD. Height change was a significant covariate ($p = 0.000$), with maturity, baseline height and inter-DXA physical activity exposure playing minor supportive roles ($p = 0.07$, $p < 0.15$); other covariates were not significantly influential ($p = 0.21$). In all dependent variable-based analyses, Levene's test indicated equivalent error variances (all $p > 0.13$, except LSBMC percent difference $p > 0.07$).

Discussion

This group of predominantly white, upper-Midwestern, adolescent girls obtained significant lumbar spine BMC and BMD benefits through two consecutive school years of a resistance training program, regardless of maturity status and background physical activity. Exposure averaged only 22 minutes per week, acquired in 1-2 sessions per week. Girls who demonstrated high intervention effort had significant advantages at both lumbar spine and femoral neck sites, gaining 5-8% more bone mass than controls. Importantly, intervention benefits were clear at sites that are prone to clinically devastating fragility fractures later in life (femoral neck, lumbar spine).

Our results are similar to those of other investigative groups who have evaluated various interventions in pre-, peri- and post-pubertal girls over interventions spanning 7 to 20 months (1–5). Following a school-based, 10-month controlled intervention in 9-10 year old girls, using aerobic impact and weight training, Morris et al. identified differential percent gains in total body (5.5%) and femoral neck (6.5%) BMC (4). However, their intervention involved 30-minute sessions, three times weekly, averaging more than 4 times the weekly intervention exposure time of our protocol. It is possible that our protocol would yield even greater benefits, if comparable exposure times were achieved. Fuchs et al. identified differential gains of 4.5% and 3.1% for femoral neck and lumbar spine BMC, respectively, in a combined sample of pre-pubescent boys and girls who participated in a 7-month

jumping intervention (100 jumps, 3 times weekly) (2). MacDonald et al. identified a 5.4% advantage in femoral neck strength in a subset of highly compliant 9-11 year old girls who participated in eleven months of a school-based exercise program, monitored over a 16-month period (5). Weeks et al. evaluated bone mass in peri- and post-pubertal boys and girls (mean age 13.8 + 0.4 yrs) following an 8-month school-based RCT that included 10 minutes of jumping, twice weekly (30). They identified significant gains in FNBM, LSBMAD and LS area for girls; however, differential gains for intervention participants vs. controls were not statistically significant (FNBM: 13.9% v 4.9%, LSBMAD: 5.2% v 1.5%, and LS area: 4.9% v. 2.0%) (30). Finally, in a group of girls who were 10 years old at baseline, MacKelvie et al. identified significant advantages in lumbar spine (3.7%) and femoral neck (4.6%) BMC gains following a high-impact, circuit-based, jumping intervention (10 minutes, 3 times a week) over 20 months (2 school years) (1).

Although difficult to compare based on intervention variation (type, intensity, duration) and group composition (gender, maturity), our 19-month differential percent gains (14 months of intervention exposure) are greater than or equal to those previously reported following 7-11 months of intervention exposure (16 months observation). Particularly when the amount of minutes engaged in the intervention is considered, our results spanning two 7-month exposure periods are notable. Moreover, our results were accomplished using resistance bands and light hand-held dumbbells, with minimal impact loading, in contrast to the previously published interventions that focused on jumping.

The group of girls included in the present analysis is not identical to the group in which we previously reported 8-month results (22). For the prior analysis, we specifically evaluated a smaller subset of the whole group, comparing HI effort participants versus controls in girls who were Tanner breast stage 2 or 3 at baseline (T2&T3, n= 48). In that subset, we identified advantages in differential gains for T3 girls at the lumbar spine (BMD, 4%) and for T2 girls at the femoral neck (width, 19%) ($p<0.03$) (22). The present analysis includes the entire group of girls who provided both baseline and 19-month follow-up data (n=62), accounting for maturational variation on the basis of menarche status progression (PRE, PERI, POST groups). In the current analysis, the significant benefits observed in intervention subjects across the total sample (lumbar spine BMC and BMD) may be attributable to cumulative benefits of a second year of intervention exposure or to exposure during a more favorable maturity phase for adaptation, when estrogen levels were likely higher. Forty-four girls in the present analysis were post-menarche by the completion of Year 2 (44/62: 71% of Year 2 analysis sample), compared to only 15 girls from the previously published analysis who were post-menarcheal at the completion of Year 1 (15/38: 39% of Year 1 analysis sample). As noted above, we have previously provided evidence of greater intervention response at the lumbar spine in more mature (T3 vs. T2) participants; a similar effect may have contributed to the significant response observed following the Year 2 intervention (22). In the present analysis, maturity status was not a significant covariate in ANCOVA, likely due to collinearity with change in height, which was a highly influential covariate in both sets of analyses. Results of the present analysis do not allow us to distinguish between cumulative and/or maturity-specific intervention effects; it is likely that both factors played a role in Year 2 bone acquisition. Future analyses, when all subjects are post-menarche, will address maturity-specific responses to intervention activity more

effectively, using biological age as a continuous function to evaluate activity exposure timing relative to menarche (see limitations section).

We were concerned that background physical activity participation, separate from the resistance training intervention, might confound our results. Therefore, we carefully recorded hours per week of non-intervention organized physical activity and calculated a mean value for the study interval. We have reported significant associations with both lumbar spine and femoral neck bone outcomes in children and adolescents using mean organized activity recorded using a similar questionnaire (16, 31). We are confident that entry of this organized activity variable as a covariate in ANCOVA captured the intended physical activity exposure variation, as it explained 5% of variance in fat mass in separate regression models (data not shown: negative factor, $p=0.004$). However, organized physical activity did not serve as a significant covariate in either ANCOVA bone model. Thus, in our cohort of 6th and 7th grade girls, the focal resistance training intervention appears to have been a more influential factor in bone acquisition than background, non-intervention organized physical activity; this lack of significance may be due to lower variance in organized activity exposure and vigor in the current study than in our previous work. While use of accelerometers may provide a better indicator of background activity vigor than activity hours alone, short-term accelerometry may not reflect long-term patterns and is burdensome for longitudinal monitoring of study participants over multiple years (13). Use of other devices to measure bone stresses directly (e.g. strain gauges) is unlikely to be approved for use in healthy pediatric populations and would deter study participation.

There was no difference in reported calcium or vitamin D intake among groups. On the whole, girls consumed the current RDA for calcium (1300 mg/day) and fell below current recommendations for vitamin D intake (600 IU/day). With no significant difference in intakes among groups, it is unlikely that variation in vitamin D consumption confounded our results, especially as we accounted statistically for intake variation in our analyses. We cannot rule out variability in serum vitamin D, as disparate sun exposure may have contributed to more disparate serum vitamin D levels than are evident based on intake indices. It is possible that a greater intervention effect might have been observed if vitamin D intakes had been within the adequate range.

The current analysis was designed to assess the effect of a resistance training intervention over two school years. Effort was carefully observed and recorded for each girl during the first intervention year, but it was only “spot-checked” during Year 2 (different observer, blind to Year 1 effort status). Because “spot-checking” indicated consistent effort level within-individuals from year to year, we did not change categorization of LO and HI effort groups from Year 1 to Year 2. It is possible that some girls classed as HI effort in Year 1 may have reduced effort while not observed by researchers during Year 2, and conversely, some girls classed as LO effort in Year 1 may have increased effort while not observed by researchers during Year 2. However, this possible scenario, yielding more similar effort levels between HI and LO groups, would have been expected to reduce differences between HI and LO intervention groups; accordingly, our view is that observed significant differences between HI and LO likely provide a conservative view of benefits attributable to enthusiastic participation in this loading intervention.

The efficacy of an exercise intervention is limited by both compliance and generalizability; programs that are not easily implemented and enthusiastically completed over a sufficient term will not be successful. In the current analysis, our data suggest that approximately 50% of the intervention participants were adequately enthusiastic to derive significant benefits at both major sites of osteoporotic fracture. Future research is needed to ascertain how to motivate all participants to engage with sufficient effort to gain significant and clinically relevant benefits. The intervention studied here was performed as part of the regular, school-based physical education program. It is important to note that *all* intervention school girls participated in these exercise sessions, as they were part of the curriculum; unfortunately, we cannot evaluate effort or benefits in these girls, as they did not enroll in the study. The sessions required only 8-12 minutes of participation, 2-3 days per week, from November through April. Equipment includes light dumbbells and wall-mounted resistance bands, which are low in cost and easy to store and maintain. The program is safe, easy to teach and to individualize in a group setting, allowing progressive resistance as strength and coordination improve. Implementation on a large scale, in a variety of schools, appears feasible, but must be tested in a large study that includes schools representing a variety of racial, ethnic and socioeconomic characteristics. Additional study, including careful assessment of maturational progression during exposure, evaluating retention of benefits after intervention cessation, is needed to assess the utility of this or similar adolescent exercise interventions in improving adult bone health.

Limitations:

As is often the case in a clinical interventional study, our study was hampered by several methodological issues. First, we successfully enrolled and studied only about 1/3 of eligible students at the study schools, minimizing representation and generalizability of results. Second, a different observer assessed intervention effort during Year 1 and 2, introducing the possibility that variation in effort was underreported. Third, the current intermediate analysis uses a simplistic repeated measures design, which does not allow for evaluation of time-varying covariates, as would be ideal for a growth analysis of this nature. We plan more sophisticated analyses when additional follow-up data have been accumulated, evaluating relevant factors that change across time (e.g. height, weight, body composition) and accounting for biological age as a continuous variable. This strategy will improve statistical power and provide us with greater capacity to evaluate the potential influence of disparate rates of growth/maturation and uneven inter-scan interval upon bone accrual rates; the current analysis was underpowered to completely address this concern and likely underpowered overall. Fourth, participants were assigned to the intervention by school site (geographic basis only); while it is possible that this methodology may have resulted in the systematic influence of factors such as socioeconomic status or racial/ethnic variation on variables of interest, our racial/ethnic chi square results do not support this concern. Finally, our metric of background physical activity relies on self-reported calendar-based data, rather than accelerometric measures; as noted, we are confident that these data capture relevant variance in organized activity exposure, as they explained significant variance in follow-up fat mass, accounting for key covariates. The observed negative association with fat mass, without a significant association with bone properties, suggests that there may have been

inadequate variance in osteogenic activity separate from the intervention to detect “background activity” bone benefits in the current sample.

Conclusion

This study provides evidence of significant benefits in lumbar spine properties as a result of a simple school-based exercise intervention, even in participants who expended low effort. Higher levels of participation compliance were associated with significant benefits at both lumbar spine and femoral neck. Observed benefits, the ease of participation and the potential generalizability of the intervention make this pilot program an attractive possibility for bone-building youth exercise. Further study is necessary, including a larger randomized trial to determine whether broad implementation is feasible and yields significant benefits under more stringent conditions. Additional follow-up is needed to determine whether long-term benefits are realized.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

Conception and study design: TAS. Data collection: Kristen Hendrickson and Brittney Bernardoni. Data analysis: JND. Data interpretation: JND, TAS, DMW, JTN. Drafting: JND. Revising critically: JND, TAS, DMW, JTN. Final approval of manuscript: JND, TAS, DMW, JTN. Responsibility for integrity: JND, TAS, DMW, JTN. The authors are deeply grateful to Kristen Hendrickson and Dr. Brittney Bernardoni who worked as study coordinators for the first and second year of this longitudinal study, respectively; their dedication provided meaningful data and remarkable study retention. We appreciate the expertise and dedication of Jessie Libber, our study DXA technician. The results of the present study do not constitute endorsement by ACSM.

Conflicts of Interest and Source of Funding: The authors have no conflicts of interest to disclose. We acknowledge funding support from the UW Institute for Clinical and Translational Research (Clinical and Translational Science Award, NIH/NCATS 9U54TR000021) and from the UWHC Sports Medicine Classic Fund.

References:

1. MacKelvie KJ, Khan KM, Petit MA, Janssen PA, McKay HA. A school-based exercise intervention elicits substantial bone health benefits: A 2-year randomized controlled trial in girls. *Pediatrics*. 2003 12;112(6 Pt 1):e447. [PubMed: 14654643]
2. Fuchs RK, Bauer JJ, Snow CM. Jumping improves hip and lumbar spine bone mass in prepubescent children: A randomized controlled trial. *J Bone Miner Res*. 2001 1;16(1):148–56. [PubMed: 11149479]
3. Weeks BK, Young CM, Beck BR. 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 7;23(7):1002–11. [PubMed: 18302501]
4. Morris FL, Naughton GA, Gibbs JL, Carlson JS, Wark JD. Prospective ten-month exercise intervention in premenarcheal girls: Positive effects on bone and lean mass. *J Bone Miner Res*. 1997 9;12(9):1453–62. [PubMed: 9286762]
5. Macdonald HM, Kontulainen SA, Petit MA, Beck TJ, Khan KM, McKay HA. Does a novel school-based physical activity model benefit femoral neck bone strength in pre- and early pubertal children? *Osteoporos Int*. 2008 10;19(10):1445–56. [PubMed: 18357403]
6. Scerpella TA, Bernardoni B, Wang S, Rathouz P, Li Q, Dowthwaite JN. Site-specific, adult bone benefits attributed to loading during youth: A preliminary longitudinal analysis. *Bone*. 2016;85:148–59. [PubMed: 26826335]

7. Scerpella TA, Dowthwaite JN, Rosenbaum PF. Sustained skeletal benefit from childhood mechanical loading. *Osteoporos Int.* 2011 9 14;22:2205–10. [PubMed: 20838772]
8. Baxter-Jones AD, Kontulainen SA, Faulkner RA, Bailey DA. A longitudinal study of the relationship of physical activity to bone mineral accrual from adolescence to young adulthood. *Bone.* 2008 12;43(6):1101–7. [PubMed: 18725335]
9. Ducher G, Eser P, Hill B, Bass S. History of amenorrhoea compromises some of the exercise-induced benefits in cortical and trabecular bone in the peripheral and axial skeleton: A study in retired elite gymnasts. *Bone.* 2009 10;45(4):760–7. [PubMed: 19573632]
10. Eser P, Hill B, Ducher G, Bass S. Skeletal benefits after long-term retirement in former elite female gymnasts. *J Bone Miner Res.* 2009 12;24(12):1981–8. [PubMed: 19453258]
11. Erlandson MC, Kontulainen SA, Chilibeck PD, Arnold CM, Faulkner RA, Baxter-Jones AD. Former premenarcheal gymnasts exhibit site-specific skeletal benefits in adulthood after long-term retirement. *J Bone Miner Res.* 2012 11;27(11):2298–305. [PubMed: 22714629]
12. Duckham RL, Baxter-Jones AD, Johnston JD, Vatanparast H, Cooper D, Kontulainen S. Does physical activity in adolescence have site-specific and sex-specific benefits on young adult bone size, content, and estimated strength? *J Bone Miner Res.* 2014 2;29(2):479–86. [PubMed: 23907819]
13. Dowthwaite J, Dunsmore K, Gero N, et al. Arm bone loading index predicts DXA musculoskeletal outcomes in two samples of post-menarcheal girls. *J Musculoskelet Neuronal Interact.* 2015;15(4):358–71. [PubMed: 26636282]
14. Dowthwaite JN, Scerpella TA. Distal radius geometry and skeletal strength indices after peripubertal artistic gymnastics. *Osteoporos Int.* 2011 Jan;22(1):207–16.
15. Dowthwaite JN, Rosenbaum PF, Scerpella TA. Mechanical loading during growth is associated with plane-specific differences in vertebral geometry: A cross-sectional analysis comparing artistic gymnasts vs. non-gymnasts. *Bone.* 2011 11;49(5):1046–54. [PubMed: 21839871]
16. Dowthwaite JN, Rosenbaum PF, Sames CA, Scerpella TA. Muscle function, dynamic loading, and femoral neck structure in pediatric females. *Med Sci Sports Exerc.* 2014;46(5):911–9. [PubMed: 24743106]
17. Dowthwaite JN, Rosenbaum PF, Scerpella TA. Site-specific advantages in skeletal geometry and strength at the proximal femur and forearm in young female gymnasts. *Bone.* 2012 5;50(5):1173–83. [PubMed: 22342799]
18. Nikander R, Sievanen H, Heinonen A, Kannus P. Femoral neck structure in adult female athletes subjected to different loading modalities. *J Bone Miner Res.* 2005 3;20(3):520–8. [PubMed: 15746998]
19. Nikander R, Sievanen H, Uusi-Rasi K, Heinonen A, Kannus P. Loading modalities and bone structures at nonweight-bearing upper extremity and weight-bearing lower extremity: A pQCT study of adult female athletes. *Bone.* 2006 10;39(4):886–94. [PubMed: 16731064]
20. Laing EM, Wilson AR, Modlesky CM, O'Connor PJ, Hall DB, Lewis RD. Initial years of recreational artistic gymnastics training improves lumbar spine bone mineral accrual in 4- to 8-year-old females. *J Bone Miner Res.* 2005 3;20(3):509–19. [PubMed: 15746997]
21. Pettersson U, Nordstrom P, Alfredson H, Henriksson-Larsen K, Lorentzon R. Effect of high impact activity on bone mass and size in adolescent females: A comparative study between two different types of sports. *Calcif Tissue Int.* 2000 9;67(3):207–14. [PubMed: 10954774]
22. Bernardoni B, Thein-Nissenbaum J, Fast J, et al. A school-based resistance intervention improves skeletal growth in adolescent females. *Osteoporos Int.* 2014 3;25(3):1025–32. [PubMed: 24114402]
23. More Maps, Race and Ethnicity in Middleton-Cross Plains School District, Wisconsin (Unified School District) [Internet]. Available from: <https://statisticalatlas.com/school-district/Wisconsin/Middleton-Cross-Plains-School-District/Race-and-Ethnicity>
24. McKay HA, Bailey DA, Mirwald RL, Davison KS, Faulkner RA. Peak bone mineral accrual and age at menarche in adolescent girls: A 6-year longitudinal study. *J Pediatr.* 1998 11;133(5):682–7. [PubMed: 9821429]
25. Bernardoni B, Scerpella T, Rosenbaum P, et al. The influence of Organized physical activity (including gymnastics) on young adult skeletal traits: Is maturity phase important? *Pediatric*

- exercise science. 2015;5;27(2):285–96. Available from: May;27(2):285–96. doi: 10.1123/pes.2014-0051. PMID: [PubMed: 25386845]
26. Rasmussen AR, Wohlfahrt-Veje C, Tefre de Renzy-Martin K, et al. Validity of self-assessment of pubertal maturation. *Pediatrics*. 2015 1;135(1):86–93. [PubMed: 25535262]
 27. Krueger D, Libber J, Binkley N. Spine trabecular bone score precision, a comparison between GE lunar standard and high-resolution densitometers. *J Clin Densitom*. 2015 2 4
 28. Rockett HR, Breitenbach M, Frazier AL, et al. Validation of a youth/adolescent food frequency questionnaire. *Prev Med*. 1997 Nov-Dec;26(6):808–16.
 29. Dowthwaite JN, Flowers PPE, Spadaro JA, Scerpella TA. Bone geometry, density and strength indices of the distal radius reflect loading via childhood gymnastic activity. *J Clin Densitom*. 2007;10:65,65–75. [PubMed: 17289528]
 30. Weeks BK, Young CM, Beck BR. 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 7;23(7):1002–11. [PubMed: 18302501]
 31. Ren J, Brann LS, Bruening KS, Scerpella TA, Dowthwaite JN. Relationships among diet, physical activity, and dual plane dual-energy X-ray absorptiometry bone outcomes in pre-pubertal girls. *Arch Osteoporos*. 2017 12;12(1):19,017-0312-9.

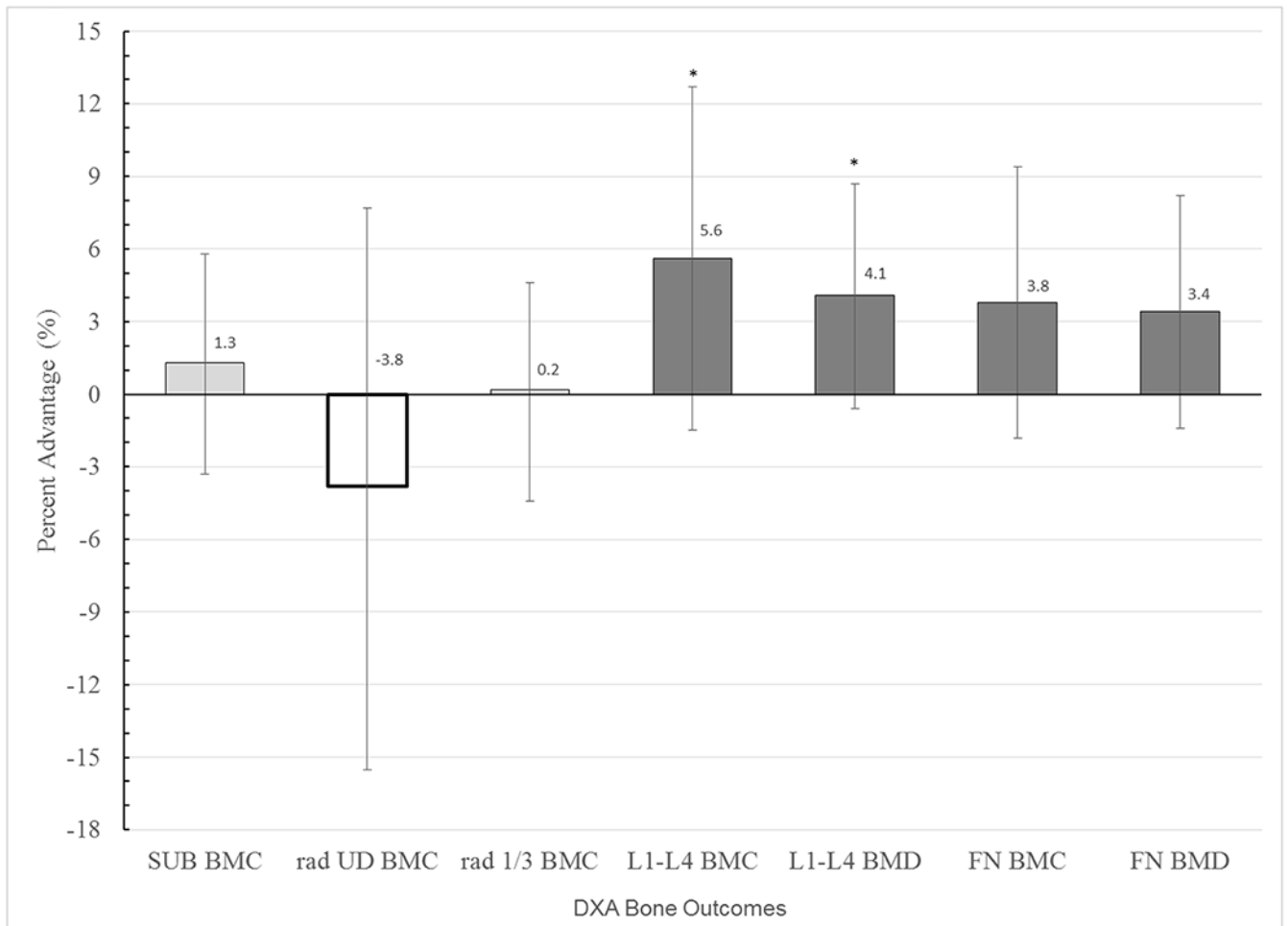


Fig. 1. Mean Bone Gain Percent Advantage for INT versus CON (intention to treat analysis) For each bone outcome, adjusted mean percent advantages in bone gain from baseline to follow-up (FU2), are presented for INT (bar) relative to CON (x axis, zero line), with 95% confidence intervals. Means are adjusted for the following covariates: inter-scan interval, Mean Non-intervention Organized Activity over the interval, Maturity level, Baseline Height, Height Change, Mean Ca and ln Mean vitamin D. * indicates $p < 0.05$

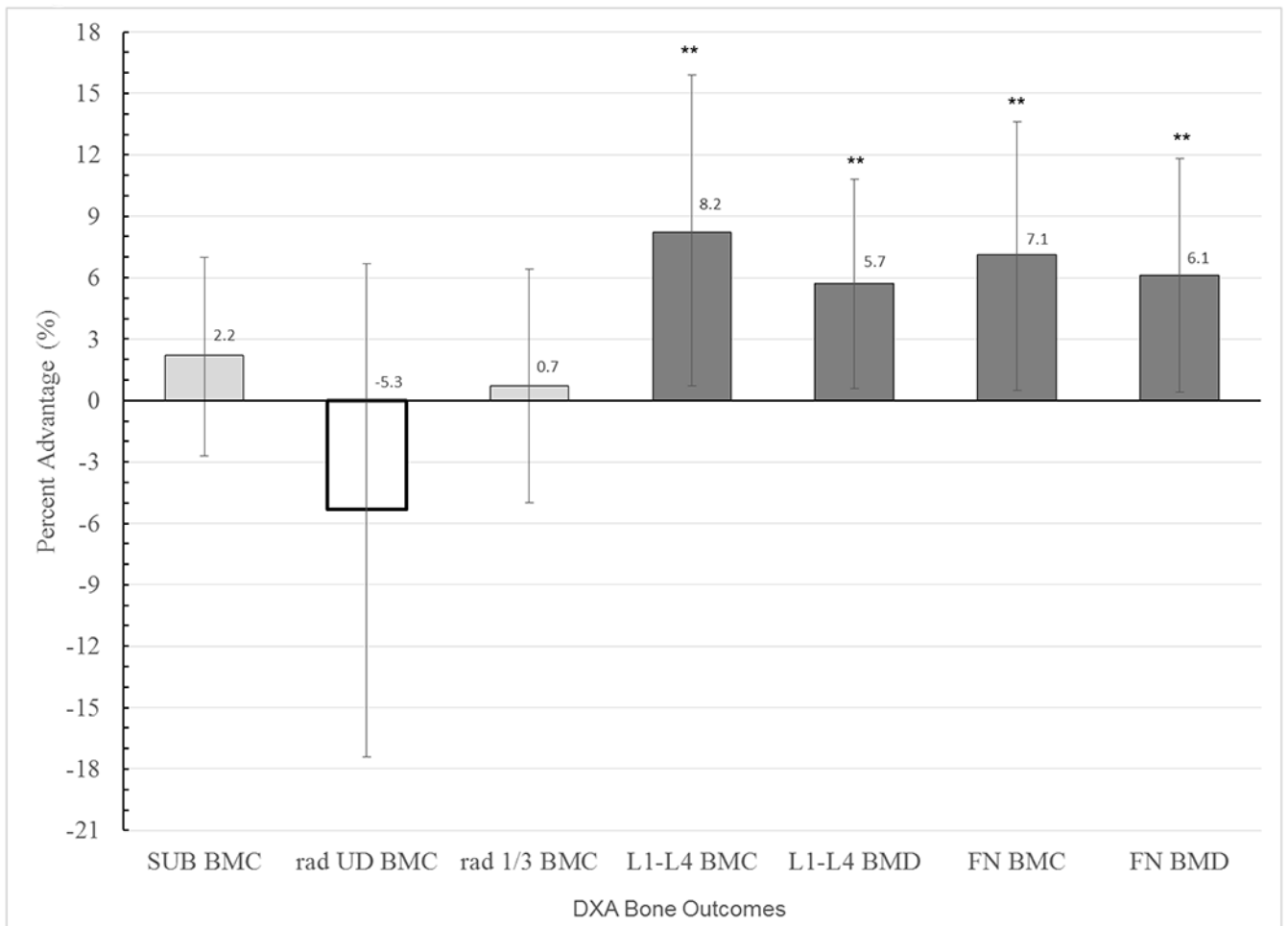


Fig. 2.

Mean Bone Gain Percent Advantage for HI versus CON (high compliance analysis)

For each bone outcome, adjusted mean percent advantages in bone gain from baseline to follow-up (FU2), are presented for HI (bar) relative to CON (x axis, zero line), with 95% confidence intervals. Means are adjusted for the following covariates: inter-scan interval, Mean Non-intervention Organized Activity over the interval, Maturity level, Baseline Height, Height Change, Mean Ca and ln Mean vitamin D. ** indicates $p < 0.01$

Table 1.

Descriptive Statistics for Independent Variables

Variable	TOTAL (n= 62)		CONTROL (n= 21)		LO-INT (n=22)		HI-INT (n=19)	
	BL	FU2	BL	FU2	BL	FU2	BL	FU2
Interval (months)	19.0 (0.8)		18.5 ^a (0.7)		19.3 (0.7)		19.2 (0.8)	
Age (years)	11.5 (0.3)	13.2 (0.3)	11.6 (0.3)	13.2 (0.3)	11.4 (0.3)	13.1 (0.3)	11.6 (0.3)	13.2 (0.2)
Height (cm)	152.2 (6.6)	161.0 (6.0)	153.1 (6.8)	161.7 (5.7)	152.4 (7.5)	162.2 (6.9)	150.9 (5.4)	159.0 (5.0)
Weight (kg)	42.9 (8.6)	51.5 (8.9)	42.7 (6.3)	51.2 (6.5)	42.7 (11.3)	51.5 (11.2)	43.2 (7.4)	51.8 (8.5)
BMI (kg/m ²)	18.4 (2.8)	19.8 (2.9)	18.2 (1.8)	19.5 (1.9)	18.2 (3.8)	19.5 (3.6)	18.9 (2.4)	20.4 (2.8)
Percent Fat (%)	27.2 (6.2)	25.9 (5.8)	25.3 (4.3)	24.7 (4.4)	28.7 (7.3)	27.1 (6.7)	27.5 (6.4)	25.8 (5.9)
Mean Ca ^b (mg)	1293 (344)		1232 (300)		1316 (345)		1333 (394)	
Mean Vit D ^b (IU)	403 (245)		396 (291)		436 (210)		374 (236)	
Mean Activity ^b (h/wk)	3.4 (2.9)		3.8 (3.1)		2.9 (2.5)		3.8 (3.3)	

^aCON inter-scan interval was significantly lower than for LO-INT and HI-INT, Kruskal-Wallis $p < 0.001$. However, significant GROUP differences were not detected for age at either time point (Kruskal-Wallis), anthropometrics (RM ANOVA: GROUP or GROUP*TIME interaction), or mean inter-scan activity and diet variables (ANOVA). For anthropometrics, significant TIME effects were detected within GROUP (RM ANOVA $p < 0.002$).

^bMean Calcium (Ca) and Vitamin D (Vit D) intake and Mean Activity are derived from self-report tools.

Table 2.

Descriptive Statistics for Unadjusted Dependent Variable Data

Variable	TOTAL (n= 62)		CONTROL (n= 21)		LO-INT (n=22)		HI-INT (n=19)	
	BL	FU2	BL	FU2	BL	FU2	BL	FU2
SUB BMC (g)	1225.9 (216.5)	1573.2 (248.1)	1261.9 (215.8)	1591.7 (240.2)	1184.1 (248.8)	1543.6 (288.0)	1234.4 (176.5)	1587.0 (214.7)
radUD BMD (g/cm ²)	0.331 (0.046)	0.364 (0.049)	0.324 (0.040)	0.360 (0.050)	0.324 (0.048)	0.362 (0.047)	0.347 (0.048)	0.371 (0.054)
rad1/3 BMD (g/cm ²)	0.653 (0.068)	0.730 (0.068)	0.650 (0.054)	0.724 (0.052)	0.645 (0.083)	0.722 (0.076)	0.666 (0.067)	0.746 (0.073)
L1-L4 BMC (g)	33.67 (8.02)	46.99 (9.54)	34.86 (9.68)	47.10 (11.39)	32.27 (8.28)	45.97 (9.75)	33.98 (5.47)	48.06 (7.15)
L1-L4 BMD (g/cm ²)	0.851 (0.117)	1.024 (0.134)	0.859 (0.137)	1.008 (0.144)	0.826 (0.116)	1.005 (0.132)	0.873 (0.093)	1.064 (0.122)
FN BMC (g)	3.46 (0.60)	4.27 (0.70)	3.57 (0.55)	4.30 (0.67)	3.26 (0.70)	4.06 (0.79)	3.55 (0.48)	4.47 (0.59)
FN BMD (g/ cm ²)	0.860 (0.102)	0.993 (0.113)	0.868 (0.082)	0.980 (0.103)	0.828 (0.120)	0.962 (0.129)	0.889 (0.092)	1.044 (0.091)

ANOVA detected no significant differences among groups for Unadjusted Baseline Data ($p>0.29$). Unadjusted dependent variables were not evaluated using RM ANOVA, because RM ANCOVA evaluated change across time in the main analyses, accounting for covariates.

Table 3.

Percentage Gains from Baseline to 19 month Follow-up

Dependent Variable	HI + LO + CON (n= 62)		HI (n= 19)		CON (n= 21)	
	Raw Mean	95% CI	Adjusted Mean	95% CI	Adjusted Mean	95% CI
SUBBMC (g)	29.0%	27.5, 30.5	29.4%	26.8, 31.9	27.2%	24.9, 29.5
UDradBMD (g/cm ²)	11.6%	7.7, 15.5	6.9%	0.6, 13.2	12.2%	6.5, 18.0
1/3radBMD (g/cm ²)	12.1%	10.6, 13.6	12.3%	9.4, 15.3	11.6%	8.9, 14.4
L1-L4BMC (g)	40.3%	37.9, 42.7	44.1% **	40.2, 48.1	35.9%	32.2, 39.5
L1-L4BMD (g/cm ²)	20.1%	18.5, 21.6	23.0% **	20.4, 25.7	17.3%	14.9, 19.8
FNBMCM (g)	23.5%	21.6, 25.4	27.4% **	23.9, 30.8	20.3%	17.2, 23.4
FNBMCM (g/cm ²)	15.2%	13.6, 16.8	18.7% **	15.7, 21.6	12.6%	9.8, 15.3

Mean percentage gains are presented for the full sample (leftmost columns: HI + LO + CON), with adjusted mean percentage gains for HI versus CON ANCOVA results (right columns). Covariates appearing in the HI versus CON model were evaluated at the following mean values: Inter-scan interval= 1.56 years, Mean Non-intervention Organized Activity over the interval= 3.8 h/wk, Maturity level = 0.8 (analogous to average maturity status of pre-menarche at baseline, late pre-menarche at follow-up, based on 0, 1, 2 for PRE, PERI, POST group assignment in ANCOVA). Baseline Height= 152.0 cm, Height Change= 8.4 cm, Mean Ca= 1294 mg, ln Mean vitamin D= 5.76 (IU).

** Significance was (p<0.01) for lumbar spine and femoral neck outcomes.

Results for Levene's tests were not significant (p = 0.07)