Combined hormonal contraceptive use is not protective against musculoskeletal conditions or injuries: a systematic review with data from 5 million females

Lynita White,[1] Justin M Losciale ,[2,3] Kipling Squier,[3,4] Sarah Guy,[5] Alex Scott,[3,4] Jerilynn C Prior,[6,7] Jackie L Whittaker [2,3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bjsports-2022-106519).

[1]Tall Tree Physiotherapy and Health Centre, Vancouver, British Columbia, Canada
[2]Arthritis Research Canada, Vancouver, British Columbia, Canada
[3]Department of Physical Therapy, Faculty of Medicine, The University of British Columbia, Vancouver, British Columbia, Canada
[4]Centre for Hip Health and Mobility, The University of British Columbia, Vancouver, British Columbia, Canada
[5]City Sport + Physiotherapy Clinic, Vancouver, British Columbia, Canada
[6]Centre for Menstrual Cycle and Ovulation Research, The University of British Columbia, Vancouver, British Columbia, Canada
[7]Women's Health Research Institute, The University of British Columbia, Vancouver, British Columbia, Canada

Correspondence to
Dr Jackie L Whittaker, Department of Physical Therapy, Faculty of Medicine, The University of British Columbia, Vancouver, Canada; jackie.whittaker@ubc.ca

Accepted 6 May 2023
Published Online First 24 May 2023

To cite: White L, Losciale JM, Squier K, *et al*. *Br J Sports Med* 2023;**57**:1195–1202.

ABSTRACT

Objective Assess the association between combined hormonal contraceptives (CHC) use and musculoskeletal tissue pathophysiology, injuries or conditions.

Design Systematic review with semiquantitative analyses and certainty of evidence assessment, guided by the Grading of Recommendations Assessment, Development and Evaluation approach.

Data Sources MEDLINE, EMBASE, CENTRAL, SPORTDiscus, CINAHL searched from inception to April 2022.

Eligibility Intervention and cohort studies that assessed the association between new or ongoing use of CHC and an outcome of musculoskeletal tissue pathophysiology, injury or condition in postpubertal premenopausal females.

Results Across 50 included studies, we assessed the effect of CHC use on 30 unique musculoskeletal outcomes (75% bone related). Serious risk of bias was judged present in 82% of studies, with 52% adequately adjusting for confounding. Meta-analyses were not possible due to poor outcome reporting, and heterogeneity in estimate statistics and comparison conditions. Based on semiquantitative synthesis, there is low certainty evidence that CHC use was associated with elevated future fracture risk (risk ratio 1.02–1.20) and total knee arthroplasty (risk ratio 1.00–1.36). There is very low certainty evidence of unclear relationships between CHC use and a wide range of bone turnover and bone health outcomes. Evidence about the effect of CHC use on musculoskeletal tissues beyond bone, and the influence of CHC use in adolescence versus adulthood, is limited.

Conclusion Given a paucity of high certainty evidence that CHC use is protective against musculoskeletal pathophysiology, injury or conditions, it is premature and inappropriate to advocate, or prescribe CHC for these purposes.

PROSPERO registration number This review was registered on PROSPERO CRD42021224582 on 8 January 2021.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Fifty per cent of young females start combined hormonal contraceptive (CHC) use for non-contraceptive reasons.
⇒ Adolescent CHC use may interfer with gain to peak bone mass and may delay the development of ovulatory menstrual cycles.
⇒ Some proponents of CHC suggest they may be protective against musculoskeletal injuries including anterior cruciate ligament tears.

WHAT THIS STUDY ADDS

⇒ There is low certainty evidence that CHC use is associated with higher future fracture risk by up to 1.20 times, and total knee arthroplasty risk by up to 1.36 times.
⇒ There is a paucity of high certainty evidence about the effects of CHC use on non-bone-related musculoskeletal injuries or conditions (ie, cartilage, ligament, muscle, tendon) or the influence of CHC use in adolescence versus adulthood.
⇒ Currently, there is insufficient high-quality evidence to make recommendations about the protective or negative effects of CHC use on musculoskeletal health.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Health care providers should use this information and consider patient's individual needs/priorities when prescribing CHC for non-contraceptive purposes.
⇒ It is premature to prescribe or pressure adolescent females to use CHCs to prevent musculoskeletal injuries or conditions.
⇒ Future research should test the cause-effect relationship between new CHC use and musculoskeletal injuries or conditions while accounting for confounding and other sources of bias that can distort findings (ie, self-selection).

BACKGROUND

In 2019, the Global Burden of Disease study estimated that 1.7 billion people, or 22% of the world's population, were affected by musculoskeletal (MSK) conditions.[1] MSK conditions reduce physical function, impair quality of life and are a significant cause of physical disability,[2] compromising mental health and increasing the risk of other chronic health conditions and opioid use.[3]

The lifelong burden of MSK conditions is greater in women than men. Women are more likely to

report MSK pain than men,[4] and they experience higher rates of chronic MSK conditions such as osteoarthritis.[5] Girls and women also have a greater risk of sport and recreation-related MSK injury.[6] For example, women are up to 2–5 times more likely to experience a knee ligament injury than men playing the same sport.[7] Once they have an MSK condition, women also account for greater expenditures related to pharmaceuticals, treatment of comorbidities and caregiving compared with men.[8] This disparity is multifactorial and associated with both biological (eg, anatomical predispositions, movement characteristics, tissue morphology) and sociocultural (eg, family roles, access to healthcare, type, and level of physical activity participation, exercise preference, and life events) factors.[9 10]

One biological factor that is alleged to contribute to the increased incidence of MSK conditions in females is their menstrual status.[11–17] This has led some to suggest that combined hormonal contraceptives (CHC) could be used to 'control' or 'stabilise' the menstrual cycle as a means to reduce the burden of MSK conditions. Unfortunately, much of the research informing non-contraceptive uses of CHC is based on hypothesis generating (ie, case series, cross-sectional, case–control or syntheses of these) not hypothesis testing studies (ie, cohort, intervention or syntheses of these).[12 14 17] In contrast, CHC use negatively impacts the ovulatory cycle,[18] peak bone mass[19] and is associated with reductions in bone mineral density (BMD), a precursor to osteoporosis and associated comorbidities.[19 20] The negative association between CHC use and bone health or other MSK tissue pathophysiology is rarely mentioned when encouraging CHC use to reduce injury risk.[12 14 21] Further, the long-term consequence of prolonged CHC use on MSK tissues and health is unknown.[22]

Given that ~50% of young females start CHC use for non-contraceptive reasons (ie, to control menstrual cycle irregularity (17%–18%), menstrual cramping (14%–26%), acne (10%–12%) and other reason including injury prevention (10%),[20] it is essential these decisions are evidence guided. The objective of this systematic review was to assess the association between CHC use and MSK tissue (ie, tendon, ligament, muscle, cartilage and bone) pathophysiology, injury and conditions based on a critical appraisal of existing studies. Given that adolescence is a key period in the life cycle for MSK tissue accrual,[23] and the effects of CHC may be unique to age groups,[19] a secondary objective was to consider adolescent (≤18 years of age) or adult (>18 years) CHC use.

METHODS

Framework

The Cochrane Handbook[24] guided the conduct, and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[25] and PRISMA-Search extension guided reporting of this review.[26]

Data sources and search

Relevant studies were identified by searching five databases (Medline-Ovid, EMBASE-Ovid, CENTRAL-Ovid, SPORTDiscus-EBSCOhost, CINAHL-EBSCOhost). Search strategies consisted of medical subject headings and text words related to CHC, the concept of MSK tissues, injuries or conditions with limitations for human participants, and the English language. The Medline search strategy was developed in consultation with a health sciences librarian and adapted for other databases. The search strategy for Medline-Ovid is available in online supplemental file 1. Searches were run on 28 December 2020, and updated on 27 April 2022. Searches were documented and reference lists of identified systematic reviews and included studies were handsearched to identify additional relevant records. Records were transferred to a reference management software (RefWorks, ProQuest, USA).

Eligibility

We included studies that reported primary data from humans that assessed the association between CHC use by postpubertal, premenopausal females with normal menstrual status (ie, not amenorrhoea or oligomenorrhoea), and MSK tissue (ie, bone, joint, ligament, muscle, tendon and associated connective tissues) pathophysiology, injuries or conditions. CHC use was operationalised as new or ongoing use of a pharmacologic dose of synthetic ethinyl estradiol (EE) in combination with progestin in pill, patch or vaginal ring forms. Studies including participants with mixed menstrual status were included if data were sufficiently disaggregated (eg, normal menses vs oligomenorrhoea). Pathophysiology was defined as disordered physiological processes, and a condition as a disease or lesion that negatively affects the structure or function of a tissue. We included all defined outcomes of MSK injury regardless of onset type (ie, acute or chronic) or time loss. Analyses reporting data from the same parent study were included if they assessed different CHC exposures (ie, preparation or dose), MSK outcomes or presented data at different time points. We excluded studies with participants who had a condition (eg, anorexia nervosa) or were undergoing treatment that could affect reproductive hormone levels (eg, Turner syndrome) that assessed an outcome of MSK performance only (eg, muscle strength, exercise-induced muscle damage, VO_2 max, electromyography activity and functional performance), and contraceptive preparations that included oestrogen only, progestin-only or were implanted or intrauterine devices (IUD), as these devices are either non-hormonal (IUD) or progestin-only (implant, IUD).

Study selection

After manual identification and removal of duplicates in Refworks (LW), records were imported into a screening and data extraction platform (Covidence, Veritas Health Innovation). Authors (LW, JML, KS, SG, AS and JLW) independently screened titles and corresponding abstracts in duplicate to determine potentially relevant records, followed by full-text review to determine final record selection. A third author resolved any disagreements at all stages if the two primary reviewers could not reach a consensus. All decisions and reasons for inclusion and exclusion were recorded in Covidence.

Data extraction

Authors (LW, JML, KS, SG, AS and JLW) independently performed data extraction, in duplicate using a structured data extraction form (Covidence). Data extraction included: study information (first author, publication date, title, location, design, population description, sample size, participants per group, funding sources and conflicts of interest); participant characteristics (sample description, and age at enrolment and subsequent follow-up); CHC details (dosage and chemical compound, method of delivery and length of prestudy and within-study use); follow-up duration (start and end date); MSK outcomes (ie, outcome and measurement method) and results (unadjusted and adjusted group level values and between-group comparisons as available).

Study quality and risk of bias

Authors (LW, JML, KS, SG, AS, JLW) independently assessed study quality and risk of bias across included studies in duplicate

using the Downs and Black Quality Assessment Tool (DBQAT).[27] This tool assigns an individual score calculated out of 32 total points for each study (11 points for reporting, 3 points for external validity, 7 points for bias, 6 points for confounding and 5 for power: see online supplemental file.[27] Disagreements were resolved through consensus or third author when needed. The potential for selection, attrition and measurement bias, and bias due to confounding and statistical analysis were rated as 'not serious' (≥12/13 points), 'serious' (11/13 points) or 'very serious' (≤10/13 points) using questions 14–26 to facilitate semiquantitative synthesis (see Semiquantitative synthesis below).[27]

Data syntheses

Data synthesis involved three steps. First, we identified and categorised unique MSK tissue (ie, bone, joint, ligament, muscle, tendon and associated connective tissues) pathophysiology, injuries or conditions assessed across the included studies. Second, within each MSK outcome category, studies were grouped based on similar statistics of effect (eg, ORs, HRs, mean difference), follow-up times and age. If no statistic of effect was reported, it was calculated (Cohen's d for continuous outcomes, ORs for dichotomous outcomes) when the necessary raw data were available. Finally, quantitative (meta-analysis) or semiquantitative synthesis was conducted. Meta-analyses were planned a priori for MSK outcomes where there were two or more studies with similar statistics of effect, with the remaining outcomes (those with insufficient data for meta-analyses) to undergo semiquantitative analyses.[28]

Quantitative synthesis

If possible, meta-analyses (random effects models with inverse variance weighting using restricted maximum likelihood estimation) were performed to estimate an overall mean difference (same unit continuous outcomes), standardised mean difference (different scale continuous outcomes), or OR (dichotomous outcomes), and a rating of overall certainty of the evidence was assigned as 'high' or downgraded to 'moderate', 'low' or 'very low' using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach.[29]

Semiquantitative synthesis

When meta-analyses were not possible, semiquantitative syntheses[28] were performed. This involved rating the certainty and confidence of evidence for CHC effect on each MSK outcome using a modified GRADE approach with adaptations to assess non-pooled data across five domains including study design (randomised controlled trials (RCTs) were assigned higher certainty than observational studies); risk of bias (DBQAT questions 14–26 rating); inconsistency (inconsistency in sample, methods and heterogeneity statistic results); indirectness (generalisability of findings to the target population and research question) and imprecision (95% CI width).[24] All domain ratings were considered when assigning an overall judgement of high, moderate, low or very low certainty of evidence[28] and a corresponding statement of the direction (considering consistency reported across studies) and magnitude of the treatment effect of CHC on each respective outcome. A similar approach has been used in a previous review paper by our group.[30]

Protocol deviations

During study selection and data synthesis, we made post hoc changes to our study protocol. Specifically, we narrowed our selection criteria to hypothesis testing study designs, (ie, cohort and intervention studies including RCTs and quasi-experimental), which are less prone to bias (eg, survivor bias or reverse causality) and provide more robust evidence to inform clinical recommendations (ie, prescription of CHC to prevent MSK conditions).[31–33] We also increased the minimum number of studies with a similar effect statistic needed for meta-analyses and semiquantitative analyses to outcomes with three or more studies. This decision was based on best practice guidance,[24] to reduce the probability of ambiguous and unclear conclusions.

Equity and diversity statement

Due to the nature of our research question, we only included studies with female participants but did not restrict on gender, geographical region, or socioeconomic or education level. The study team included diverse perspectives including those of women and men, clinicians (physiotherapists) and clinician scientists with a diversity of career stages (PhD candidates through to professor), persons of colour and members of the two-spirit, lesbian, gay, bisexual, transgender, queer and/or questioning, intersex, asexual and additional sexual orientations and gender identities (2SLGBTQIA+) community. We acknowledge that we lack perspectives of persons from middle-income to low-income geographical regions.

Patient and public Involvement

There was no patient or public involvement in this study.

RESULTS

Of 5438 potential records identified from database searches, 50 were included (figure 1)[34–83] incorporating data from 5 695 908 participants from 48 unique cohorts.[34 35 38–66 69–83]

Study characteristics

Included studies are summarised in online supplemental file 1. Three were RCTs (6%),[35 43 74] 14 were quasi-experimental (28%),[46 47 49 50 52 56 62 64–66 76 77 79 83] 25 were prospective cohort studies (50%)[38–42 44 45 48 51 53–55 57 60 61 63 69–72 75 78 80–82] and 8 were retrospective cohort studies (16%).[34 36 37 58 59 67 68 73]

Six studies (12%)[39 40 64 66 71 76] assessed participants who took CHC under the age of 18 (adolescent) exclusively and 44 studies (88%)[34–38 41–63 65 67–70 72–75 77–83] assessed participants across various age ranges without stratification by age groups. The length of CHC interventions across studies ranged from 28 days51[61] to 14.5 years.[59] Prior CHC exposure across participants in the CHC intervention groups ranged from no prior use[43 46 47 49 52 57 59 64–66 71 74 77–80 83] to >97 months.[36 37] Comparison conditions included a variety of 'non-CHC users', including participants with no lifetime CHC use[80] to no CHC use 2–6 months before enrolment.[82 83] Follow-up ranged from 25 days[51] to 26 years,[36 37] with 12 studies (24%)[40 46 47 49 50 62 64–66 71 77 82] concluding at 1 year.

MSK outcomes

Thirty unique MSK outcomes spanning five tissue types (ie, bone, tendon, joint, ligament and muscle) were identified (table 1). Seventeen outcomes (57%) were assessed in three or more studies. Of these, 14 outcomes (82%) were bone related, with lumbar spine BMD (n=23 studies, 46%)[35 39 42 46–50 52 58 62–66 71 73–75 77–79] and femoral neck BMD (n=14 studies, 28%)[35 40 42 48 50 52 58 62–64 66 71 75 78] being the most common.

Study quality and risk of bias

The results of the study quality and risk of bias assessment are summarised in online supplemental file 1. The median DBQAT score was 15 (8-24). Only 9 studies[34 39 42 53 71–73 78 80] (18%) were judged to be at 'not serious' risk of bias, while 26 (52%)[39 40 42 43 45 53 55 57–60 63 64 67 68 71–76 78–82] were judged to

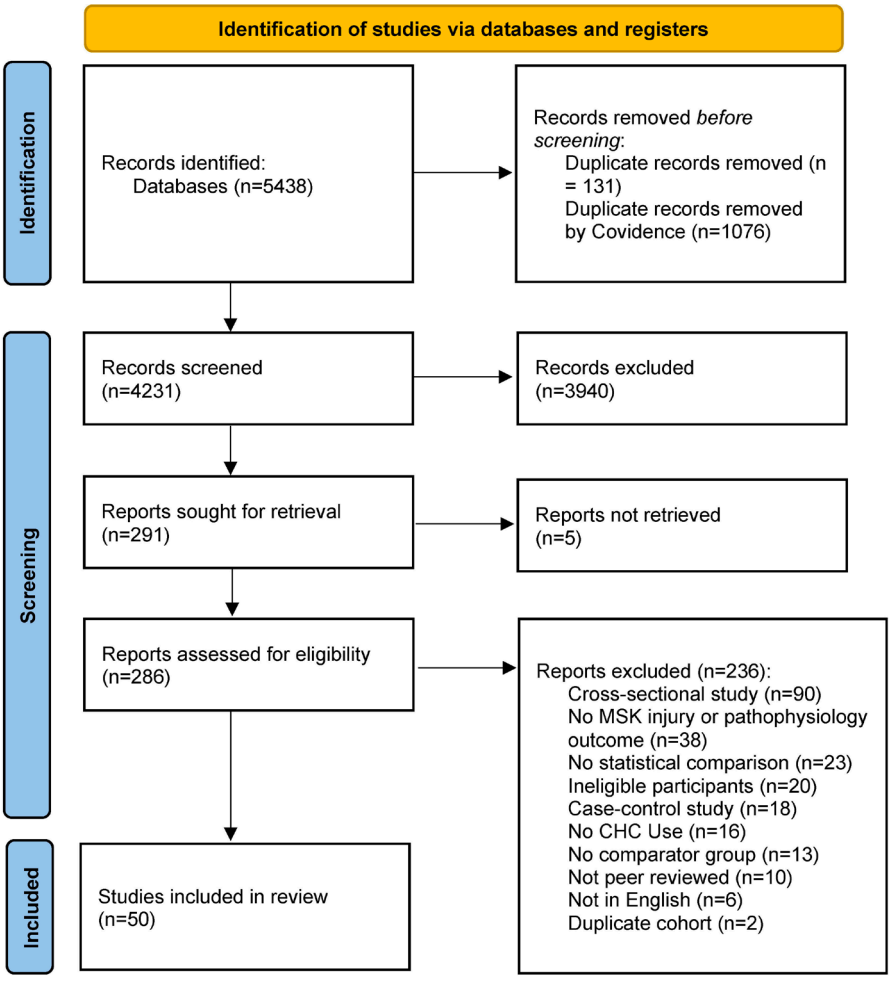

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses[25] flow chart. CHC, combined hormonal contraceptive MSK, musculoskeletal.

adequately adjust for confounding. Concerns for data dredging were judged in 20 studies (40%),[35–37 43 45 48 55–60 62–66 69 74 81] and for selection biases in 7 studies (14%).[35 41 50 51 60 62 63]

Effect of CHC use on MSK outcomes

All estimates of CHC effect are summarised in online supplemental file 1. Twenty-one studies (42%)[35 38 40 42 44–46 49 51 52 56 61 63 65 66 69–71 73 82 83] provided a comparison of group means, 7 (14%)[39 43 50 63 67 74 75] provided a comparison of pre-post change scores, 13 (26%)[39 42 48 62–64 66 69 70 77–80] comparison of percent pre–post change, 1 (2%)[68] an OR, 2 (4%)[57 72] rate ratios, 3 (6%)[36 37 53] risk ratios, 6 (12%)[34 55 59 60 74 81] HRs and 2 (4%)[41 47] only reported a p value.

Meta-analyses

Despite 17 outcomes being assessed in three or more studies (table 1), meta-analyses were not possible due to poor outcome reporting (ie, missing group means and/or SD), differences in estimate statistics and significant heterogeneity in comparison condition.

Semiquantitative analyses

The results of the semiquantitative analyses of outcomes assessed in three or more studies are summarised in table 2, with additional detail in online supplemental file 1. Due to limited stratification of data in the source studies, it was not possible to consider adolescent versus adult use. Modified GRADE ratings were consistently downgraded for risk of bias (81.3% with very serious limitations), indirectness (56.3%) and imprecision (87.5%). The evidence for two outcomes (total knee arthroplasty and any fracture) was rated low certainty, while the evidence for all other outcomes was rated very low certainty. There is low certainty evidence that CHC use may be associated with increased future fracture risk by up to 1.20 times (risk ratios range from 1.02 to 1.20),[34 37 72 81] and total knee arthroplasty by up to 1.36 times (risk ratios range from 1.00 to 1.36; see online supplemental file).[53 55 60]

The effect of adolescent CHC use

Six studies exclusively assessed females ≤18 years old across nine outcomes.[40 64 66 71 76 80] Four studies assessed the effect of CHC use on lumbar spine and femoral neck BMD,[40 64 66 71] each, while one study assessed the other seven outcomes each (ie, whole body BMD, radius BMD, deoxypyridinoline, bone alkaline phosphatase, lumbar spine bone mineral content (BMC), whole body BMC, lean body mass). Semiquantitative synthesis indicate there is very low certainty evidence of an unclear association between CHC use and lumbar spine and femoral neck BMD.

DISCUSSION

Semiquantitative analyses reveal low certainty evidence that CHC use may be associated with higher future fracture risk (up to 1.2 times), and total knee arthroplasty risk (up to 1.36 times). Beyond

Table 1 Musculoskeletal outcome by number of studies

Outcome tissue (type)	Outcome	No of studies
Bone (density)	Lumbar BMD	23[35 39 40 42 46–50 52 58 62–66 71 73–75 77–79]
Bone (density)	Femur BMD	14[35 40 42 48 50 52 58 62–64 66 71 75 78]
Bone (density)	Whole Body BMD	8[39 42 45 58 64 73 77 82]
Bone (turnover; resorption)	D-PYD	6[40 46 47 49 65 67]
Bone (content)	Lumbar BMC	5[35 58 62 64 77]
Bone (content)	Whole Body BMC	5[35 58 64 74 77]
Bone (turnover; resorption)	CTX	5[41 45 51 61 82]
Muscle/bone/ligament/tendon (mass)	LBM	5[43 70 73 76 82]
Bone (density)	Total Hip BMD	4[39 73–75]
Bone (density)	Radius BMD	4[35 48 62 80]
Bone (turnover; formation)	BAP	4[40 41 45 51]
Bone (turnover; formation)	P1NP	4[45 51 61 83]
Bone (turnover; formation)	BGP	4[46 47 49 65]
Bone (turnover; resorption)	PYD	4[46 47 49 65]
Bone (injury)	Any fracture	4[34 37 72 81]
Muscle (structure)	Quadriceps morphology*	4[38 44 69 70]
Joint (condition)	TKA	3[53 55 60]
Bone (density)	Trochanter BMD	2[35 63]
Bone (density)	Calcaneus BMD	2[67 68]
Bone (density)	Ward's Triangle BMD	2[35 63]
Bone (density)	Tibia BMD	2[45 62]
Bone (content)	Femur BMC	2[58 64]
Bone (content)	Radius BMC	2[35 63]
Bone (injury)	Stress fracture	2[57 74]
Joint (condition)	THA	2[53 60]
Ligament (injury)	ACL Injury	2[54 59]
Bone (density)	Subtotal BMD	1[77]
Bone (content)	Subtotal BMC	1[77]
Bone (turnover; resorption)	IGF-1	1[83]
Tendon (structure)	Tendon morphology†	1[70]
Joint (injury)	Back disorder‡	1[36]
Joint (motion)	Anterior tibial translation	1[56]

*Muscle fibre cross-sectional area, whole muscle cross-sectional area, fibre type, myonuclei content, muscle thickness.
†Tendon cross-sectional area, collagen concentration, collagen cross-linking.
‡Tendon cross-sectional area, collagen concentration, collagen cross-linking.
ACL, anterior cruciate ligament; BAP, bone alkaline phosphatase; BGP, osteocalcin; BMC, bone mineral content; BMD, bone mineral density; CTX, C-terminal peptide; D-PYD, deoxypyridinoline; IGF-1, insulin-like growth factor 1; LBM, lean body mass; P1NP, procollagen type 1 terminal peptide; THA, total hip arthroplasty; TKA, total knee arthroplasty.

this, there is very low certainty evidence of unclear relationships between CHC use and a wide range of bone turnover and bone health outcomes, and a paucity of evidence about the effect of CHC use on other MSK tissue (ie, tendon, ligament, muscle, cartilage) physiology, injury or conditions. Despite the importance of the adolescence period for MSK tissue accrual,[23] and evidence that the effect of CHC use may be unique to life stage,[19] stratification by adolescent or adult CHC use was not possible.

Our results build on the findings of past reviews that report on the short-term associations between CHC use and specific MSK injuries[12] or BMD.[19] By assessing the effect of CHC use on a broader spectrum of MSK tissue pathophysiology, injuries and associated conditions without restriction by follow-up length, our review has been able to consider MSK conditions which are more prevalent in women that typically appear later in life (eg, osteoarthritis, frailty fracture).[5 37] This more inclusive approach, combined with an assessment of the certainty of the evidence, provides a comprehensive overview of the MSK considerations of CHC use across the lifespan that can be used by females and their healthcare providers to inform decisions about CHC prescription.

Across 50 studies and 32 unique outcomes included in this review, the majority were related to bone structure or bone physiology. Despite this large evidence base, there is still a lot that is unclear. Although there is evidence that past CHC use is associated with higher future fracture risk at any site, this was judged to be low certainty evidence suggesting that the estimate of effect is likely to change with future research. Similarly, there is only very low certainty evidence of unclear or absent relationships between CHC use and most other bone-related outcomes. Outside of bone-related outcomes, we identified an elevated risk for total knee arthroplasty (a common end-stage treatment for knee osteoarthritis) in past CHC users.

An important finding from this review is the paucity of evidence assessing CHC use and tendon, ligament, or muscle-related outcomes (eg, tendinopathy, ligament ruptures or sprains, muscle strains), which is foundational information needed before encouraging CHC use for injury prevention. This is in direct contrast to previous reviews that suggest CHC use may decrease the risk of ACL laxity and ACL tears,[12 14 21] which, in turn, could reduce the prevalence of one of the most burdensome MSK conditions—knee osteoarthritis.[30 84] There are important methodological differences between the current and past reviews that may explain the discrepancy in findings. Specifically, conclusions of past reviews are based on selective interpretation of case series and case–control studies,[12 14] and either did not rate evidence certainty (ie, GRADE) or did not follow best practice to summarises the totality of the evidence for each outcome, and indicate how likely the findings are to change with future research.[24 28]

Although these case-series and case–control study designs are important for generating hypotheses, they are without strict controls that make them prone to bias (error that consistently increases or decreases the effect of an intervention) and confounding (distortion of the effect of an intervention by a third factor) which require cautious interpretation of their results.[32] To highlight this, Herzog *et al*[59] conducted a population cohort study with nearly 3 million females and a nested case–control analyses (see figure 2). The cohort study demonstrated no difference in ACL tear risk (adjusted HR 0.95; 95% CI 0.89 to 1.01) between new CHC users and non-CHC (ie, IUD) users, while the case–control analyses identified that participants who used CHC at any time in the past 5 years had lower odds of an ACL tear (adjusted OR 0.90; 95% CI 0.85 to 0.94).[59] The discrepancy demonstrates that case–control analyses may be influenced by bias, including selection bias (selected based on outcome vs exposure) or comparator bias (active contraceptive seeking control vs non-contraceptive seeking control) and/or confounding. This skillful illustration of the limitations of hypothesis generating study designs to understand associations supports our decision to limit our inclusion to cohorts and intervention study designs.

A secondary aim of this review was to consider the differential effects of CHC use on MSK outcomes stratified by adolescent (≤18 years of age) or adult (>18 years) use. This aim was based on the fact that adolescence is a critical period for the developing MSK system,[19 85] and a previous review that demonstrated adolescent CHC use may have a detrimental effect on BMD accrual.[19] While speculative, our finding of an elevated risk of

Table 2 Semiquantitative synthesis with modified GRADE rating

Outcome	Studies (n)	Participants (n)	Follow-up (years)	Study design	Study limitations	Inconsistency	Indirectness	Imprecision	Modified GRADE rating	Direction and magnitude (without stratifying for age)
Lumbar BMD[35 39 40 42 46–50 52 58 62–66 71 73–75 77–79]	23	4484	1–20	+++	XX	✓	X	X	+	CHC use may not be associated with lumbar BMD
Femoral Neck BMD[35 40 42 48 50 52 58 62–64 66 71 75 78]	14	3314	1–20	++++	XX	✓	X	X	+	CHC use may not be associated with femoral neck BMD
Whole Body BMD[39 42 45 58 64 73 77 82]	8	1914	0.8–20	+++	XX	X	X	X	+	Association between CHC use and whole body BMD is unclear
D-PYD[40 46 47 49 65 67]	6	668	1–2	+++	XX	X	✓	X	+	Association between CHC use and DPD is unclear
CTX[41 45 51 61 82]	5	180	0.08–1	+++	XX	✓	X	X	+	Association between CHC use and CTX is unclear
LBM[43 70 73 76 82]	5	713	0.18–2.2	++++	XX	X	X	X	+	Association between CHC use and LBM is unclear
Lumbar BMC[35 58 62 64 77]	5	1230	1–20	+++	XX	X	X	X	+	Association between CHC use and lumbar BMC is unclear
Whole body BMC[35 58 64 74 77]	5	1321	1–20	++++	XX	X	✓	X	+	Association between CHC use and whole body BMC is unclear
Any fracture[34 37 72 81]	4	1 663 062	2.5–26	+++	X	✓	✓	✓	++	CHC use may ↑ fracture risk by 1.20 times (RR 95% CI 1.08 to 1.34)
P1NP[45 51 61 83]	4	119	0.08–0.8	+++	XX	✓	X	X	+	Association between CHC use and P1NP is unclear
PYD[46 47 49 65]	4	233	1	++	XX	✓	✓	X	+	Association between CHC use and PYD is unclear
BAP[40 41 45 51]	4	397	0.08–1	+++	XX	X	X	X	+	Association between CHC use and BAP is unclear
BGP[46 47 49 65]	4	233	1	++	XX	X	✓	X	+	Association between CHC use and BPG is unclear
Radius BMD[35 48 62 80]	4	655	1–5	++++	XX	✓	✓	X	+	Association between CHC use and radius BMD is unclear
Total Hip BMD[39 73–75]	4	933	2–7	++++	X	X	X	X	+	CHC use may not be associated with Total Hip BMD
TKA[53 55 60]	3	1 345 078	6–15	+++	X	✓	✓	✓	++	CHC use may ↑ TKR risk by 1.36 times (RR 95% CI 1.00 to 1.86)

✓, no limitation; +++, moderate-quality evidence; +, very low quality evidence; ++, low-quality evidence; ++++, high-quality evidence; BAP, bone alkaline phosphatase; BGP, Osteocalcin; BMC, bone mineral content; BMD, bone mineral density; CHC, combined hormonal contraceptive; CTX, C-terminal peptide; D-PYD, deoxypyridinoline; GRADE, Grading of Recommendations Assessment, Development and Evaluation; LBM, lean body mass; P1NP, procollagen type 1 terminal peptide; RR, risk ratio; TKA, total knee arthroplasty; X, serious limitation; XX, very serious limitation.

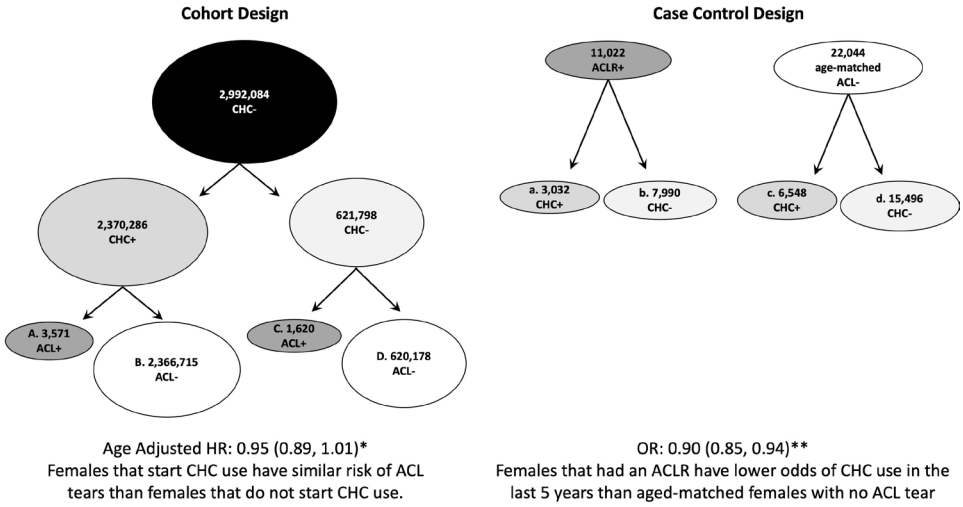

Figure 2 Comparison of cohort versus case–control design by Herzog *et al*[59] ACL (anterior cruciate ligament), ACL+ (ACL tear), ACL– (no ACL tear), ACLR (ACL reconstruction), ACLR+ (had an ACL reconstruction), CHC (combined hormonal contraceptive), CHC+ (CHC use), CHC– (no CHC use), HR, OR. *HR: The risk of an ACL tear among females who initiated CHC use relative to the risk of anterior cruciate ligament tear among women who did not initiate combined hormonal contraceptives adjusted for age. HR=(A/(A+B))/(C/(C+D)). **OR: The odds of using CHC over the previous 5 years among females who had an ACLR relative to the odds of taking CHC over the same period among age-matched females who did not have an ACL tear. OR=(a/b)/(c/d).

future fracture in CHC users could be related to poor BMD accrual in adolescence. Unfortunately, 88% of studies included in this review had participants that spanned adolescence through adulthood and did not provide stratified results which interfered with our ability to fully explore the difference in the effect of CHC between these two life stages. Across the two outcomes (lumbar spine and femoral neck BMD) where semiquantitative analyses were possible, there was very low quality evidence of an unclear association with CHC use.

Clinical implications

Currently, there is insufficient evidence to recommend CHC use to protect MSK health including the prevention of ACL tears. In contrast, CHC use could increase the risk of future fractures and total knee arthroplasty. Females and their healthcare providers can use this information to inform decisions about CHC prescriptions.

Recommendations for future research

Given the paucity of high certainty evidence about the effect of CHC use on MSK outcomes, in particular non-bone-related outcomes, there are many opportunities to contribute to the field through rigorously designed prospective cohort studies or RCTs. Future studies should follow established reporting guidelines for cohort and RCTs (Strengthening the Reporting of Observational Studies in Epidemiology), Consolidated Standards of Reporting Trials (CONSORT)).[86 87] This includes reporting results in their native units, avoiding selective reporting of p values, and reporting data only in figures. Investigators are also encouraged to clearly define the non-user comparator group and any washout period used to facilitate study pooling. For bone-related outcomes, we encourage investigators to disaggregate their results by life stage (eg, adolescent vs adult) and consider the influence of menstrual status (eg, normal menses, amenorrhoea or oligomenorrhoea) to facilitate better understanding of the relationship between CHC use and BMD.[19]

Strengths and limitations

We followed best practices for systematic reviews including a priori protocol registration, a comprehensive search strategy developed in collaboration with a librarian scientist, and grading risk of bias plus the certainty of evidence. These efforts facilitated an extensive synthesis and analysis across many relevant outcomes. Post hoc protocol changes allowed us to focus on rigorous study designs that can directly inform treatment decisions (ie, RCTs and cohort studies) while avoiding the inherent biases of cross-sectional studies (eg, reverse-causality bias) or case–control studies (eg, incidence-prevalence bias). The decision against pooling data across different estimate statistics (ie, per cent change and mean change) and life stages reduced the ambiguity of our conclusions. We chose not to contact individual authors for missing data which may have prevented us from performing further meta-analyses. This decision was made based on the broad scope of the review, many source studies being published more than 5 years previously, and past experiences where efforts infrequently result in helpful clarification.[88]

Despite our extensive search strategy, it is important to acknowledge the possibility of omitting a relevant study. Although approximately two-thirds of the studies included in this review were conducted in North America or Europe, the remaining one-third represents data from Africa, Asia and South America suggesting that our findings may be considered applicable beyond white communities in high-income countries. A few of the included studies reported data about the education level or other determinants of health, it is not possible to comment on the generalisability of our findings in this respect. The semiquantitative GRADE approach relies on the judgement of the research team and the ratings may reflect implicit biases. This likely resulted in more downgrading of evidence-certainty ratings to avoid overstating findings without supporting quantitative estimates. We chose to only synthesise outcomes with three or more studies of similar outcomes based on our previous experience using this approach and the high likelihood that the ratings based on two studies would be downgraded due to uncertainty. Despite this, we have included the full findings of all studies in online supplemental file 1. The decision to synthesise studies with different comparison conditions (eg, never vs new user) can increase heterogeneity and lead to evidence-certainty downgrading. The decision to restrict study inclusion to

designs (ie, only intervention or cohort studies) that are hypothesis testing, led to case–control studies being omitted from our review and is one potential reason why our findings may differ from past reviews. Finally, the decision to not restrict study follow-up length enhances the generalisability of these findings, but also creates variability that could lead to evidence downgrading.

CONCLUSION

There is insufficient evidence to support the use of CHC to prevent MSK injuries in females, including ACL tears. Low certainty evidence suggests that past CHC use may be associated with a slightly elevated risk of future fracture and total knee arthroplasty. Very low certainty evidence indicates that the association between CHC use and BMD, BMC, and other biomarkers of bone physiology is unclear or absent.

Given a paucity of high certainty evidence that CHC use is protective against MSK pathophysiology, injury or conditions, it is premature and inappropriate to advocate or prescribe CHC for these purposes.

Twitter Justin M Losciale @JayLos18 and Jackie L Whittaker @jwhittak_physio

Acknowledgements JLW is supported by a Michael Smith Foundation for Health Research a Scholar Award (SCH-2020-0403) and an Arthritis Society STAR Career Development Award (STAR-19-0493). JML is supported by the Arthritis Society (TGP 21-0000000086) and the University of British Columbia Four-Year Fellowship.

Contributors LW and JLW designed the systematic review. JLW coordinated the study. LW, JML, KS, SG, AS and JLW screened titles and abstracts as well as full-text records. LW, JML, KS, SG, AS and JLW performed data extraction and risk of bias assessment. LW, JML and JLW performed semiquantitative synthesis. LW, JML and JLW wrote the first draft of the manuscript. JCP provided perspectives on CHC and bone variables. All authors contributed to reviewing, editing and revising the manuscript and approved the final submitted version. JLW is the guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests JLW is an associate editor of the *British Journal of Sports Medicine*.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

ORCID iDs
Justin M Losciale http://orcid.org/0000-0001-5135-1191
Jackie L Whittaker http://orcid.org/0000-0002-6591-4976

REFERENCES

1 Cieza A, Causey K, Kamenov K, *et al*. Global estimates of the need for rehabilitation based on the global burden of disease study 2019: A systematic analysis for the global burden of disease study 2019. *Lancet* 2021;396:S0140-6736(20)32340-0:2006–17.:.
2 Sprangers MA, de Regt EB, Andries F, *et al*. Which chronic conditions are associated with better or poorer quality of life *J Clin Epidemiol* 2000;53:895–907.
3 Briggs AM, Cross MJ, Hoy DG, *et al*. Musculoskeletal health conditions represent a global threat to healthy aging: A report for the 2015 world health organization world report on ageing and health. *Gerontologist* 2016;56 Suppl 2:S243–55.
4 Rollman GB, Lautenbacher S. Sex differences in musculoskeletal pain. *Clin J Pain* 2001;17:20–4.
5 Srikanth VK, Fryer JL, Zhai G, *et al*. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis and Cartilage* 2005;13:769–81.
6 Emery CA. Risk factors for injury in child and adolescent sport: a systematic review of the literature. Clinical Journal of Sport Medicine 2003;13:256–68.
7 Montalvo AM, Schneider DK, Webster KE, *et al*. Anterior cruciate ligament injury risk in sport: a systematic review and meta-analysis of injury incidence by sex and sport classification. J Athl Train 2019;54:472–82.
8 Public Health Agency of Canada. *Economic Burden of Illness in Canada 2003-2008*. Ottawa: Minister of Health, 2014.
9 Laird Y, Fawkner S, Niven A. A grounded theory of how social support influences physical activity in adolescent girls. Int J Qual Stud Health Well-being 2018;13:1435099.
10 Telford RM, Telford RD, Olive LS, *et al*. Why are girls less physically active than boys? findings from the look longitudinal study. PLoS One 2016;11:e0150041.
11 Hewett TE, Zazulak BT, Myer GD. Effects of the menstrual cycle on anterior cruciate ligament injury risk: a systematic review. *Am J Sports Med* 2007;35:659–68.
12 Konopka JA, Hsue LJ, Dragoo JL. Effect of oral contraceptives on soft tissue injury risk, soft tissue laxity, and muscle strength: a systematic review of the literature. Orthop J Sports Med 2019;7:232596711983106.
13 Bokshan SL, Mehta S, DeFroda SF, *et al*. What are the primary cost drivers of anterior Cruciate ligament reconstruction in the United States? A cost-Minimization analysis of 14,713 patients. *Arthroscopy* 2019;35:S0749-8063(18)31190-3:1576–81.:.
14 Herzberg SD, Motu'apuaka ML, Lambert W, *et al*. The effect of menstrual cycle and contraceptives on ACL injuries and laxity: a systematic review and meta-analysis. Orthopaedic Journal of Sports Medicine 2017;5:232596711771878.
15 Drinkwater BL, Bruemner B, Chesnut CH. Menstrual history as a determinant of current bone density in young athletes. *JAMA* 1990;263:545–8.
16 Kubo K, Miyamoto M, Tanaka S, *et al*. Muscle and tendon properties during Menstrual cycle. *Int J Sports Med* 2009;30:139–43.
17 Teal S, Edelman A. Contraception selection, effectiveness, and adverse effects: A review. *JAMA* 2021;326:2507–18.
18 Prior J, Whittaker JL, Scott AW. Adolescent combined hormonal contraceptives and surgical repair of anterior Cruciate tears: A risky recommendation based on an Unproven causal relationship. *Phys Sportsmed* 2019;47:240–1.
19 Goshtasebi A, Subotic Brajic T, Scholes D, *et al*. Adolescent use of combined hormonal contraception and peak bone mineral density accrual: A meta-analysis of international prospective controlled studies. *Clin Endocrinol (Oxf)* 2019;90:517–24.
20 Prior JC. Adolescents' use of combined hormonal contraceptives for Menstrual cycle–related problem treatment and contraception: Evidence of potential lifelong negative reproductive and bone effects. *Women's Reproductive Health* 2016;3:73–92.
21 DeFroda SF, Bokshan SL, Worobey S, *et al*. Oral contraceptives provide protection against anterior cruciate ligament tears: a national database study of 165,748 female patients. *Phys* Sportsmed 2019;47:416–20.
22 Trémollieres F. Impact of oral contraceptive on bone metabolism. Best Pract Res Clin Endocrinol Metab 2013;27:47–53.
23 Weaver CM, Gordon CM, Janz KF, *et al*. The National Osteoporosis Foundation's position statement on peak bone mass development and lifestyle factors: a systematic review and implementation recommendations. *Osteoporos Int* 2016;27:1281–386.
24 Higgins JPT, Thomas J, Chandler J, *et al*. Cochrane Handbook for systematic reviews of interventions. In: *Cochrane Handbook for Systematic Reviews of Interventions*. 23 September 2019.
25 Page MJ, McKenzie JE, Bossuyt PM, *et al*. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
26 Rethlefsen ML, Kirtley S, Waffenschmidt S, *et al*. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev* 2021;10:39.
27 Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377–84.
28 Huguet A, Hayden JA, Stinson J, *et al*. Judging the quality of evidence in reviews of Prognostic factor research: Adapting the GRADE framework. *Syst Rev* 2013;2:71.
29 Atkins D, Best D, Briss PA, *et al*. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
30 Whittaker JL, Losciale JM, Juhl CB, *et al*. Risk factors for knee osteoarthritis after traumatic knee injury: a systematic review and meta-analysis of randomised controlled trials and cohort studies for the OPTIKNEE consensus. *Br J Sports Med* 2022;56:1406–21.
31 Celentano DD, Szklo M, Gordis L. *Gordis Epidemiology*6th ed. Philadelphia, PA: Elsevier, 2019: 149–75.
32 Delgado-Rodríguez M, Llorca J, Bias LJ. Bias. *J Epidemiol Community Health* 2004;58:635–41.
33 Hayden JA, Côté P, Steenstra IA, *et al*. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol* 2008;61:552–60.
34 Yoo JE, Shin DW, Han K, *et al*. Association of female reproductive factors with incidence of fracture among postmenopausal women in Korea. *JAMA Netw Open* 2021;4:e2030405.
35 Weaver CM, Teegarden D, Lyle RM, *et al*. Impact of exercise on bone health and contraindication of oral contraceptive use in young women. Med Sci Sports Exerc 2001;33:873–80.

36 Vessey M, Painter R, Mant J. Oral contraception and other factors in relation to back disorders in women: findings in a large cohort study. Contraception 1999;60:331–5.
37 Vessey M, Mant J, Painter R. Oral contraception and other factors in relation to hospital referral for fracture. findings in a large cohort study. Contraception 1998;57:231–5.
38 Sung E-S, Han A, Hinrichs T, *et al*. Effects of oral contraceptive use on muscle strength, muscle thickness, and fiber size and composition in young women undergoing 12 weeks of strength training: a cohort study. BMC Womens Health 2022;22.
39 Scholes D, Hubbard RA, Ichikawa LE, *et al*. Oral contraceptive use and bone density change in adolescent and young adult women: a prospective study of age, hormone dose, and discontinuation. J Clin Endocrinol Metab 2011;96:E1380–7.
40 Rome E, Ziegler J, Secic M, *et al*. Bone biochemical markers in adolescent girls using either depot medroxyprogesterone acetate or an oral contraceptive. J Pediatr Adolesc Gynecol 2004;17:373–7.
41 Reiger J, Yingling VR. The effects of short-term jump training on bone metabolism in females using oral contraceptives. J Sports Sci 2016;34:259–66.
42 Reed SD, Scholes D, LaCroix AZ, *et al*. Longitudinal changes in bone density in relation to oral contraceptive use. Contraception 2003;68:177–82.
43 Procter-Gray E, Cobb KL, Crawford SL, *et al*. Effect of oral contraceptives on weight and body composition in young female runners. Med Sci Sports Exerc 2008;40:1205–12.
44 Oxfeldt M, Dalgaard LB, Jørgensen EB, *et al*. Molecular markers of skeletal muscle hypertrophy following 10 wk of resistance training in oral contraceptive users and nonusers. J Appl Physiol 2020;129:1355–64.
45 O'Leary TJ, Wardle SL, Gifford RM, *et al*. Tibial Macrostructure and Microarchitecture Adaptations in Women During 44Weeks of Arduous Military Training. Journal of Bone and Mineral Research 2021;36:1300–15.
46 Nappi C, Sardo ADS, Greco E, *et al*. Effects of an oral contraceptive containing drospirenone on bone turnover and bone mineral density. Obstetrics & Gynecology 2005;105:53–60.
47 Nappi C, Di Spiezio Sardo A, Acunzo G, *et al*. Effects of a low-dose and ultra-low-dose combined oral contraceptive use on bone turnover and bone mineral density in young fertile women: a prospective controlled randomized study. *Contraception* 2003;67:355–9.
48 Mazess RB, Barden HS. Bone density in premenopausal women: effects of age, dietary intake, physical activity, smoking, and birth-control pills. *Am J Clin* Nutr 1991;53:132–42.
49 Massaro M, Di Carlo C, Gargano V, *et al*. Effects of the contraceptive patch and the vaginal ring on bone metabolism and bone mineral density: a prospective, controlled, randomized study. Contraception 2010;81:209–14.
50 Massai R, Mäkäräinen L, Kuukankorpi A, *et al*. The combined contraceptive vaginal ring and bone mineral density in healthy pre-menopausal women. Hum Reprod 2005;20:2764–8.
51 Martin D, Cooper SB, Tang JCY, *et al*. Bone metabolic marker concentrations across the Menstrual cycle and phases of combined oral contraceptive use. *Bone* 2021;145:.:S8756-3282(21)00026-0.
52 LIU R-jian, Gai L, JIA Y-fang, Liu R, Jia Y, *et al*. Effect of two kinds of different combined oral contraceptives used on bone mineral density in women of reproductive age. *Journal of Reproduction and Contraception* 2011;22:145–51.
53 Liu B, Balkwill A, Cooper C, *et al*. Reproductive history, hormonal factors and the incidence of hip and knee replacement for osteoarthritis in middle-aged women. *Ann Rheum Dis* 2009;68:1165–70.
54 Liederbach M, Dilgen FE, Rose DJ. Incidence of anterior Cruciate ligament injuries among elite ballet and modern dancers: A 5-year prospective study. *Am J Sports Med* 2008;36:1779–88.
55 Leung Y-Y, Talaei M, Ang L-W, *et al*. Reproductive factors and risk of total knee replacement due to severe knee osteoarthritis in women, the Singapore Chinese Health study. *Osteoarthritis Cartilage* 2019;27:1129–37.
56 Lee H, Petrofsky JS, Yim J. Do oral contraceptives alter knee ligament damage with heavy exercise? *Tohoku J Exp Med* 2015;237:51–6.
57 Kelsey JL, Bachrach LK, Procter-Gray E, *et al*. Risk factors for stress fracture among young female cross-country runners. *Med Sci Sports Exerc* 2007;39:1457–63.
58 Jackowski SA, Baxter-Jones ADG, McLardy AJ, *et al*. The associations of exposure to combined hormonal contraceptive use on bone mineral content and areal bone mineral density accrual from adolescence to young adulthood: a longitudinal study. Bone Rep 2016;5:e333–41.
59 Herzog MM, Young JC, Lund JL, *et al*. Oral contraceptive use and anterior cruciate ligament injury: comparison of active comparator new user cohort and case-control study designs. Inj. Epidemiol. 2020;7.
60 Hellevik AI, Nordsletten L, Johnsen MB, *et al*. Age of menarche is associated with knee joint replacement due to primary osteoarthritis (The HUNT Study and the Norwegian Arthroplasty Register). *Osteoarthritis and Cartilage* 2017;25:1654–62.
61 He W, Dam TV, Thøgersen R, *et al*. Fluctuations in metabolites and bone markers across the menstrual cycle in eumenorrheic women and oral contraceptive users. *The Journal of Clinical Endocrinology Metabolism* 2022;107:1577–88.
62 Hartard M, Kleinmond C, Luppa P, *et al*. Comparison of the skeletal effects of the progestogens desogestrel and levonorgestrel in oral contraceptive preparations in young women: controlled, open, partly randomized investigation over 13 cycles. Contraception 2006;74:367–75.
63 Hansen MA, Overgaard K, Riis BJ, *et al*. Potential risk factors for development of postmenopausal osteoporosis--examined over a 12-year period. Osteoporos Int 1991;1:95–102.
64 Gersten J, Hsieh J, Weiss H, *et al*. Effect of Extended 30 µg Ethinyl Estradiol with Continuous Low-Dose Ethinyl Estradiol and Cyclic 20 µg Ethinyl Estradiol Oral Contraception on Adolescent Bone Density: A Randomized Trial. J Pediatr Adolesc Gynecol 2016;29:635–42.
65 Gargano V, Massaro M, Morra I, *et al*. Effects of two low-dose combined oral contraceptives containing drospirenone on bone turnover and bone mineral density in young fertile women: a prospective controlled randomized study. Contraception 2008;78:10–15.
66 Gai L, Jia Y, Zhang M, *et al*. Effect of two kinds of different combined oral contraceptives use on bone mineral density in adolescent women. Contraception 2012;86:332–6.
67 Elgán C, Samsioe G, Dykes AK. Influence of smoking and oral contraceptives on bone mineral density and bone remodeling in young women: a 2-year study. Contraception 2003;67:439–47.
68 Elgán C, Dykes A-K, Samsioe G. Bone mineral density changes in young women: a two year study. Gynecol Endocrinol 2004;19:169–77.
69 Dalgaard LB, Jørgensen EB, Oxfeldt M, *et al*. Influence of second generation oral contraceptive use on adaptations to resistance training in young untrained women. J Strength Cond Res 2022;36:1801–9.
70 Dalgaard LB, Dalgas U, Andersen JL, *et al*. Influence of oral contraceptive use on adaptations to resistance training. Front Physiol 2019;10:824.
71 Cromer BA, Bonny AE, Stager M, *et al*. Bone mineral density in adolescent females using injectable or oral contraceptives: a 24-month prospective study. Fertil Steril 2008;90:2060–7.
72 Cooper C, Hannaford P, Croft P, *et al*. Oral contraceptive pill use and fractures in women: a prospective study. *Bone* 1993;14:41–5.
73 Cobb KL, Kelsey JL, Sidney S, *et al*. Oral contraceptives and bone mineral density in white and black women in cardia. Osteoporosis International 2002;13:893–900.
74 Cobb KL, Bachrach LK, Sowers M, *et al*. The effect of oral contraceptives on bone mass and stress fractures in female runners. *Med Sci Sports Exerc* 2007;39:1464–73.
75 Combined hormonal contraceptives use and bone mineral density changes in adolescent and young women in a prospective population-based Canada-wide observational study. J Musculoskelet Neuronal Interact 2018;18:227–36.
76 Bonny AE, Secic M, Cromer BA. A longitudinal comparison of body composition changes in adolescent girls receiving hormonal contraception. Journal of Adolescent Health 2009;45:423–5.
77 Biason TP, Goldberg TBL, Kurokawa CS, *et al*. Low-Dose combined oral contraceptive use is associated with lower bone mineral content variation in adolescents over a 1-year period. BMC Endocr Disord 2015;15:15.
78 Berenson AB, Rahman M, Breitkopf CR, *et al*. Effects of depot medroxyprogesterone acetate and 20-microgram oral contraceptives on bone mineral density. Obstet Gynecol 2008;112:788–99.
79 Berenson AB, Breitkopf CR, Grady JJ, *et al*. Effects of hormonal contraception on bone mineral density after 24 months of use. *Obstet* Gynecol 2004;103:899–906.
80 Beksinska ME, Kleinschmidt I, Smit JA, *et al*. Bone mineral density in a cohort of adolescents during use of norethisterone enanthate, depot-medroxyprogesterone acetate or combined oral contraceptives and after discontinuation of norethisterone enanthate. *Contraception* 2009;79:345–9.
81 Barad D, Kooperberg C, Wactawski-Wende J, *et al*. Prior oral contraception and postmenopausal fracture: A women's health initiative observational cohort study. *Fertil Steril* 2005;84:374–83.
82 Almstedt HC, Cook MM, Bramble LF, *et al*. Oral contraceptive use, bone mineral density, and bone turnover markers over 12 months in college-aged females. *J Bone Miner Metab* 2020;38:544–54.
83 Allaway HCM, Misra M, Southmayd EA, *et al*. Are the effects of oral and vaginal contraceptives on bone formation in young women mediated via the growth hormone-IGF-I axis *Front Endocrinol (Lausanne)* 2020;11:334.
84 Poulsen E, Goncalves GH, Bricca A, *et al*. Knee osteoarthritis risk is increased 4-6 fold after knee injury - a systematic review and meta-analysis. *Br J Sports Med* 2019;53:1454–63.
85 Berger C, Goltzman D, Langsetmo L, *et al*. Peak bone mass from longitudinal data: Implications for the prevalence, pathophysiology, and diagnosis of osteoporosis. *J Bone Miner Res* 2010;25:1948–57.
86 Vandenbroucke JP, von Elm E, Altman DG, *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. *PLoS Med* 2007;4:e297e297.
87 Moher D, Hopewell S, Schulz KF, *et al*. Explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1–37.
88 Patel AD, Bullock GS, Wrigley J, *et al*. Does sex affect second ACL injury risk? A systematic review with meta-analysis. *Br J Sports Med* 2021;55:873–82.

Supplemental File: Extra Web Material

1. **Medline-Ovid Search Strategy**
2. **Downs and Black Quality Assessment Tool**
3. **Study Characteristics**
4. **Downs and Black Quality Assessment Tool Ratings**
5. **Semi-quantitative Analyses**

1. Medline-Ovid Search Strategy

1 exp contraceptives, oral/ or exp contraceptives, oral, combined/ or exp contraceptives, oral, hormonal/ or exp contraceptives, oral, sequential/ or exp contraceptives, oral, synthetic/ (49877)
Annotation: includes non mesh drug terms from each. Can review w/ Jerilynn
2 Hormonal Contraception/ (38)
3 ((combined or hormon* or oral) adj3 contracep*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (41640)
4 (birth control adj3 pill?).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (562)
5 1 or 2 or 3 or 4 [CHC mesh and keyword] (62439)
6 "bone and bones"/ or exp "bones of lower extremity"/ or exp "bones of upper extremity"/ or epiphyses/ or growth plate/ or exp rib cage/ or exp skull/ or exp spine/ (599910)
7 Bone Diseases, Metabolic/ or bone demineralization, pathologic/ or decalcification, pathologic/ or osteoporosis/ or osteoporosis, postmenopausal/ or bone resorption/ or osteochondritis/ or osteochondritis dissecans/ or osteochondrosis/ or spinal osteochondrosis/ or spinal diseases/ or intervertebral disc degeneration/ or intervertebral disc displacement/ or "ossification of posterior longitudinal ligament"/ or spinal osteophytosis/ or osteoarthritis, spine/ or spondylosis/ or spondylolysis/ or spondylolisthesis/ or osteosclerosis/ or exp Fractures, Bone/ or heel spur/ or osteophyte/ or Bone Density/ (330498)
8 (((bone* or hip or pelv* or humer* or femur or femoral or wrist or tibia* or fibula* or vertebr*) adj3 (fracture* or break* or broken)) or bone demineralization or bone decalcification or osteoporos* or bone resporption or osteochondr* or disc degeneration or degenerative disc or disc displacement or heel spur or osteosclero* or spondyl* or osteoarthritis or osteopathy* or ossification or disc displace* or bone mass or bone loss or bone densit* or bone mineral density or bone health).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (482133)
9 6 or 7 or 8 [bone mesh and keywords] (967426)
10 tendons/ or achilles tendon/ or hamstring tendons/ or patellar ligament/ or rotator cuff/ (41967)
11 tendinopathy/ or elbow tendinopathy/ or tennis elbow/ or enthesopathy/ or tendon entrapment/ or de quervain disease/ or trigger finger disorder/ or tenosynovitis/ (11085)
12 (tendon* or rotator cuff or patellar ligament* or tendin* or tenosynovitis or tennis elbow or enthesopathy or enthesitis or de quervain disease or trigger finger or epicondyl*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (106729)
13 10 or 11 or 12 [tendon mesh and keyword] (106729)
14 exp Muscle, Skeletal/ or fascia/ or fascia lata/ (274584)
15 muscular diseases/ or anterior compartment syndrome/ or ischemic contracture/ or fibromyalgia/ or medial tibial stress syndrome/ or Sarcopenia/ or Fasciitis, Plantar/ or Iliotibial Band Syndrome/ (39386)
16 (muscl* or muscl* or fascia* or rectus abdomin* or paraspinal or deltoid or gracilis or hamstring* or pectoral* or psoas or iliopsoas or quadricep* or tensor fascia lata or iliotibial band or it band or ITB or ITBS or compartment syndrome or contracture or medial tibial stress syndrome or MTSS or sarcopen* or

fasciitis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (1004055)
17 14 or 15 or 16 [muscle and fascia mesh and keyword] (1045510)
18 exp joints/ or exp Fibrocartilage/ [all joints, synovial bursa, articular cartilage, articular ligaments or intervertebral discs, meniscus, tfcc, palmar plate, plantar plate] (259368)
19 joint diseases/ or osteoarthritis/ or osteoarthritis, hip/ or osteoarthritis, knee/ or osteoarthritis, spine/ or bursitis/ or periarthritis/ or contracture/ or hip contracture/ or femoracetabular impingement/ or hallux limitus/ or hallux rigidus/ or joint dislocations/ or diastasis, bone/ or pubic symphysis diastasis/ or exp fracture dislocation/ or hip dislocation/ or knee dislocation/ or patellar dislocation/ or shoulder dislocation/ or joint instability/ or joint loose bodies/ or patellofemoral pain syndrome/ or shoulder impingement syndrome/ or synovitis/ or temporomandibular joint disorders/ or temporomandibular joint dysfunction syndrome/ or cartilage diseases/ or chondromalacia patellae/ or osteochondritis/ (183144)
20 ((joint# adj3 (disloc* or impinge* or sublux* or diastasis or instabil* or loose bodies)) or temporomandibular joint or TMJ or cartilag* or fibrocartilag* or menisc* or chondromalacia* or osteochondritis or ligament* or ACL or MCL or PCL or LCL or labrum or labral or articular or osteoarthritis or bursitis or periarthritis or impingement syndrome or hallux limitus or hallux rigidus or patellofemoral pain syndrome or PFPS or synovitis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (409560)
21 18 or 19 or 20 [joints, cartilage, and associated terms mesh and keyword] (554426)
22 9 or 13 or 17 or 21 [all msk injuries and conditions] (2237463)
23 5 and 22 [chc and all msk injuries and conditions] (2638)

2. Downs and Black Quality Assessment Tool[21]

Category	Item	Question	Scoring
Reporting	1	Is the hypothesis/aim clearly described?	Yes (1), No (0)
	2	Are the main outcomes clearly described?	Yes (1), No (0)
	3	Are the characteristics of patients included clearly described?	Yes (1), No (0)
	4	Are interventions clearly described?	Yes (1), No (0)
	5	Are the distributions of principal confounders clearly described?	Yes (2), Partially (1), No (0)
	6	Are the main findings clearly described?	Yes (1), No (0)
	7	Does the study provide estimates of random variability for main outcomes?	Yes (1), No (0)
	8	Have all important adverse events been reported?	Yes (1), No (0)
	9	Have the characteristics of participants lost to follow-up been described?	Yes (1), No (0)
	10	Have actual p-values been reported?	Yes (1), No (0)
External Validity	11	Were participants representative of the entire population?	Yes (1), No/unclear (0)
	12	Were people prepared to participate representative of the entire population?	Yes (1), No/unclear (0)
	13	Were staff/facilities used representative of the treatment majority of persons receive?	Yes (1), No/unclear (0)
Internal Validity	14	Were participants blinded to the intervention?	Yes (1), No/unclear (0)
	15	Were assessors blinded to intervention group?	Yes (1), No/unclear (0)
	16	Was data dredging made clear?	Yes (1), No/unclear (0)
	17	Were different follow up lengths adjusted for?	Yes (1), No/unclear (0)
	18	Were statistical tests appropriate?	Yes (1), No/unclear (0)
	19	Was compliance measured reliably?	Yes (1), No/unclear (0)
	20	Were main outcomes valid and reliable?	Yes (1), No/unclear (0)
	21	Were participants recruited from the same population?	Yes (1), No/unclear (0)
	22	Were participants recruited over the same time period?	Yes (1), No/unclear (0)
	23	Were participants randomized?	Yes (1), No/unclear (0)
	24	Was random assignment concealed?	Yes (1), No/unclear (0)
	25	Was there adequate adjustment for confounding?	Yes (1), No/unclear (0)
	26	Were losses to follow-up considered?	Yes (1), No/unclear (0)
Power	27	Did the study have sufficient power?	≤70% (0), 80% (1), 85% (2), 90% (3), 95% (4), 99% (5)

3. Study Characteristics

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Allaway, 2020 (Quasi, USA)	Never user: n (8 (23.6±1.0) Users: n (17 CHC Oral: n (9 (22.3±1.3) CHC Ring: n (8 (23.1±1.4)	Oral; EE/DG (30µg/150µg) 42-45 days (0 days) Ring; EE/DG (15µg/120µg) 42-45 days (0 days)	Never users (0 days)	IGF-I (serum) ng∗d/mL	Mean pre-post Baseline Never user: 154.7 ± 36.0 CHC Oral: 173.3 ± 28.2 CHC Ring: 117.8 ± 11.5 During Intervention (50-87 days) Never user: NR CHC Oral: NR CHC Ring: NR	NR	15
				P1NP (serum) ng/mL	Mean pre-post Baseline: Never user: 7.34 ± 2.15 CHC Oral: 11.93 ± 3.27 CHC Ring: 13.38 ± 4.97 During Intervention (50-87 days) Never user: 8.99 ± 2.09 CHC Oral: 4.61 ± 2.27 CHC Ring:4.98 ± 1.22	NR	
Almstedt, 2020 (PC, USA)	Never user: n (28 (19.3±.6) Ongoing user: n (34 (19.2±.5)	Oral; EE (20-35µg) 12 mo (1.9 ± 1.4 yr)	Never users (no use in past year)	LBM (DXA) kg	Time point mean (baseline vs. 12-mo) Never user: 39.9±4.6 vs. 40.1±4.5 Ongoing user: 42.0±4.6 vs. 42.5±4.4	NR	13
				CTX (serum) ng/ml	Time point mean (baseline vs. 6-mo) Never user: 13.8±5.3 vs. 14.2±8.5 Ongoing user: 18.6±8.2 vs. 20.4 ± 0.3	p (0.018)	
				WHOLE BODY BMD (DXA) g/cm^2	Time point mean (baseline vs. 12-mo) Never user: 1.043±0.01 vs. 1.055±0.01 Ongoing user: 1.037±0.01 vs. 1.041±0.01	NR	
Barad, 2005 (PC, USA)	Never users: n (47,922 (65.9±6.9) Previous Users: n (33,025 (60.0±6.5)	Oral; NR (NR) NR (NR)	NR	First fracture (self-report)	Crude rate (per 1000 person-years) Never user: 24 Previous user <5 years: 22 Previous ≥ 5 years: 20	Adjusted HR (95%CI) Overall: 1.07 (1.01,1.15) <5 years: 1.09 (1.01,1.18) 5-10 years: 1.07 (.96,1.20) ≥10 years: 1.02 (.91,1.14)	19
Beksinska, 2009 (PC, South Africa)	Never user: n (96 (17.4±1.2) New user: n (59 (17.8±1.0)	Oral; estrogen (93% used 30 and 40 µg) Up to 5 years (0 days)	Never users (0 days)	RADIUS BMD (DXA) g/cm^2	Adjusted mean % change Never user: 1.49 (1.25-1.72) New user: 0.84 (0.39-1.28)	p =0.01	17

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Berenson, 2004 (Quasi, USA)	Never user: n (44 (25.5±4.3) CHC A: n (25 (26.1±3.9) CHC B: n (42 (25.4±4.4)	CHC A: Oral EE/NO (0.035mg/1mg) 24 months (no use within 1 mo) CHC B: Oral EE/DG (0.030mg/ 0.15mg) 24 months (no use within 1 mo)	Never users (NR)	LUMBAR SPINE BMD (DXA) g/cm^2	Adjusted mean % change (baseline to 12-mo) Never user: -0.44 (-2.06, 1.16) CHC A: 2.12 (0.30, 3.93) CHC B: 0.17 (-1.56, 1.90) Adjusted mean % change (baseline to 24-mo) Never user: 1.80 (-0.33, 3.92) CHC A: -1.53 (-3.80, 0.73) CHC B: -2.57 (-4.63, -0.51)	NR Mean % change (95%CI) difference Never user vs CHC A: 0.67 (-1.54, 2.88) Never user vs CHC B: 1.51 (-0.40, 3.42)	20
Berenson, 2008 (PC, USA)	Never user: n (51 (16-33) New user: n (77 (16-33)	Oral; DG/EE2, placebo, EE2 (0.15mg/20µg 21 days, 2 days,10µg 5 days) 336 months (no use within 3 mo)	Never users (No use within 3 mo)	LUMBAR SPINE BMD (DXA) g/cm^2 FEMORAL NECK BMD (DXA) g/cm^2	Adjusted mean % change from baseline 6-mo: Never user: 0.51, New user: 0.18 12-mo: Never user: 0.91, New user: 0.20 18-mo: Never user: 1.33, New user: 0.08 24 mo: Never user: 1.66, New User: -0.01 30-mo: Never user: 1.93, New user: -0.19 36-mo: Never user: 1.94, New user: -0.54 Adjusted mean % change from baseline 6-mo: Never user: 0.05, New user: -0.22 12-mo: Never user: 0.15, New user: -0.30 18-mo: Never user: 0.29, New user: -0.54 24-mo: Never user: 0.54, New User: -0.76 30-mo: Never user: 0.66, New user: -1.00 36-mo: Never user: 0.61, New user: -1.29	6 mo p<.001 12 mo p<.001 18 mo p<.001 24 mo p<.001 30 mo p<.001 36 mo p<.001 6 mo p>.05 12 mo p<.05 18 mo p<.001 24 mo p<.001 30 mo p<.001 36 mo p<.001	20
Biason, 2015 (Quasi, Brazil)	Never user: n (26 (15.6; 14.7-16.1) New user: n (35 (15.8; 11.8-19.5)	Oral; DG/EE (50µg/20µg) 12 months (0 days)	Never users (0 days)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 12.16% New user: 2.07%	Mean difference in % changes 10.09%, p=0.056	15
				LUMBAR SPINE BMC(DXA) g	Mean % change (baseline to 12-mo) Never user: 16.84% New user: 1.57%	Mean difference in % changes 15.27%, p=0.014	
				WHOLE BODY BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 5.28% New user: 0.84%	Mean difference in % changes 4.44%, p=0.15	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				WHOLE BODY BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 11.34% New user: 1.22%	Mean difference in % changes 10.12%, p (0.031	
				SUBTOTAL BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 5.28% New user: 0.56%	Mean difference in % changes 4.72%, p (0.15	
				SUBTOTAL BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 16.04% New user: 1.18%	Mean difference in % changes 14.86%, p (0.033	
Bonny, 2009 (Quasi, USA)	Never user: n (18 (15.7±1.8) New user: n (18 (15.6±1.6)	Oral; NR (NR) (no use for 3-months)	Never users (no use for 3-months)	LBM (DXA) kg	Mean % change (baseline to 6-mo) Never users: 0.6% ± 3.4% New users: 0.6% ± 4.7%	p=0.07	14
				LUMBAR SPINE BMD (DXA) g/cm^2	NR	Mean difference in change (95% CI) 0.002 (-0.104, 0.091)	
Brajic, 2018 (PC, Canada)	Never user: n (78 (18.5 [18.0, 19.1]) Ongoing user: n (229 (19.8 [9.5, 20.2])	Oral, Ring; estrogen (avg 26.5µg/day, range 15-35) (mean age of starting CHC 17.5)	Never users (0-days)	FEMORAL NECK BMD (DXA) g/cm^2	NR	Mean difference in change (95% CI) -0.001 (-0.010, 0.008)	18
				TOTAL HIP BMD (DXA) g/cm^2	NR	Mean difference in change (95% CI) -0.001 (-0.009, 0.006)	
Cobb, 2002 (RC, USA)	Black never user: n (56 (31.2±4.0) past user: n (204 (31.5±3.6) White	Oral; EE(37.3 ± 11.5 µg) N/A 4.1 (IQR 7.1) years	Never users 4.1 (IQR 7.1) years	LUMBAR SPINE BMD (DXA) g/cm^2	Adjusted mean at 7 years Black Never user: 1.12 ± 0.11 Past user: 1.12 ± 0.13 White	Beta (± SE, R^2) -0.000005 ± 0.0002, 0%	16

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
	never user: n (60 (33.2±3.3) past user: n (156 (32.4±3.7)				Never user: 1.06 ± 0.11 Past user: 1.04 ± 0.12		
				WHOLE BODY BMD (DXA) g/cm^2	Adjusted mean at 7 years Black Never user: 1.16 ± 0.09 Past user: 1.16 ± 0.10 White Never user: 1.08 ± 0.07 Past user: 1.10 ± 0.08	Beta (± SE, R^2) -0.000054 ± 0.00012, 0.1%	
				TOTAL HIP BMD (DXA) g/cm^2	Adjusted mean at 7 years Black Never user: 1.03 ± 0.12 Past user: 1.04 ± 0.14 White Never user: 0.94 ± 0.11 Past user: 0.98 ± 0.11	Beta (± SE, R^2) -0.000012 ± 0.0002, 0%	
				LBM (DXA) kg	Adjusted mean at 7 years Black Never user: 44.6 ± 7.1 Past user: 44.4 ± 6.0 White Never user: 42.8 ± 5.4 Past user: 42.6 ± 4.6	NR	
				WHOLE BODY BMC (DXA) g	Yearly rate of change Eumenorrheic Never user: 3.7 ± 3.4 New user: 9.9 ± 3.9	Difference in mean yearly change rate (± SE) Eumenorrheic 6.2±5.2	
Cobb, 2007 (RCT, USA)	Never user: n (81 (21.9±2.6) New user: n (69 (22.3±2.7)	Oral; EE/NG (30 µg/ 0.3mg) 2 years (no use for 6-months)	Never users (no use for 6-months)	LUMBAR SPINE BMD (DXA) g/cm^2	Yearly rate of change Eumenorrheic Never user: 0.0002 ± 0.0016 New user: 0.0022 ± 0.0019	Difference in mean yearly change rate (± SE) Eumenorrheic 0.0020±0.0025	21
				TOTAL HIP BMD (DXA) g/cm^2	Yearly rate of change Eumenorrheic Never user: -0.0023 ± 0.0015 New user: 0.0013 ± 0.0017	Difference in mean yearly change rate (± SE) Eumenorrheic 0.0035±0.0022	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Cooper, 1993 (PC, UK)	n (NR (29)	Oral; NR (NR) 3.7 years (NR)	Never users (0 days)	stress fracture (questionnaire)	Incidence rate per 100 women-years Never users: 9.2 New user: 5.8	HR (95%CI) 0.57 (0.18, 1.83)	18
				Any Fracture (national database)	Incidence rate per 1000 women-years Never user: 2.6 Ongoing user: 2.99	Adjusted RR (95%CI) 1.20 (1.08,1.34)	
				Forearm Fracture (national database)	Incidence rate per 1000 women-years Never user: 0.67 Ongoing user: 0.66	Adjusted RR (95%CI) 1.06 (0.95,1.32)	
Cromer, 2008 (PC, USA)	Never user: n (95 (14.8±1.9) New user: n (62 (16.0±1.4)	Oral; EE/LNG (20µg/100µg) 2 years (no use for 3 months)	Never users (no use for 3 months)	LUMBAR SPINE BMD (DXA) g/cm^2	Adjusted mean (baseline vs. 6-mon vs. 12-mon vs. 18-mon vs. 24-mon) Never user: 0.98±0.01 vs. 1.00±0.01 vs. 1.02±0.01 vs. 1.03±0.01 vs. 1.04±0.01 New user: 1.01±0.01 vs. 1.02±0.01 vs. 1.03±0.01 vs. 1.03±0.01 vs. 1.03±0.01	Adjusted % Change (± SE), from baseline to 24 mo Never user: 6.3% ± 0.5% New user: 4.2% ± 0.7%	17
				FEMORAL NECK BMD (DXA) g/cm^2	Adjusted mean (baseline vs. 6-mon vs. 12-mon vs. 18-mon vs. 24-mon) Never user: 0.92±0.01 vs. 0.93±0.01 vs. 0.94±0.01 vs. 0.95±0.01 vs. 0.96±0.01 New user: 0.96±0.01 vs. 0.96±0.01 vs. 0.96±0.01 vs. 0.97±0.01 vs. 0.97±0.01	Adjusted % Change (± SE), from baseline to 24 mo Never user: 3.8% ± 0.8% New user: 3.0% ± 1.0%	
Dalgaard, 2019 (PC, DEN)	Never user: n (14 (24±1) Ongoing user: n (14 (24±1)	Oral; n (7 EE/GD (30µg/75µg) Oral; n (5 EE/GD (20µg/75µg) Oral; n (2 EE/DGn (20µg/150µg) 10 weeks (6.1 ± 5 years prior use)	Never users (NR)	Quadriceps CSA (MRI) mm^2	Mean % change (baseline to 10-weeks) Never user: 7.9% ± 0.1% Ongoing user: 10.8% ± 1.3%	Group-by-time interaction p=0.06	15
				Quadriceps Fiber Type CSA (Biopsy) µm^2	Mean pre-post (baseline vs. 10-weeks) Type I Never user: 4020±348 vs. 3777±354 Ongoing user: 3821±197 vs. 4490±313 Type II Never user: 3239±344 vs. 3691±361 Ongoing user: 3452±242 vs. 3891±387	Group-by-time interaction p=0.98	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				Fiber type composition (biopsy) %	Mean pre-post (baseline vs. 10 weeks) Type I Never user: 53.4%±2.7% vs. 52.4%±2.9% Ongoing user:46.9%±2.8% vs. 48.3%±2.5% Type IIa Never user: 39.4%±2.6% vs. 42.8%±2.1% Ongoing user: 42.6%±2.5% vs. 47.7%±2.5% Type Iix Never user: 7.1%±2.1% vs. 4.8%±1.2% Ongoing user: 10.5%±2.2% vs. 3.9%±1.5%	Group-by-time interaction Type I: p=0.52 Type IIa: p=0.64 Type Iix: p=0.05	
				tendon CSA (MRI) mm^2	Mean pre-post (baseline vs. 10 weeks) Proximal Never user: 77±3 vs. 85±5 Ongoing user: 81±6 vs. 87±5 Middle Never user: 80±3 vs. 97±7 Ongoing user: 78±4 vs. 90±4 Distal Never user: 100±5 vs. 109±5 Ongoing user: 95±5 vs. 101±5	Group-by-time interaction proximal: p=0.70 middle: p=0.57 distal: p=0.57	
				tendon collagen concentration (biopsy) mg/mg d.w; dry weight	Mean pre-post (baseline vs. 10 weeks) Never user: 0.61±0.03 vs. 0.62±0.04 Ongoing user: 0.62±0.02 vs. 0.64±0.02	Group-by-time interaction p=0.72	
				tendon collagen cross-linking (biopsy) pmol/pmol	Mean pre-post Baseline Collagen concentration 0.62±0.02 Hydroxylysyl pyridinoline/Collagen 0.73±0.06 Lysyl pyridinoline/Collagen 0.03±0.02 Pentosidine/Collagen 0.012±0.001 10 weeks Collagen concentration0.64/0.02 Hydroxylysyl pyridinoline/Collagen 0.80±0.05 Lysyl pyridinoline/Collagen 0.03±0.00 Pentosidine/Collagen 0.012±0.001 Baseline Collagen concentration 0.61±0.03 Hydroxylysyl pyridinoline/Collagen 0.63±0.06 Lysyl pyridinoline/Collagen 0.04±0.01 Pentosidine/Collagen 0.011±0.001 10 weeks Collagen concentration 0.62±0.04	Group-by-time interaction HP/Collagen: p=0.56 LyP/Collagen: p=0.13 Pentosidine/Collagen: p=0.44	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
					Hydroxylysyl pyridinoline/Collagen 0.65±0.06 Lysyl pyridinoline/Collagen 0.03±0.00 Pentosidine/Collagen 0.012±0.001		
Dalgaard, 2020 (PC, DEN)	Never user: n (18 (24.3±2.5) Ongoing user: n (20 (24.2±2.0)	Oral; EE(30-35µg) 10 weeks (6.5 ± 2.5 yrs prior use)	Never users (0 yrs)	Quadriceps CSA (MRI) at 10 cm above lateral epicondyle, cm^2	Mean pre-post (baseline vs. 10 weeks) Never user 33.1±4.2 vs. 36.3±5.2 Ongoing user 35.7±4.7 vs. 39.4±5.5 Mean % change (baseline vs. 10 weeks) Never user 9.7±4.9% Ongoing user 10.6±4.8%	p=0.46	16
				Quadriceps CSA (MRI) at 20 cm above lateral epicondyle, cm^2	Mean pre-post (baseline vs. 10 weeks) Never user: 54.5±5.2 vs. 59.5±5.6 Ongoing user: 54.4±9.5 vs. 59.6±10.3 Mean % change (baseline vs. 10 weeks) Never user: 9.2 ± 5.0% Ongoing user: 9.5 ± 6.0%	p=0.81	
				Quadriceps CSA (MRI) at 30 cm above lateral epicondyle, cm^2	Mean pre-post (baseline vs. 10 weeks) Never user: 53.7±7.8 vs. 58.6±7.5 Ongoing user: 52.1±9.1 vs. 57.8±10 Mean % change (baseline vs. 10 weeks) Never user: 9.2 ± 5.0% Ongoing user: 11.0 ± 6.0%	p=0.37	
				Quadriceps Type I CSA (biopsy), $µm^2$	Mean % change (baseline vs. 10 weeks) Never user: 6.4 ± 7.4% Ongoing user: 8.8 ± 7.6%	NS	
				Quadriceps Type II CSA (biopsy), $µm^2$	Mean % change (baseline vs. 10 weeks) Never user: 16.6 ± 7.2% Ongoing user: 19.9 ± 7.9%	NS	
				LBM (DXA) kg	Mean pre-post Baseline Never user: 43.9±5.0 Ongoing user: 42.9±5.0 10 weeks Never user: 45.1±5.0 Ongoing user: 44.6±5.0	p=0.08	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Elgan, 2003 (RC, Sweden)	CHC-nonsmoker: n (35 (18-26) CHC-smoker: n (9 (18-26) CHC+nonsmoker: n (57 (18-26) CHC+smoker: n (17 (18-26)	Oral; NR(NR) Users: 2 yrs (4.3 ± 2.3 yrs)	Never users (NR)	CALCANEUS BMD (DXA) g/cm^2	Mean change CHC-nonsmoker: 0.0048±0.0312 CHC-smoker: -0.0330±0.0300 CHC+nonsmoker: -0.0069±0.0365 CHC+smoker: -0.0116±0.0428	Multivariable linear regression (CHC - nonsmoker is reference, ± SE) CHC-smoker: -0.03 ± 0.01, p (0.02 CHC+nonsmoker: -0.01 ± 0.01, p (0.07 CHC+smoker: -0.02 ± 0.01, p (0.01	14
				D-PYD (urine) nmol/L	Mean change CHC-nonsmoker: 0.5394±2.8025 CHC-smoker: -2.0000±2.8000 CHC+nonsmoker: -0.3679±1.7303 CHC+smoker: -0.5286±2.2812	Multivariable linear regression (CHC - nonsmoker is reference) CHC-smoker: -3.26 ± 0.92, p (0.001 CHC+nonsmoker: -1.50 ± 0.49, p (0.003 CHC+smoker: -1.72 ± 0.74, p (0.022	
Elgan, 2004 (RC, Sweden)	n (72 (21.5±2.2)	Oral; NR(NR) NR (NR)	Never users (NR)	CALCANEUS BMD (DXA) g/cm^2	NR	OR (95%CI) ≥5% BMD loss vs. ≥5% BMD gain 6.3 (1.6,25.7)	14
Gai, 2012 (Quasi, China)	Never user: n (115 (17.13±0.78) CHC A: n (127 (17.1±0.8) CHC B: n (134 (17.1±0.8)	CHC A: Oral; EE/DG (30µg/0.15mg) CHC B: Oral; EE/CA (35µg/2mg) 24-months (0 days)	Never users (0 days)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean pre-post (baseline vs. 12-mo vs. 24-mo) Never user: 1.01±0.11 vs. 1.02±0.11 vs. 1.03±0.11 CHC A: 1.01±0.11 vs. 1.01±0.11 vs. 1.01±0.11 CHC B: 1.01±0.11 vs. 1.01±0.11vs. 1.01±0.11 Mean % change (baseline vs. 24-mo) Never user: 1.88% CHC A: -0.30% CHC B: 0.30%	Baseline: p=0.99 12-mo: p=0.75 24-mo: p=0.34	15
				FEMORAL NECK BMD (DXA) g/cm^2	Mean pre-post (baseline vs. 12-mo vs. 24- mo) Never user: 0.82 ± 0.09 vs. 0.82 ± 0.09 vs. 0.82 ± 0.09 CHC A: 0.82 ± 0.09 vs. 0.82 ± 0.09 vs. 0.81 ± 0.09 CHC B: 0.82 ± 0.09 vs. 0.82 ± 0.09 vs. 0.82 ± 0.09 Mean % change (baseline vs. 24-mo) Never user: 0.98% CHC A: -0.61% CHC B: 0.49%	Baseline: p=0.97 12-mo: p=0.93 24-mo: p=0.56	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Gargano, 2008 (Quasi, Italy)	Never user: n (20 (25.7±6.4) CHC A: n (20 (26.1±4.9) CHC B: n (21 (28.1±3.7)	CHC A: Oral; EE/DP (30µg/3mg) CHC B: Oral; EE/DP (20µg/3mg) 12-months (NR)	Never users (NR)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean pre-post (baseline vs. 12-mo) Never user: 1.041±0.08 vs. 1.042±0.02 CHC A: 1.040±0.06 vs. 1.041±0.11 CHC B: 1.042±0.17 vs. 1.040±0.19	NS	13
				BGP (serum)	NR	NR	
				PYD (urine)	NR	NR	
				D-PYD (urine)	NR	NR	
Gersten, 2016 (Quasi, USA)	Never user: n (372 (14.8±1.72) CHC A: n (247 (16±1.61) CHC B: n (240 (15.9±1.71)	CHC A: Oral; 84 days EE/LNG (30µg/150µg), then 7 days EE (10µg) CHC B: Oral; 21 days EE/LNG (20µg/100µg) 12-months (NR)	Never users (NR)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 2.50±0.14% CHC A: 2.26±0.17% CHC B: 1.45±0.17%	Mean difference in % Change (95%CI) Never user vs. CHC A: 0.23 (-0.20, 0.67) Never user vs. CHC B: 1.05 (0.61, 1.49)	23
				LUMBAR SPINE BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 3.80±0.19 % CHC A: 3.53±0.23 % CHC B: 2.34±0.24%	Mean difference in % Change (95%CI) Never user vs. CHC A: 0.27 (-0.33, 0.87) never user vs. CHC B: 1.45 (0.85, 2.06)	
				FEMORAL NECK BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 1.12±0.13% CHC A: 1.77±0.15% CHC B: 1.80±0.16%	Mean difference in % Change (95%CI) Never user vs. CHC A: -0.65 (-1.05, -0.25) Never user vs. CHC B: -0.32 (-0.09, 0.72)	
				FEMORAL NECK BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 1.51±0.18% CHC A: 1.99±0.22% CHC B: 1.02±0.23	Mean difference in % Change (95%CI) Never user vs. CHC A: -0.48 (-1.05, 0.09) Never user vs. CHC B: 0.49 (-0.09, 1.07)	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Hansen, 1991 (PC, DEN)	Never user: n (90 (51±2) Previous user: n (31 (51±2)	Oral; NR(NR) 12-years (36±36mo)	Never user (NR)	WHOLE BODY BMD (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 1.75±0.14% CHC A: 1.32±0.14% CHC B: 1.35±0.14%	Mean difference in % Change (95%CI) Never user vs. CHC A: 0.43 (0.03, 0.82) Never user vs. CHC B: 0.40 (0.01, 0.80)	10
				WHOLE BODY BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 3.84±0.33% CHC A: 3.31±0.35% CHC B: 2.83±0.35%	Mean difference in % Change (95%CI) Never user vs. CHC A: 0.53 (-0.43, 1.48) Never user vs. CHC B: 1.01 (0.05, 1.96)	
				LUMBAR SPINE BMD (DXA) g/cm^2	Mean value (12 years after baseline) Never user: 0.88 ± 0.16 Previous user: 0.85 ± 0.14	NS	
				FEMORAL NECK BMD (DXA) g/cm^2	Mean value (12 years after baseline) Never user: 0.68 ± 0.10 Previous user: 0.64 ± 0.09	NS	
				TROCHANTER BMD (DXA) g/cm^2	Mean value (12 years after baseline) Never user: 0.59 ± 0.10 Previous user: 0.59 ± 0.09	NS	
				WARD'S TRIANGLE BMD (DXA) g/cm^2	Mean value (12 years after baseline) Never user: 0.48 ± 0.10 Previous user: 0.43 ± 0.09	p < 0.05	
				RADIUS BMC (SPA)	Mean value (12 years after baseline) Never user: 30.9 ± 5.9 Previous user: 31.8 ± 5.9	NS	
				RADIUS BMC early postmenopausal change (SPA)	Mean change Never user: -1.7 ± 1.9% Previous user: -2.3 ± 1.9%	NS	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Hartard, 2006 (Quasi, GER)	Never user: n (17 (21.1±1.5) CHC A: n (22 (20.6±1.7) CHC B: n (20 (20.8±2)	CHC A: Oral; EE/DG (20µg/150µg) CHC B: Oral; EE/LNG (20µg/100µg) User CHC A: 12 mo (2.4 ± 1.2yrs) User CHC B: 12 mo (1.7 ± 1.8yrs)	Never user (0.4 ± 1.2 yrs)	RADIUS BMC subsequent postmenopausal change (SPA)	Mean change Never user: -1.7 ± 0.8 Previous user: -1.9 ± 0.7%	NS	12
				aBMD1 (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: 0.47 ± 2.91% CHC A: -1.52 ± 1.80% CHC B: -0.11 ± 3.01%	CHC A vs. Never user: p<0.05	
				LUMBAR SPINE BMC (DXA) g	Mean % change (baseline to 12-mo) Never user: 0.62 ± 3.06% CHC A: -1.10 ± 2.24% CHC B: -0.52 ± 2.68	CHC A vs. Never user: p<0.05	
				aBMD2 (DXA) g/cm^2	Mean % change (baseline to 12-mo) Never user: -.69 ± 3.62% CHC A: -0.30 ± 3.83% CHC B: -0.22 ± 4.38%	NS	
				aBMD5 (DXA) mg/cm^3	Mean % change (baseline to 12-mo) Never user: -1.03 ± 2.97% CHC A: -0.35 ± 4.70% CHC B: -1.95 ± 3.15%	NS	
				aBMD10 shank 4% (DXA) mg/cm^3	Mean % change (baseline to 12-mo) Never user: 0.38 ± 2.50% CHC A: -0.83 ± 1.96% CHC B: -1.04 ± 2.59%	NS	
				aBMD10 shank 14% (DXA) mg/cm^3	Mean % change (baseline to 12-mo) Never user: 0.59 ± 1.24% CHC A: 0.45 ± 0.96% CHC B: -0.41 ± 1.33%	CHC B vs. Never user: p<0.05	
				aBMD10 shank 38% (DXA) mg/cm^3	Mean % change (baseline to 12-mo) Never user: 0.57 ± 0.63% CHC A: 0.22 ± 0.68% CHC B: 0.36 ± 0.57%	NS	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Hellevik, 2017 (PC, NOR)	Never user: n (6,202 (55.7±15.2) Previous user: n (11,924 (55.7±15.2)	Oral; NR(NR) User: N/A (90,646 person years)	Never user (NR)	TKR (medical records) number of cases	Never user: 130 Previous user: 103	Adjusted HR (95%CI) vs. never users 1.36 (1.00, 1.86)	18
				THR (medical records) number of cases	Never users: 193 Previous user: 133	Adjusted HR (95%CI) vs. never users 1.03 (0.79, 1.35)	
Herzog, 2020 (RC, USA)	Never user: n (621,798 (32.4±6.8) New user: n (2,370,286 (26.7±8.1)	Oral; EE (≤35µg) up to 14.5 years (no use ≥180 days)	Never users (no use ≥180 days)	ACL injury (clinical diagnosis, reconstruction) cases	Number of cases (%) Never user: 1620 (0.26%) New user: 3571 (0.15%)	Adjusted HR (95%Ci) 0.95 (0.89, 1.01)	19
Jackowski, 2016 (RC, Canada)	Never user: n (43 (16.3±5.6) Ongoing user: n (67 (18.0±6.1)	Oral; NR(NR) Users: N/A (4.9 ± 3.9 yrs)	Never Users (0 days)	aBMD1 (DXA) g/cm^2	NR	NS	15
				LUMBAR SPINE BMC (DXA) g	NR	NS	
				aBMD2 (DXA) g/cm^2	NR	NS	
				FEMORAL NECK BMC (DXA) g	NR	NS	
				aBMD3 (DXA) g/cm^2	NR	mean (± SE) -0.0099 ± 0.0042	
				WHOLE BODY BMC (DXA) g	NR	NS	
Kelsey, 2007 (PC, USA)	n (127(22.0±2.6)	Oral; NR(NR) 2 years (no use within 6 months)	Never Users (no use within 6 months)	stress fracture (imaging)	NR	Adjusted rate ratio (95%CI) 2.22 (0.65, 7.69)	15
Lee, 2015 (Quasi, USA)	Never user: n (25(25.2±1.6) Ongoing user: n (15 (25.1±2.8)	Oral; EE (30-55µg) 5 days (at least 1 yr)	Never Users (NR)	Anterior Tibial Translation (KT-2000) mm	Baseline mean Never user: 5.3 ± 1.0 Ongoing user: 4.5 ± 0.6	p=0.01	12
Leung, 2019 (PC, Singapore)	Never users: n (25,905 (57±8.3) Previous user: n (9,280 (53.3±6.2)	Oral; NR(NR) Users: N/A (NR)	Never Users (NR)	TKR (medical record) count	Number of cases Never users: 1163 Previous users: 482	Adjusted HR (95%CI), never user reference 1.18 (1.05, 1.32)	18

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Liederbach, 2008 (PC, USA)	Never user: 47 (18-41) Ongoing user: 136 (18-41)	Oral; NR(NR) NR (NR)	Never Users (NR)	ACL injury (clinical exam/imaging) count	Number of cases Never user: 5 Ongoing user: 5	p=0.13	14
Liu, 2011 (Quasi, China)	Never user: n (53 (29.9±4.0) CHC A: n (46 (29.3±4.1) CHC B: n (55 (29.0±3.9)	CHC A:Oral; EE/DG (30µg/0.15mg) CHC B: Oral; EE/CA (35µg/2mg) 2 years (no use ≥ 6 months)	Never Users (no use ≥ 6 months)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean pre-post (baseline vs. 24-mo) Never user: 1.109±0.112 vs. 1.108±0.109 CHC A: 1.110±0.114 vs. 1.106±0.109 CHC B: 1.109±0.111 vs. 1.110±0.111	Baseline: p=0.99 24 month: p=0.98	17
				FEMORAL NECK BMD (DXA) g/cm^2	Mean pre-post (baseline vs. 24-mo) Never user: 0.913±0.088 vs. 0.913±0.091 CHC A: 0.914±0.089 vs. 0.899 ± 0.092 CHC B: 0.912±0.091 vs. 0.912±0.091	Baseline: p=0.99 24 months: p=0.70	
Liu, 2009 (PC, UK)	Never user: n (519,734 (56.0±4.7) Previous user: n (772,033 (56.0±4.7)	Oral; NR(NR) N/A (NR)	Never Users (NR)	TKR (medical record) count	Number of cases Never user: 5025 Previous user: 4774	adjusted RR (95%CI) 1.00 (0.96,1.04)	18
				THR (medical record) count	Number of cases Never user: 5850 Previous user: 6118	adjusted RR (95%CI) 1.02 (0.98,1.06)	
Massai, 2005 (Quasi; Finland, Chile, the Netherlands)	Never user: n (31 (29.1±4.1) Ongoing user: n (76 (26.6±4.9)	Ring; EE/ET (15µg/120µg) 24 months (no use ≥ 1 month)	Never Users (NR)	LUMBAR SPINE BMD (DXA) g/cm^2	Z-score change (baseline to 12-mo) Never user: 0.212 ± 0.254 New User: 0.058 ± 0.212 Z-score change (baseline to 24-mo) Never user: 0.257 ± 0.328 New User: 0.093 ± 0.278	Difference of mean change (95%CI) 12-months: 12 −0.222 (−0.369, −0.076), p (0.003 24 months: −0.341 (−0.473, −0.208), p< 0.0001	14

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Massaro, 2010 (Quasi, Italy)	Never user: n (17 (25.2±6.4) CHC patch: n (16 (27.3±2.7) CHC Ring: n (16 (26.0±5.4)	patch; EE/NGMN (20µg/150µg) Ring; EE/ET (15µg/120µg) 12 months (NR)	Never Users (NR)	FEMORAL NECK BMD (DXA) g/cm²	Z-score change (baseline to 12-mo) Never user: 0.085 ± 0.336 New User: 0.057 ± 0.233 Z-score change (baseline to 24-mo) Never user: 0.223 ± 0.286 New User: 0.061 ± 0.284	Difference of mean change (95%CI) 12-months: −0.156 (−0.332, 0.019), p(0.080 24-months: −0.267 (−0.383, −0.151), p< 0.0001	17
				LUMBAR SPINE BMD (DXA) g/cm²	Mean (Baseline vs. 12-month) Never user: 1.041±0.08 vs. 1.042±0.02 CHC patch: 1.040±0.12 vs. 1.041±0.0 CHC Ring: 1.042±0.15 vs. 1.041±0.18	NS	
				BGP (serum)	NR	p<0.05	
				PYD (urine)	NR	p<0.05	
				D-PYD (urine)	NR	NS	
				LUMBAR SPINE BMD (dual-photon absorptiometry) g/cm²	Mean % Change (baseline to 24-mo) Never user: 0.33 ± 4.2% < 5 yrs CHC use: 0.09 ± 3.2% > 5 yrs CHC use: -0.02 ± 4.0 %	NS	
Mazess, 1991 (PC, USA)	n (300 (20-39)	Oral; NR(NR) NR(NR)	Never Users (NR)				8
				RADIUS BMD one-third (single-photon absorptiometry) g/cm²	Mean % Change (baseline to 20-mo) Never user: -1.12 ± 4.7% < 5 yrs: 0.42 ± 4.2% > 5 yrs: -0.84 ± 5.0%	NS	
Nappi, 2003 (Quasi, Italy)	Never user: n (19 (29.2±4.8)	CHC A: Oral; EE/GD (20µg/75µg)		LUMBAR SPINE BMD (DXA) g/cm²	NR	NS	16

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Nappi, 2005 (Quasi, Italy)	CHC A: n (19 (28.7±6.2) CHC B: n (18 (29±5.8)	CHC B: Oral; EE/GD (15µg/60µg) 12 months (NR)	Never users (NR)	BGP (serum)	NR	NS	24
				PYD (urine)	NR	p <0.05	
				D-PYD (urine)	NR	p <0.05	
	Never user: n (22 (28.1±6.1) CHC A: n (23 (27.2±5.3) CHC B: n (22 (26.9±5.5)	CHC A: Oral; EE/DP (30µg/3mg) CHC B: EE/GD (30µg/75µg) 12 months (NR)	Never Users (NR)	LUMBAR SPINE BMD (DXA) g/cm^2	Mean (baseline vs. 12-mo) Never user: 1.042±0.16 vs. 1.039±0.09 CHC A: 1.039±0.08 vs.1.065±0.11 CHC B: 1.041±0.09 vs. 1.047±0.10	NS	
				PYD (urine)	NR	p <0.05	
				D-PYD (urine)	NR	p<0.05	
				BGP (serum)	NR	NR	
Procter-Gray, 2008 (RCT, USA)	Never user: n (53 (21.9±2.6) New user: n (48 (22.3±2.7)	Oral; EE/NG (30µg/0.3mg) 24 months (no use ≥ 6 months)	Never users (no use ≥ 6 months)	LBM (DXA) kg/yr	Mean annual rate of change irregular menstrual group Never user 0.30±0.28 New user: 0.32±0.29 regular menstrual group Never user: -0.10±0.14 New user: 0.77±0.17	Mean difference in change rate ± SE Irregualr group: 0.02 ± 0.35, p (0.96 Regualr group: 0.77 ± 0.17, p< 0.0001	22
Reed, 2003 (PC, USA)	Never user: n (114 (18-39) Ongoing user: n (64 (18-39)	Oral; EE (30-35µg) 36 months (3.7 years [0.1 to 15 yrs])	Never Users (no use ≥12 months)	LUMBAR SPINE BMD (DXA) g/cm^2	Adjusted Mean (baseline to 36-mo) Never user: 1.06 Ongoing user: 1.06 % Change (baseline to 36-mo) Never user: 1.34% Ongoing user: 1.61%	p=0.65 p=0.73	16
				FEMORAL NECK BMD (DXA) g/cm^2	Adjusted Mean (baseline to 36-mo) Never user: 0.95 Ongoing user: 0.95 % Change (baseline to 36-mo) Never user: 0.12% Ongoing user: 0.48%	p=0.60 p=0.55	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Reiger, 2016 (PC, USA)	Never user: n (10 (20.2±1.0) Ongoing user: n (13 (20.5±1.8)	Oral; EE/PG (20µg-35µg/100µg-1000µg) 3 weeks (2.7 ± 1.9 yrs)	Never Users (NR)	WHOLE BODY BMD (DXA) g/cm^2	Adjusted Mean (baseline to 36-mo) Never user: 1.1 Ongoing user: 1.1 % Change (baseline to 36-mo) Never user: 0.66% Ongoing user: 0.68%	p=0.90 p=0.96	12
				BAP (serum)	(baseline) NR	NS	
				CTX (serum)	(baseline) NR	NS	
				LUMBAR SPINE BMD (DXA) g/cm^2	NR	NR	
				FEMORAL NECK BMD (DXA) g/cm^2	NR	NR	
Rome, 2004 (PC, USA)	Never user: n (152 (14.8±1.5) New user: n (165 (16±1.4)	Oral; NR(NR) 12 months (no use in past 6 months)	Never Users (no use in past 6 months)	BAP (serum)	Mean (12-mo, adjusted for baseline) Never user: 40.4±1.03 New user: 35.7±1.03	p=0.004	13
				D-PYD (urine) nmol/mmol	Mean (12-mo, adjusted for baseline) Never user: 9.8 ±1.03 New user: 9.0 ±1.03	p=0.08	
Scholes, 2011 (PC, USA)	Adolescent Never user: n (28 (16.4±0.1) Ongoing user: n (49 (16.8±0.1) Young women Never user: n (18 (24.1±0.3) Ongoing user: n (44 (24.6±0.3)	Oral; EE <30µg or 30-35µg Adolescent Users: 36 months (9.0 [0.8] months) Young Women Users: 36 months (19.2 [2.5] months)	Never user (no use for 2 yrs)	LUMBAR SPINE BMD (DXA) g/cm^2	Adjusted % change (baseline to 24-mo), Adjusted mean change (baseline to 36-mo) Adolescents Never user: 2.26%, 0.0216 Ongoing user (30-35 dose): 1.32%, 0.0115 Young women Never user: 0.35% Ongoing user: NR	Adolescents: NR Young Women: NS	17

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				WHOLE BODY BMD (DXA) g/cm^2	Adjusted % change (baseline to 24-mo), Adjusted mean change (0 to 36-mo) Adolescents Never user: 2.03%, 0.0214 Ongoing user (30-35 dose): 1.45%, 0.0146 Young women Never user: 0.90% Ongoing user: NR	Adolescents: NR Young Women: NS	
				TOTAL HIP BMD (DXA) g/cm^2	Adjusted % change (baseline to 24-mo) Adolescents Never user: 0.67% Ongoing user (30-35 dose): NR Young women Never user: -0.42% Ongoing user: NR	Adolescents: NR Young Women: NS	
Vessey ,1999 (RC, UK)	n (NR (25-39)	Oral; estrogen (≥50µg) N/A (5 to ≥97)	Never Users (0 days)	Spinal OA (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.3 (0.9, 1.7) Recently used: 1.0 (0.6, 1.6) Used in past: 1.3 (1.0, 1.8)	10
				Displaced cervicsal disc (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.5 (0.9, 2.5) Recently used: 1.3 (0.7, 2.6) Used in past: 1.6 (0.9, 2.8)	
				Displaced lumbar disc (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.1 (0.9, 1.4) Recently used: 1.1 (0.8, 1.5) Used in past: 1.1 (0.8, 1.4)	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Vessey, 1998 (RC, UK)	Never user: 123,000 woman-years (25-39) Ongoing user: 187,000 woman-years (25-39)	Oral; NR(NR) N/A (5 to ≥97)	Never Users (0 days)	Other displaced disc (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.0 (0.8, 1.3) Recently used: 1.0 (0.8, 1.4) Used in past: 1.0 (0.8, 1.3)	10
				Cervicalgia (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 0.9 (0.7, 1.1) Recently used: 1.0 (0.7, 1.4) Used in past: 0.8 (0.6, 1.0)	
				Backache (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.1 (0.9, 1.2) Recently used: 0.9 (0.7, 1.1) Used in past: 1.2 (1.0, 1.3)	
				Sprains/strains (medical record/referrals)	NR	adjusted RR (95%CI) Ever used: 1.0 (0.8, 1.2) Recently used: 1.0 (0.8, 1.4) Used in past: 0.9 (0.7, 1.2)	
				Any fracture (medical record/referrals)	NR	adjusted RR (95%CI) ≤1 year use: 0.8 (0.5, 1.2) 13-24 months: 0.9 (0.6, 1.3) 25-48 months: 1.2 (1.0, 1.5) 49-72 months: 1.2 (0.9, 1.4) 73-96 months: 1.2 (1.0, 1.5) ≥97 months: 1.2 (1.1, 1.4)	
				Forearm Fracture (medical record/referrals)	NR	adjusted RR (95%CI) ≤1 year use: 1.1 (0.3, 2.8) 13-24 months: 1.8 (0.8, 3.8) 25-48 months: 1.3 (0.7, 2.2) 49-72 months: 1.1 (0.6, 2.0) 73-96 months: 1.1 (0.6, 2.1) ≥97 months: 1.5 (1.1, 2.1)	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				Ankle Fracture (medical record/referrals)	NR	adjusted RR (95%CI) ≤1 year use: 0.7 (0.1, 2.1) 13-24 months: 1.6 (0.7, 3.2) 25-48 months: 0.9 (0.4, 1.6) 49-72 months: 0.7 (0.3, 1.3) 73-96 months: 1.3 (0.7, 2.3) ≥97 months: 1.0 (0.7, 1.5)	
				Tarsal/metatarsals (medical record/referrals)	NR	adjusted RR (95%CI) ≤1 year use: 0.4 (0.0, 1.5) 13-24 months: 0.9 (0.3, 2.2) 25-48 months: 1.2 (0.7, 2.0) 49-72 months: 1.2 (0.7, 2.0) 73-96 months: 1.2 (0.6, 2.0) ≥97 months: 0.8 (0.5, 1.2)	
				LUMBAR SPINE BMD (DXA) g/cm^2	Mean (baseline) Never users, Ex-: 1.28 ± 0.03 Never users, Ex+: 1.25 ± 0.02 Ongoing users, Ex-: 1.23 ± 0.02 Ongoing users, Ex+: 1.25 ±0.02	NS	
Weaver, 2001 (RCT, USA)	Never user, Ex-: n (24 (24.1±0.8) Never user, Ex+: n (37 (23.9 ± 0.7) Ongoing user, Ex-: n (40 (24.3±0.6) Ongoing user, Ex+: n (40 (24.1± 0.6)	Oral; EE (≤50µg) 24 months (NR)	Never Users (NR)	LUMBAR SPINE BMC (DXA) g	Mean (baseline) Never users, Ex-: 53.02 ± 2.06 Never users, Ex+: 50.13 ± 1.14 Ongoing users, Ex-: 48.84 ± 1.61 Ongoing users, Ex+: 49.88 ± 1.34	NS	12
				FEMORAL NECK BMD (DXA) g/cm^2	Mean (baseline) Never users, Ex-: 1.04 ± 0.03 Never users, Ex+: 1.02 ± 0.02 Ongoing users, Ex-: 1.00 ± 0.02 Ongoing users, Ex+: 1.01 ± 0.02	NS	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				WHOLE BODY BMC (DXA) g	Mean (baseline) Never users, Ex-: 2663 ± 77 Never users, Ex+: 2584 ± 45 Ongoing users, Ex-: 2502 ± 69 Ongoing users, Ex+: 2507 ± 54	NS	
				RADIUS BMD (DXA) g/cm^2	Mean (baseline) Never users, Ex-: 0.70 ± 0.01 Never users, Ex+: 0.96 ± 0.01 Ongoing users, Ex-: 1.00 ± 0.02 Ongoing users, Ex+: 1.01±0.02	NS	
				RADIUS BMC (DXA) g/cm	Mean (baseline) Never users, Ex-: 0.90 ± 0.02 Never users, Ex+: 0.85 ± 0.01 Ongoing users, Ex-: 0.86 ± 0.02 Ongoing users, Ex+: 0.85 ± 0.02	NS	
				TROCHANTER BMD (DXA) g/cm^2	Mean (baseline) Never users, Ex-: 0.81 ± 0.02 Never users, Ex+: 0.79 ± 0.02 Ongoing users, Ex-: 0.79 ± 0.02 Ongoing users, Ex+: 0.78 ± 0.02	NS	
				WARD'S TRIANGLE BMD (DXA) g/cm^2	Mean (baseline) Never users, Ex-: 1.01 ± 0.03 Never users, Ex+: 0.99 ± 0.02 Ongoing users, Ex-: 0.96 ± 0.03 Ongoing users, Ex+: 0.98 ± 0.02	NS	
				Studies Added in Updated Search			
He, 2022 (PC, DEN)	Never user/previous user: n (28 (23.8±2.7)	Oral; EE/LNG (30µg/150µg)	Never	PINP (serum biomarker)	Average PINP concentration lower duRing menstrual/pill cycle in ongoing users	p=0.108	12

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
	Ongoing user: n (10 (23.7±2.0)	Ongoing user: 28 days (64.5 ± 26.2 months)	user/previous user: 0 (31.9 ± 44.1 months)	CTX (serum biomarker)	Average CTX concentration lower duRing the menstrual/pill cycle in ongoing users	p <0.05	
				PINP (serum biomarker) $ng \cdot mL^{-1}$	Mean values across menstural cycle/CHC cycle Never user: 64.9±21.9 Ongoing user: 62.9±22.1	p=0.81	
Martin, 2021 (PC, UK)	Never user: n (14 (21±2) Ongoing user: n (14 (22±4)	Oral; NR(NR) Users: 28 days (≥ 6 mo)	Never Users (NR)	β-CTX (serum biomarker) $ng \cdot L^{-1}$	Mean values across menstural cycle/CHC cycle Never user: 560±180 Ongoing user: 500±200	p=0.37	8
				Bone ALP (serum biomarker) $U \cdot L^{-1}$	Mean values across menstural cycle/CHC cycle Never user: 18.9±5.4 Ongoing user: 17.6±3.8	p=0.47	
				Type I fiber CSA (biopsy) μm^2	Mean pre-post (baseline vs. 10 weeks) Never user 4,658 ± 200 5,056 ± 225 Ongoing user 4,418 ± 187 4,850 ± 269	p=0.97	
				Type 2 fiber CSA (biopsy) μm^2	Mean pre-post (baseline vs. 10 weeks) Never user 4,753 ± 254 5,431 ± 244 Ongoing user 4,241 ± 202 5,125 ± 220	p=0.5	
Oxfeldt, 2020 (PC, DEN)	Never user: n (18 (24±3) Ongoing user: n (20 (24±2)	Oral; NR(NR) Users: 10 weeks (NR)	Never Users (NR)	Myonuclei total fiber (biopsy) per fiber	Mean pre-post (baseline vs. 10 weeks) Never user 1.72 ± 0.13 1.88 ± 0.16 Ongoing user 1.53 ± 0.14 1.64 ± 0.13	p=0.94	10
				Myonuclei Type I (biopsy) per fiber	Mean pre-post (baseline vs. 10 weeks) Never user 3.03 ± 0.18 3.14 ± 0.18 Ongoing user 2.85 ± 0.16 2.79 ± 0.12	p=0.58	
				Myonuclei Type II (biopsy) per fiber	Mean pre-post (baseline vs. 10 weeks) Never user 3.49 ± 0.19 3.88 ± 0.27 Ongoing user 3.41 ± 0.22 3.76 ± 0.23	p=0.95	
				Myonuclear domain Type I (biopsy) μm^2/myonuclei	Mean pre-post (baseline vs. 10 weeks) Never user 667 ± 55.9 628 ± 34.4 Ongoing user 647 ± 27.6 599 ± 25.9	p=0.64	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
Sung, 2022 (PC, GER)	Never user: (muscle thickness group n (40, fibre composition group n (14) 25.00±4.56 Ongoing user: (muscle thickness group n (34, fibre composition group n (12) 22.39±2.30	Oral; EE (20-30µg) Users: 20 weeks (minimum 12 months)	Never users (no use in past year)	Myonuclear domain Type II (biopsy) µm²/myonuclei	Mean pre-post (baseline vs. 10 weeks) Never user 821 ± 52.6 772 ± 52.6 Ongoing user 763 ± 46.7 731 ± 52.8	p=0.99	9
				Myosin heavy chain protein distribution Type I (biopsy) %	Mean pre-post (baseline vs. 10 weeks) Never user 51.1 ± 2.2 53.3 ± 1.2 Ongoing user 52.6 ± 2.2 49.8 ± 1.8	p=0.08	
				Myosin heavy chain protein distribution Type IIa (biopsy) %	Mean pre-post (baseline vs. 10 weeks) Never user 45.0 ± 2.3 44.9 ± 1.2 Ongoing user 39.9 ± 1.5a 46.8 ± 1.4	p<0.01	
				Myosin heavy chain protein distribution Type IIx (biopsy) %	Mean pre-post (baseline vs. 10 weeks) Never user 3.8 ± 0.9 1.8 ± 0.6 Ongoing user 7.5 ± 1.3 3.4 ± 0.7	p=0.57	
				Muscle thickness of rectus femoris, vastus intermedius, vastus lateralis (ultrasound) cm²	Mean pre-post (baseline vs. 12 weeks) Never user: 6.13±1.08 vs 6.61±1.16 Ongoing user 5.98±0.57 vs 6.48±0.77	p=0.89	
				Muscle fibre thickness Type I (biopsy) µm	Mean pre-post (baseline vs. 12 weeks) Never user: 53.43±6.51 vs 56.83±6.51 Ongoing user: 53.45±6.33 vs 54.29±5.95	p=0.43	
				Muscle fibre thickness Type II (biopsy) µm	Mean pre-post (baseline vs. 12 weeks) Never user: 46.24±7.67 vs 53.39±6.63 Ongoing user: 53.45±6.33 vs 54.29±5.95	p=0.43	
				Muscle nucleus-to-fibre Type I (biopsy) ratio	Mean pre-post (baseline vs. 12 weeks) Never user: 3.04±0.63 vs 3.65±1.02 Ongoing user: 3.20±0.65 vs 3.35±0.77	p=0.26	
				Muscle fibre ratio Type I (biopsy) %	Mean pre-post (baseline vs. 12 weeks) Never user: 42.67±12.52 vs 40.81±12.61 Ongoing user: 44.12±15.00 vs 35.95±13.37	p=0.84	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				Muscle fibre ratio Type II (biopsy) %	Mean pre-post (baseline vs. 12 weeks) Never user: 57.33±12.52 vs 59.19±12.61 Ongoing user: 55.88±15.00 vs 60.05±13.37	p=0.84	
Yoo, 2021 (RC, Korea)	n (1 272 115 (61.0±8.1)	Oral; NR(NR)	Never users (never use: 79.8% of participants <1y: 9.2% of participants 1y+: 6.1% of participants unknown: 4.9% of participants)	Incident fracture (medical record) count	Number of cases Any fractures (189 883 (14.9%) Vertebral fractures (72 732 Hip fractures (11 153 Others fractures (106 895	OC use for 1 year or longer any fracture: aHR 1.03 (1.01-1.05) vertebral fracture: aHR 1.06 (1.03-1.09) hip fracture: aHR 1.06 (0.97-1.15) other fracture: aHR 1.03 (1.00-1.02)	23
				total vBMD10 4% site (HRpQCT) mg HA/cm^3	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 245±24 vs. 248±23 vs. 250±25 vs. 253±23 Ongoing user: 240 ± 21 vs. 243 ± 21 vs. 246 ± 23 vs. 250 ± 21	p≥0.3	
O'Leary, 2021 (PC, UK)	Never user: 11 Ongoing user: 18 (24±2)	Oral; NR(NR) Users: 44 weeks (NR)	Never Users (NR)	trabecular vBMD10 4% site (HRpQCT) mg HA/cm^3	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 203±24 vs. 205±23 vs. 207±24 vs. 210±22 Ongoing user: 197±18 vs. 199±16 vs. 202±17 vs. 204±15	p≥0.3	15
				cortical vBMD10 4% site (HRpQCT) mg HA/cm^3	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 745±34 vs. 744±33 vs. 739±42 vs. 741±34 Ongoing user: 748±48 vs. 750±48 vs. 745±57 vs. 754±52	p≥0.3	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				tibial trabecular area 4% site (HRpQCT) mm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 962±110 vs. 961±110 vs. 959±109 vs. 957±110 Ongoing user: 947±127 vs. 946±127 vs. 945±127 vs. 944±128	p≥0.19	
				tibial trabecular bone volume 4% site (HRpQCT) %	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 29.1±3.5 vs. 29.5±3.3 vs. 29.6±3.5 vs. 30.1±3.4 Ongoing user: 27.5±2.9 vs. 27.8±2.6 vs. 28.1±2.8 vs. 28.5 ± 2.5	p≥0.19	
				tibial cortical area 4% site (HRpQCT) mm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 82±11 vs. 83±10 vs. 85±10 vs. 86±11 Ongoing user: 81±12 vs. 83±13 vs. 84±13 vs. 85±13	p≥0.19	
				tibial cortical thickness 4% site (HRpQCT) mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.71±0.11 vs. 0.72±0.10 vs. 0.72±0.09 vs. 0.74± 0.10 Ongoing user: 0.72±0.14 vs. 0.73±0.15 vs. 0.74±0.16 vs. 0.75± 0.16	p≥0.19	
				tibial cortical perimeter 4% site (HRpQCT) mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 129.7±7.7 vs. 129.5±7.5 vs. 131.6±8.8 vs. 130.9±8.2 Ongoing user: 127.7±8.4 vs. 127.6±8.3 vs. 128.6±9.2 vs. 127.7±8.6	p≥0.19	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				tibial trabecular thickness 4% site (HRpQCT) mm	Median (IQR) pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.239 (0.230, 0.245) vs. 0.242 (0.231, 0.249) vs. 0.251 (0.237, 0.254) vs. 0.248 (0.234, 0.258) Ongoing user: 0.230 (0.226, 0.244) vs. 0.231 (0.225, 0.240) vs. 0.237 (0.230, 0.257) vs. 0.238 (0.232, 0.251)	p≤0.05 contraception × time interaction Trabecular thickness increased in COCP users from week 1 to week 28 (0.005 [95% CI, 0.002–0.009] mm, p=0.04 and week 44 (0.006 [95% CI, 0.004–0.009] mm, p=0.005, and from week 14 to week 28 (0.006 [95% CI, 0.002–0.010] mm, p=0.04	
				tibial trabecular number 4% site (HRpQCT) 1/mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.68±0.23 vs. 1.72±0.24 vs. 1.80±0.24 vs. 1.76±0.20 Ongoing user: 1.77±0.16 vs. 1.79±0.16 vs. 1.85±0.20 vs. 1.85± 0.17	p≥0.16	
				tibial trabecular spacing 4% site (HRpQCT) mm	Median (IQR) pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.554 (0.473, 0.593) vs. 0.544 (0.466, 0.591) vs. 0.520 (0.447, 0.567) vs. 0.509 (0.463, 0.560) Ongoing user: 0.534 (0.474, 0.546) vs. 0.524 (0.483, 0.550) vs. 0.502 (0.452, 0.543) vs. 0.511 (0.463, 0.522)	p≥0.16	
				tibial cortical porosity 4% site (HRpQCT) %	Median (IQR) pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.1 (0.9, 1.5) vs. 1.0 (1.0, 1.6) vs. 1.0 (0.8, 1.4) vs. 1.0 (1.0, 1.6) Ongoing user: 1.0 (0.7, 1.2) vs. 1.1 (0.7, 1.3) vs. 0.9 (0.5, 1.3) vs. 1.1 (0.6, 1.4)	p≥0.70	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				tibial cortical pore diameter 4% site (HRpQCT) mm	Median (IQR) pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.173 (0.163, 0.182) vs. 0.177 (0.165, 0.185) vs. 0.168 (0.161, 0.176) vs. 0.167 (0.166, 0.185) Ongoing user: 0.179 (0.168, 0.189) vs. 0.177 (0.169, 0.190) vs. 0.168 (0.158, 0.185) vs. 0.176 (0.164, 0.187)	p ≥ .161, training did not change cortical pore diameter size in any contraceptive group but was higher in nonusers compared with COCP users at week 1, and higher in nonusers than COCP users at week 28 p≤0.024	
				total vBMD10 30% site (HRpQCT) mg HA/cm^3	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 786±42 vs. 780±43 vs. 787±35 vs. 789±41 Ongoing user: 779±46 vs. 778±49 vs. 784±45 vs. 783±47	p≥0.30	
				cortical vBMD10 30% site (HRpQCT) mg HA/cm^3	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1016±21 vs. 1014±20 vs. 1019±19 vs. 1025±19 Ongoing user: 1012±16 vs. 1009±17 vs. 1016±20 vs. 1019±27	p≥0.30	
				tibial cortical area 30% site (HRpQCT) mm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 252±35 vs. 250±32 vs. 253±37 vs. 253±38 Ongoing user: 246±31 vs. 247±20 vs. 248±31 vs. 247±31	p≥0.19	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				tibial cortical thickness 30% site (HRpQCT) mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 5.67±0.58 vs. 5.64±0.54 vs. 5.69±0.63 vs. 5.69±0.63 Ongoing user: 5.58±0.45 vs. 5.62±0.46 vs. 5.60±0.45 vs. 5.60± 0.45	p≥0.19	
				tibial cortical perimeter 30% site (HRpQCT) mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 73.3±5.1 vs. 72.5±4.9 vs. 73.9±5.3 vs. 73.8 ± 5.2 Ongoing user: 72.1±4.2 vs. 72.3 ± 4.2 vs. 72.6±4.3 vs. 72.5±3.9	p≥0.19	
				tibial cortical porosity 30% site (HRpQCT) %	Median (IQR) pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.7 (0.6, 1.1) vs. 0.6 (0.4, 0.9) vs. 0.6 (0.5, 1.1) vs. 0.6 (0.4, 1.0) Ongoing user: 0.7 (0.4, 0.9) vs. 0.7 (0.5, 0.9) vs. 0.6 (0.4, 0.8) vs. 0.7 (0.3, 0.9)	p≤ 0.05	
				tibial cortical pore diameter 30% site (HRpQCT) mm	Median (IQR) Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.270 (0.245, 0.361) vs. 0.223 (0.201, 0.280) vs. 0.243 (0.220, 0.321) vs. 0.228 (0.205, 0.256) Ongoing user: 0.223 (0.210, 0.235) vs. 0.218 (0.179, 0.244) vs. 0.208 (0.190, 0.216) vs. 0.208 (0.180, 0.229)	p≤ 0.05	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				tibial failure load under uniaxial compression 4% site (HRpQCT) kN	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 10.3±2.0 vs. 10.4±1.8 vs. 9.9±1.8 vs. 10.3±1.9 Ongoing user: 8.9±2.2 vs. 9.1±1.9 vs. 9.1±2.2 vs. 9.2±1.6	p≥0.17	
				tibial stiffness 4% site (HRpQCT) kN/mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 191±36 vs. 191±36 vs. 182±35 vs. 190±39 Ongoing user: 163±42 vs. 166±37 vs. 159±32 vs. 169±32	p≥0.17	
				tibial failure load under uniaxial compression 30% site (HRpQCT) kN	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 14.9±2.1 vs. 15.1±1.7 vs. 15.4±2.2 vs. 15.4±2.2 Ongoing user: 14.6±1.7 vs. 14.7±1.7 vs. 14.9±1.6 vs. 14.9±1.7	p≥0.17	
				tibial stiffness 30% site (HRpQCT) kN/mm	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 266±39 vs. 267±38 vs. 273±41 vs. 274±40 Ongoing user: 259±33 vs. 261±32 vs. 263±32 vs. 258±36	p≥0.17	
				aBMD arms (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.87±0.11 vs. 0.89±0.11 vs. 0.92±0.08 vs. 0.85±0.12 Ongoing user: 0.88±0.09 vs. 0.88±0.10 vs. 0.84±0.12 vs. 0.78±0.13	p≥0.11	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				aBMD legs (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.24±0.10 vs. 1.24±0.08 vs. 1.23±0.07 vs. 1.24±0.07 Ongoing user: 1.25±0.08 vs. 1.25±0.10 vs. 1.24±0.08 vs. 1.24±0.09	p≥0.11	
				aBMD trunk (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.02±0.06 vs. 1.02±0.06 vs. 1.02±0.06 vs. 1.02±0.06 Ongoing user: 1.02±0.10 vs. 1.02±0.10 vs. 1.02±0.10 vs. 1.02±0.10	p≥0.11	
				aBMD ribs (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 0.87±0.06 vs. 0.86±0.06 vs. 0.88±0.05 vs. 0.87±0.05 Ongoing user: 0.87±0.07 vs. 0.86±0.08 vs. 0.86±0.08 vs. 0.87±0.08	p≥0.11	
				aBMD pelvis (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.08±0.07 vs. 1.10±0.07 vs. 1.09±0.07 vs. 1.10±0.07 Ongoing user: 1.11±0.14 vs. 1.11±0.13 vs. 1.12±0.13 vs. 1.11±0.12	p≥0.11	
				aBMD spine (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.11±0.09 vs. 1.13±0.09 vs. 1.12±0.08 vs. 1.12±0.10 Ongoing user: 1.09±0.11 vs. 1.10±0.10 vs. 1.10±0.09 vs. 1.08±0.11	p≥0.11	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				aBMD whole body (DXA) g/cm2	Mean pre-post (week 1 vs. week 14 vs. week 28 vs. week 44) Never user: 1.21±0.08 vs. 1.21±0.07 vs. 1.21±0.06 vs. 1.20±0.07 Ongoing user: 1.22±0.09 vs. 1.22±0.10 vs. 1.21±0.10 vs. 1.19±0.10	p≥0.11	
				Bone-specific alkaline phosphatase ALP (serum biomarker) µg/L-1	Median (IQR) pre-post (week 1 vs. week 28 vs. week 44) Never user: 19.1 (17.7, 21.7) vs. 20.4 (16.4, 24.2) vs. 21.0 (15.9, 26.4) Ongoing user: 18.1 (15.6, 18.7) vs. 18.4 (17.6, 22.2) vs. 20.1 (16.7, 24.4)	p≥0.05	
				Sclerostin (serum biomarker) pmol/L-1	Median (IQR) pre-post (week 1 vs. week 28 vs. week 44) Never user: 36.7 (31.5, 39.6) vs. 35.0 (32.1, 43.4) vs. 36.9 (29.0, 45.4) Ongoing user: 33.0 (28.9, 40.6) vs. 36.9 (31.3, 47.9) vs. 30.8 (27.8, 41.7)	p≥0.05	
				P1NP (plasma) µg/L-1	Median (IQR) pre-post (week 1 vs. week 28 vs. week 44) Never user: 68.2 (58.1, 84.9) vs. 84.4 (63.7, 105.1) vs. 73.7 (64.6, 80.3) Ongoing user: 61.3 (50.5, 77.5) vs. 65.9 (54.6, 93.5) vs. 67.7 (57.1, 79.4)	p< 0.05 contraception × time interaction P1NP was higher in progestin only contraceptive users than CHC users at week 1 p=0.01, d (1.022) No interaction for CHC vs nonusers of contraception	
				β-CTX (plasma) µg/L-1	Median (IQR) pre-post (week 1 vs. week 28 vs. week 44) Never user: 0.55 (0.42, 0.59) vs. 0.53 (0.36, 0.60) vs. 0.55 (0.44, 0.66) Ongoing user: 0.49 (0.38, 0.59) vs. 0.43 (0.33, 0.60) vs. 0.49 (0.40, 0.59)	p≥.053	

Author, year (design, country)	**Participants** n (Mean±SD, Median (min-max), or Mean (95%CI) age years)	**CHC Intervention** method; compound (dose µg) Duration (prior use)	**Comparison Condition** (prior CHC exposure)	**Outcome** (method, unit)	**Group Results**	**Between Group Comparison**	**DB Score (0-32)**
				Phosphate (serum biomarker) nmol/L-1	Mean pre-post (week 1 vs. week 28 vs. week 44) Never user: 1.59±0.18 vs. 1.62±0.17 vs. 1.63±0.16 Ongoing user: 1.56±0.10 vs. 1.53±0.23 vs. 1.55±0.14	p≥0.05	
				Albumin-adjusted calcium (serum biomarker) nmol/L-1	Mean pre-post (week 1 vs. week 28 vs. week 44) Never user: 2.48±0.12 vs. 2.50±0.07 vs. 2.57±0.12 Ongoing user: 2.48±0.10 vs. 2.55±0.09 vs. 2.53±0.11a	p≥0.05	
				Total 25(OH)D (serum biomarker) nmol/L-1	Mean pre-post (week 1 vs. week 28 vs. week 44) Never user: 57.0±16.7 vs. 69.7±20.8 vs. 53.9±14.8 Ongoing user: 77.9±31.0 vs. 79.4±24.9 vs. 70.5±19.8	p≥0.05	

*Mean and standard error

aBMD (areal bone mineral density), ALP (alkaline phosphatase), BAP (Bone Alkaline Phosphotase), BGP (Osteocalcin), BMC (bone mineral content), BMD (bone mineral density), CA (cyproterone acetate), CHC+ (CHC user), CHC- (CHC nonuser), CTX (C-terminal peptide), DEN (Denmark), DG (desogestrel), DGn (desogestren), DP (drospirenone), D-PYD (Deoxypyridinoline), EE2 (ethinyl E2), ET (etonogestrel), FSR (fractional synthesis rate), GD (gestoden/gestodene), GER (Germany), HRpQCT (high-resolution peripheral quantitative computed tomography), LNG (levonorgestrel), MRI (Magnetic Resonance Imaging), NG (norgestrel), NGMN (norelgestromin, NO (norethindrone), NOR (Norway), NR (Not reported), PG (progesterone), PYD (Pyridinoline), RCT (randomized controlled trial), UK (United Kingdom), USA (United States of America), vBMD (volumetric bone mineral density

4. Downs and Black Quality Assessment Tool Ratings

	Reporting										*External validity*			*Internal validity – Bias*							*Internal validity –confounding*						*Power*	*Total*
Study Year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**(0-32)**
Hansen 1991	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	10
Mazess 1991	0	1	1	0	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	8
Cooper 1993	1	1	0	0	0	1	1	0	1	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	5	18
Vessey 1998	0	1	0	0	0	1	1	0	0	1	1	1	0	0	0	0	1	0	0	0	1	1	0	0	0	1	0	10
Vessey 1999	0	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	10
Weaver 2001	1	0	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	1	0	12
Cobb 2002	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	0	16
Elgan 2003	1	1	1	0	2	1	1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	0	0	1	0	0	14
Nappi 2003	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	1	0	1	0	0	1	1	1	0	0	1	0	16
Reed 2003	1	1	1	0	1	1	1	0	0	1	1	0	0	0	0	1	1	0	0	1	1	1	0	0	1	1	1	16
Berenson 2004	1	1	1	0	1	1	1	0	0	1	0	0	1	1	0	1	1	1	0	1	1	1	1	0	1	1	0	18
Elgan 2004	1	1	1	0	2	1	1	0	0	1	0	0	0	0	0	1	0	1	0	0	1	1	0	0	1	1	0	14
Rome 2004	1	1	1	0	1	1	0	0	0	0	1	1	0	0	0	1	1	1	0	1	1	0	0	0	1	0	0	13
Barad 2005	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	1	1	1	0	0	1	1	3	19
Massai 2005	1	1	1	1	0	1	1	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	1	1	14
Nappi 2005	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	0	0	4	24
Hartard 2006	1	1	1	1	2	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	12
Cobb 2007	1	1	1	1	2	1	1	1	1	0	0	0	1	0	0	0	1	1	1	1	1	1	1	0	1	1	1	21
Kelsey 2007	1	1	1	0	2	1	1	0	1	1	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	15
Berenson 2008	1	1	1	1	2	0	0	0	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	1	1	1	20
Cromer 2008	1	1	1	0	2	1	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	0	17
Gargano 2008	1	1	1	1	1	1	1	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	13
Liederbach 2008	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	0	1	0	14
Procter-Gray 2008	1	1	1	1	2	1	1	1	0	1	1	1	0	0	1	0	1	1	1	1	1	1	1	0	1	1	0	22
Beksinska 2009	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	1	1	1	17
Bonny 2009	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	1	1	1	0	1	1	0	0	1	0	0	14
Liu 2009	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	1	1	1	18
Massaro 2010	1	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	0	1	0	0	1	1	1	0	0	0	0	17
Liu 2011	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	1	0	1	1	0	1	0	1	0	0	1	0	17
Scholes 2011	1	1	1	0	2	1	1	0	0	0	1	1	0	0	0	1	1	1	0	1	1	1	0	0	1	1	0	17
Gai 2012	1	1	1	1	2	1	1	1	0	1	0	0	1	0	0	0	0	1	0	0	1	0	1	0	0	1	0	15
Biason 2015	1	1	0	0	1	1	1	0	0	1	0	0	1	0	1	1	1	0	0	1	1	1	0	0	0	1	1	15
Lee 2015	1	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	1	0	12
Gersten 2016	1	1	1	1	2	1	1	1	0	1	0	0	1	0	1	0	1	1	1	0	1	1	1	0	1	1	3	23
Jackowski 2016	1	1	1	0	2	1	1	0	0	0	1	0	0	0	0	0	1	1	0	1	1	1	0	0	1	1	0	15
Reiger 2016	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0	0	12
Hellevik 2017	1	1	1	0	2	1	1	0	0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	1	1	5	18
Brajic 2018	1	1	1	0	2	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	1	1	0	0	1	1	0	18
Dalgaard 2019	1	1	1	0	1	1	1	0	1	1	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	1	0	15
Leung 2019	0	1	1	0	1	1	1	0	0	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	1	1	5	18
Allaway 2020	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	1	0	15
Almstedt 2020	1	1	1	0	2	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	0	1	0	0	13
Dalgaard 2020	1	1	1	0	2	1	1	0	1	1	0	1	0	0	0	0	1	1	0	1	1	1	0	0	0	1	1	16
Herzog 2020	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	0	1	1	0	1	1	1	0	0	1	1	5	19
Oxfeldt 2020	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	1	1	0	1	1	0	0	0	0	0	0	10
Martin 2021	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	8
O'Leary 2021	0	1	1	0	1	1	1	0	1	0	0	1	0	0	0	0	1	1	0	1	1	1	0	0	0	0	3	15
Yoo 2021	1	1	1	0	2	1	1	0	0	1	1	1	0	0	0	1	1	1	0	1	1	1	0	0	1	1	5	23
He 2022	1	1	1	0	2	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	0	0	0	1	0	12
Sung 2022	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	9
Proportion fully meeting (%)[1]	90.0	98.0	82.0	30.0	38.0	94.0	94.0	22.0	20.0	72.0	34.0	34.0	24.0	2.0	14.0	60.0	72.0	84.0	18.0	72.0	84.0	68.0	24.0	4.0	52.0	74.0	34.0*	

[1]For item 5, score of 2 is fully meeting, all others 1 is fully meeting, *proportion of scores ≥1, blue columns are applicable to intervention studies only.

5.Semi-quantitative Analyses

Outcome	Author, Year	Age Range	Follow Up	Sample Size	Comparison	Between Group Estimate	Study Design	Study Limitations[1]	Inconsistency[2]	Indirectness	Imprecision	Modified GRADE Rating
Lumbar Spine BMD	Berenson 2008	16-33	36-months	128	new vs non-user	p<0.001	PC	Not serious	Consistent	Indirect	unclear	Very Low
	Brajic 2018	16-24	12-months	307	ongoing vs non-user	0.002 (-0.104, 0.091)	PC	Serious			Precise	
	Cobb 2002	18-30	?	476	past vs non-user	-0.000005 ± 0.0002	RC	Not Serious			Precise	
	Cobb 2007	18-26	24-months	150	new vs non-user	0.0020±0.0025	RCT	Very Serious			Precise	
	Hartard 2006	18-24	12-months	59	ongoing vs non-user	CHC A: d=-0.85 (-1.51, -0.19) CHC B: d=-0.20 (-0.84, 0.45)	Quasi	Very Serious			Precise	
	Jackowski 2016	8-33	20-years	110	ongoing vs non-user	NS	RC	Serious			unclear	
	Massai 2005	18-35	24-months	107	ongoing vs non-user	−0.341 (−0.473, −0.208)	Quasi	Very Serious			precise	
	Mazess 1991	20-39	12-months	300	ongoing vs non-user	d= 0.08 (-0.25, 0.40)	PC	Very Serious			Precise	
	Nappi 2003	22-34	12-months	56	new vs non-user	NS	Quasi	Very Serious			unclear	
	Reed 2003	18-39	36-months	178	ongoing vs non-user	p=0.73	PC	Not Serious			unclear	
	Rome 2004	12-18	12-months	317	new vs non-user	NR	PC	Very Serious			unclear	
	Scholes 2011	14-30	36-months	139	ongoing vs non-user	NS	PC	Not serious			unclear	
	Weaver 2001	18-31	?	141	ongoing vs non-user	NS	RCT	Very Serious			unclear	
	Hansen 1991	post menopausal	12-years	121	previous vs non-user	NS	PC	Very Serious			unclear	
	Berenson 2004	18-33	24-months	111	new vs non-user	CHC A: 0.67 (-1.54, 2.88) CHC B: 1.51 (-0.40, 3.42)	Quasi	Very Serious			unclear imprecise	
	Biason 2015	12-19	12-months	61	new vs non-user	0.101	Quasi	Very Serious			unclear	
	Cromer 2008	12-18	24-months	157	new vs non-user	NR	PC	Not Serious			unclear	
	Gai 2012	16-18	24-months	376	new vs non-user	CHC A: d= -0.09 (-0.33, 0.14) CHC B: d= -0.07 (-0.30, 0.17)	Quasi	Very Serious			Precise	
	Gargano 2008	21-34	12-months	61	new vs non-user	CHC A: d= -0.01 (-0.62, 0.59) CHC B: d= -0.02 (-0.61, 0.58)	Quasi	Very Serious			Imprecise	
	Gersten 2016	12-18	12-months	859	new vs non-user	CHC A: 0.23 (-0.20, 0.67) CHC B: 1.05 (0.61, 1.49)	Quasi	Very Serious			precise	
	Liu 2011	25-40	24-months	154	new vs non-user	CHC A: d=-0.02 (-0.41, 0.38) CHC B: d=0.02 (-0.36, 0.40)	Quasi	Very Serious			precise	
	Massaro 2010	23-34	12-months	49	new vs non-user	NS	Quasi	Very Serious			unclear	

	Nappi 2005	22-34	12-months	67	new vs non-user	NS	Quasi	Very Serious			unclear		
	Overall			**4484**			**Mod Quality +3**	**Very Serious -2**	**0**	**-1**	**-1**	**-2**	
Femoral Neck BMD	Berenson 2008	16-33	36-months	128	new vs non-user	p<0.001	PC	Not serious			unclear		
	Brajic 2018	16-24	24-months	307	ongoing vs non-user	-0.001 (-0.010, 0.008)	PC	Serious			precise		
	Cromer 2008	12-18	24-months	157	new vs non-user	NR	PC	Not Serious			unclear		
	Gai 2012	16-18	24-months	376	new vs non-user	CHC A: d=-0.12 (-0.38, 0.13) CHC B: d=-0.02 (-0.27, 0.23)	Quasi	Very Serious			Precise		
	Gersten 2016	12-18	12-months	859	new vs non-user	CHC A: -0.65 (-1.05, -0.25) CHC B: -0.32 (-0.09, 0.72)	Quasi	Very Serious			precise		
	Hansen 1991	post menopausal	12-years	121	previous vs non-user	NS	PC	Very Serious			unclear		
	Hartard 2006	18-24	12-months	59	ongoing vs non-user	CHC A: d= 0.10 (-0.53, 0.74) CHC B: d= 0.12 (-0.53, 0.76)	Quasi	Very Serious	Consiste nt	Indirect	imprecise	Very Low	
	Jackowski 2016	8-33	20-years	110	ongoing vs non-user	NS	RC	Serious			unclear		
	Liu 2011	25-40	24-months	154	new vs non-user	CHC A: d=-0.15 (-0.55, 0.24) CHC B: d= -0.01 (-0.39, 0.37)	Quasi	Very Serious			precise		
	Massai 2005	18-35	24-months	107	ongoing vs non-user	-0.27 (-0.38, -0.15)	Quasi	Very Serious			precise		
	Mazess 1991	20-39	24-months	300	ongoing vs non-user	d=-0.19 (-0.57, 0.19)	PC	Very Serious			Precise		
	Reed 2003	18-39	36-months	178	ongoing vs non-user	p=0.55	PC	Not Serious			unclear		
	Rome 2004	12-18	12-months	317	new vs non-user	NR	PC	Very Serious			unclear		
	Weaver 2001	18-31	?	141	ongoing vs non-user	NS	RCT	Very Serious			unclear		
	Overall			**3314**			**High Quality +4**	**Very Serious -2**	**0**	**-1**	**-1**	**0**	
Whole Body BMD	Biason 2015	12-19	12-months	61	new vs non-user	0.0444	Quasi	Very Serious			Unclear		
	Cobb 2002	18-30	?	476	past vs non-user	Beta -0.000054 ± 0.00012	RC	Not Serious			Precise		
	Gersten 2016	12-18	12-months	859	new vs non-user	CHC A: 0.43 (0.03, 0.82) CHC B: 0.40 (0.01, 0.80)	Quasi	Very Serious	Unclear	Indirect	precise	Very Low	
	Jackowski 2016	8-33	20-years	110	ongoing vs non-user	NS	RC	Serious			Unclear		
	Reed 2003	18-39	36-months	178	ongoing vs non-user	p=0.96	PC	Not Serious			Unclear		
	Scholes 2011	14-30	36-months	139	ongoing vs non-user	NS	PC	Not serious			Unclear		
	Almstedt 2020	18-20	12-months	62	ongoing vs non-user	NR	PC	Very Serious			Unclear		

	O'Leary 2021	19-30	44-weeks	29	ongoing vs non-user	p≥0.11	PC	Very Serious			Unclear		
	Overall			**1914**			**Mod Quality +3**	**Very Serious -2**	**-1**	**-1**	**-1**	**-2**	
D-PYD	Elgan 2003	18-26	24-months	118	new vs. ongoing user	- 1.5 ± 0.49	RC	Very Serious	Unclear	Direct	Unclear	Very Low	
	Nappi 2003	22-34	12-months	56	new vs. non-user	p<0.05	Quasi	Very Serious					
	Rome 2004	12-18	12-months	317	new vs. non-user	p=0.08	PC	Very Serious					
	Nappi 2005	22-34	12-months	67	new vs. non-user	p<0.05	Quasi	Very Serious					
	Gargano 2008	21-34	12-months	61	new vs. non-user	NR	Quasi	Very Serious					
	Massaro 2010	23-34	12-months	49	new vs. non-user	NS	Quasi	Very Serious					
	Overall			**668**			**Mod Quality +3**	**Very Serious -2**	**-1**	**0**	**-1**	**-1**	
CTX	Reiger 2016	28-25	Baseline	23	ongoing vs. non-user	NS	PC	Very Serious	Consistent	Indirect	Unclear	Very Low	
	Almstedt 2020	18-20	12-months	62	ongoing vs. non-user	NS	PC	Very Serious					
	Martin 2021	?	One pill cycle	28	ongoing vs. non-user	p=0.37	PC	Very Serious					
	O'Leary 2021	19-30	44-weeks	29	ongoing vs. non-user	p≥0.13	PC	Very Serious					
	He 2022	young	One pill cycle	38	ongoing vs. previous user	p<0.01	PC	Serious					
	Overall			**180**			**Mod Quality +3**	**Very Serious -2**	**0**	**-1**	**-1**	**-1**	
LBM	Procter-Gray 2008	18-26	26.6-months	101	new vs non-user	0.77 ± 0.17	RCT	Very Serious	Unclear	Indirect	Unclear	Very Low	
	Almstedt 2020	18-20	12-months	62	ongoing vs non-user	NR	PC	Very Serious					
	Bonny 2009	12-18	6-months	36	new vs non-user	p=0.07	Quasi	Very Serious					
	Cobb 2002	18-30	?	476	past vs non-user	NR	RC	Not Serious					
	Dalgaard 2020	18-30	10-weeks	38	ongoing vs non-user	d= -0.1 (-0.74, 0.54)	PC	Very Serious					
	Overall			**713**			**High Quality +4**	**Very Serious -2**	**-1**	**-1**	**-1**	**-1**	
Lumbar Spine BMC	Biason 2015	12-19	12-months	61	new vs non-user	0.153	Quasi	Very Serious	Inconsistent	Indirect	Unclear	Very Low	
	Gersten 2016	12-18	12-months	859	new vs non-user	CHC A: 0.27 (-0.33, 0.87) CHC B: 1.45 (0.85, 2.06)	Quasi	Very Serious			Precise		
	Hartard 2006	18-24	12-months	59	ongoing vs non-user	CHC A: d= -0.65 (-1.30, -0.01) CHC B: d= -0.40 (-1.05, 0.25)	Quasi	Very Serious			Imprecise		
	Jackowski 2016	8-33	20-years	110	ongoing vs non-user	NS	RC	Serious			Unclear		
	Weaver 2001	18-31	?	141	ongoing vs non-user	NS	RCT	Very Serious			Unclear		

	Overall			**1230**			**Mod Quality +3**	**Very Serious -2**	**-1**	**-1**	**-1**	**-2**	
Whole Body BMC	Biason 2015	12-19	12-months	61	new vs non-user	0.1012	Quasi	Very Serious	Unclear	Direct	Unclear	Very Low	
	Cobb 2007	18-26	24-months	150	new vs non-user	6.2±5.2	RCT	Very Serious			Precise		
	Gersten 2016	12-18	12-months	859	new vs non-user	CHC A: 0.53 (-0.43, 1.48) CHC B: 1.01 (0.05, 1.96)	Quasi	Very Serious			Precise		
	Jackowski 2016	8-33	20-years	110	ongoing vs non-user	NS	RC	Serious			Unclear		
	Weaver 2001	18-31	?	141	ongoing vs non-user	NS	RCT	Very Serious			Unclear		
	Overall			**1321**			**High Quality +4**	**Very Serious -2**	**-1**	**0**	**-1**	**0**	
Any Fracture	Cooper 1993	29	?	NR	ongoing vs. non-user	aRR: 1.20 (1.08,1.34)	PC	Not serious	Consistent	direct	precise	Low	
	Vessey 1998	25-39	26 years	310000 person years	ongoing vs. non-user	aRR: 1.2 (1.1, 1.4)	RC	Very Serious			precise		
	Barad 2005	50-79	~ 2.5 years	80947	ongoing vs. non-user	aHR: 1.02 (.91,1.14)	PC	Serious			precise		
	Yoo 2021	?	99.6 (96-103.2)months	1272115	ongoing vs. non-user	aHR: 1.03 (1.01-1.05)	RC	Not serious			precise		
	Overall			**1663062**			**Mod Quality +3**	**Serious -1**	**0**	**0**	**0**	**+2**	
P1NP	Allaway 2020	18-30	~87-days	24	User vs. non-user	NR	Quasi	Very Serious	Consistent	Indirect	Unclear	Very Low	
	Martin 2021	18-35	One pill cycle	28	ongoing vs. non-user	p=0.81	PC	Very Serious					
	O'Leary 2021	19-30	44-weeks	29	ongoing vs. non-user	p=0.10	PC	Very Serious					
	He 2022	young	One pill cycle	38	ongoing vs. previous user	p=0.11	PC	Very Serious					
	Overall			**119**			**Mod Quality +3**	**Very Serious -2**	**0**	**-1**	**-1**	**-1**	
PYD	Nappi 2003	22-34	12-months	56	new vs. non-user	p<0.05	Quasi	Very Serious	Consistent	direct	Unclear	Very Low	
	Nappi 2005	22-34	12-months	67	new vs. non-user	p<0.05	Quasi	Very Serious					
	Gargano 2008	21-34	12-months	61	new vs. non-user	NR	Quasi	Very Serious					
	Massaro 2010	23-34	12-months	49	new vs. non-user	p<0.05	Quasi	Very Serious					
	Overall			**233**			**Low Quality +2**	**Very Serious -2**	**0**	**0**	**-1**	**-1**	
BAP	O'Leary 2021	19-30	44-weeks	29	ongoing vs non-user	p>0.05	PC	Very Serious	Unclear	Indirect	Unclear	Very Low	
	Martin 2021	18-35	One pill cycle	28	ongoing vs non-user	p=0.47	PC	Very Serious					
	Reiger 2016	28-35	Baseline	23	ongoing vs non-user	NR	PC	Very Serious					
	Rome 2004	12-18	12-months	317	new vs non-user	p=0.004	PC	Very Serious					
	Overall			**397**			**Mod Quality +3**	**Very Serious -2**	**-1**	**-1**	**-1**	**-2**	
BGP	Nappi 2003	22-34	12-months	56	new vs. non-user	NS	Quasi	Very Serious	Unclear	Direct	Unclear	Very Low	

	Nappi 2005	22-34	12-months	67	new vs. non-user	NR	Quasi	Very Serious					
	Gargano 2008	21-34	12-months	61	new vs. non-user	NR	Quasi	Very Serious					
	Massaro 2010	23-34	12-months	49	new vs. non-user	p<0.05	Quasi	Very Serious					
	Overall			**233**			**Low Quality +2**	**Very Serious -2**	**-1**	**0**	**-1**	**-2**
Radius BMD	Beksinska 2009	15-19	60-months	155	new vs non-user	p=0.01	PC	Not serious	Consistent	Direct	Unclear	Very Low
	Hartard 2006	18-24	12-months	59	ongoing vs non-user	CHC A: d=0.17 (-0.47, 0.80) CHC B: d= -0.3 (-0.96, 0.36)	Quasi	Very Serious			Imprecise	
	Mazess 1991	20-39	24-months	300	ongoing vs non-user	d= 0.0 (-0.33, 0.33)	PC	Very Serious			Precise	
	Weaver 2001	18-31	?	141	ongoing vs non-user	NS	RCT	Very Serious			Unclear	
	Overall			**655**			**High quality +4**	**Very Serious -2**	**0**	**0**	**-1**	**1**
Total Hip BMD	Brajic 2018	16-24	24-months	307	ongoing vs non-user	-0.001 (-0.009, 0.006)	PC	Serious	Inconsistent	Indirect	Imprecise	Very Low
	Cobb 2002	18-30	7-years	476	past vs non-user	Beta -0.000012 ± 0.0002	RC	Not Serious			Imprecise	
	Scholes 2011	14-30	36-months	139	ongoing vs non-user	NS	PC	Not serious			Unclear	
	Cobb 2007	18-26	24-months	150	new vs non-user	0.0035±0.0022	RCT	Very Serious			Precise	
	Overall			**1072**			**High Quality +4**	**Serious -1**	**-1**	**-1**	**-1**	**0**
Total Knee Arthroplasty	Liu 2009	Middle aged	6.1 person-years	1291767		aRR: 1.00 (0.96, 1.04)	PC	Not serious	Consistent	direct	precise	Low
	Hellevik 2017	≥20	8.3 ± 4.5 years	18126	Previous user vs. non-user	aHR: 1.36 (1.00, 1.86)	PC	Serious			precise	
	Leung 2019	45-74	14.8 years	35185		aHR: 1.18 (1.05, 1.32)	PC	Serious			precise	
	Overall			**1345078**			**Mod Quality +3**	**Serious -1**	**0**	**0**	**0**	**(+2)**

aHR (adjusted hazard ratio), aRR (adjusted risk ratio), BAP (bone alkaline phosphatase), BGP (Osteocalcin), BMC (bone mineral content), BMD (bone mineral density), CHC (combined hormonal contraceptive), CTX (C-terminal peptide), d (Cohen's d effect size), D-PYD (Deoxypyridinoline), GRADE (Grading of Recommendations Assessment, Development and Evaluation), LBM (lean body mass), Mod (moderate), NR (not reported), NS (not significant, only used when that is what was reported) P1NP (Procollagen type 1 terminal peptide), PC (prospective cohort), PYD (Pyridinoline), TKA (total knee arthroplasty), Quasi (Quasi experimental study), RC (retrospective cohort), RCT (randomized controlled trial)

[1]not serious (≥12/13), serious (11/13), Very serious (≤10/13) based on questions 14 to 26 on the Downs and Black Tool

[2]Consistency based on overlap of the 95%CI for similar statistics, approximately two-thirds need to overlap to be consistent.