

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 resorption 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 displac* 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 peri-arthritis/ 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 peri-arthritis 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 Tool²¹

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	NR	15
				P1NP (serum) ng/mL	During Intervention (50-87 days) Never user: NR CHC Oral: NR CHC Ring: NR Mean pre-post Baseline: Never user: 7.34 ± 2.15 CHC Oral: 11.93 ± 3.27 CHC Ring: 13.38 ± 4.97	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: Oral EE/NO (0.035mg/1mg) 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)	NR	20
	CHC A: n (25 (26.1±3.9)) CHC B: n (42 (25.4±4.4))	CHC B: Oral EE/DG (0.030mg/ 0.15mg) 24 months (no use within 1 mo)			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)	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)	
Berenson, 2008 (PC, USA)	Never user: n (51 (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	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	6 mo p<.001 12 mo p<.001 18 mo p<.001 24 mo p<.001 30 mo p<.001 36 mo p<.001	20
	New user: n (77 (16-33))			FEMORAL NECK BMD (DXA) g/cm ²	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>.05 12 mo p<.05 18 mo p<.001 24 mo p<.001 30 mo p<.001 36 mo p<.001	
Biaison, 2015 (Quasi, Brazil)	Never user: n (26 (15.6; 14.7-16.1))	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
	New user: n (35 (15.8; 11.8-19.5))			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)
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)	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)	14
				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)	
				LBM (DXA) kg	Mean % change (baseline to 6-mo) Never users: 0.6% ± 3.4% New users: 0.6% ± 4.7%	p=0.07	
				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)	
				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%	
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 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) -0.000054 ± 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	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 Eumenorrhoeic Never user: 3.7 ± 3.4 New user: 9.9 ± 3.9	Difference in mean yearly change rate (± SE) Eumenorrhoeic 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 Eumenorrhoeic Never user: 0.0002 ± 0.0016 New user: 0.0022 ± 0.0019	Difference in mean yearly change rate (± SE) Eumenorrhoeic 0.0020±0.0025	21
				TOTAL HIP BMD (DXA) g/cm ²	Yearly rate of change Eumenorrhoeic Never user: -0.0023 ± 0.0015 New user: 0.0013 ± 0.0017	Difference in mean yearly change rate (± SE) Eumenorrhoeic 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)
					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%		
				Fiber type composition (biopsy) %	Type Iia Never user: 39.4%±2.6% vs. 42.8%±2.1% Ongoing user: 42.6%±2.5% vs. 47.7%±2.5%	Group-by-time interaction Type I: p=0.52 Type Iia: p=0.64 Type Iix: p=0.05	
				tendon CSA (MRI) mm ²	Type Iix Never user: 7.1%±2.1% vs. 4.8%±1.2% Ongoing user: 10.5%±2.2% vs. 3.9%±1.5% 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 concentration 0.64/0.02 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		
				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	p=0.46	
				Quadriceps CSA (MRI) at 20 cm above lateral epicondyle, cm ²	Mean % change (baseline vs. 10 weeks) Never user 9.7±4.9% Ongoing user 10.6±4.8% 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	p=0.81	
				Quadriceps CSA (MRI) at 30 cm above lateral epicondyle, cm ²	Mean pre-post (baseline vs. 10 weeks) Never user: 9.2 ± 5.0% Ongoing user: 9.5 ± 6.0% 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	p=0.37	
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 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	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)
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±0.11 vs. 1.01±0.11	Baseline: p=0.99 12-mo: p=0.75 24-mo: p=0.34	15
				Femoral neck BMD (DXA) g/cm ²	Mean % change (baseline vs. 24-mo) Never user: 1.88% CHC A: -0.30% CHC B: 0.30% 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	Baseline: p=0.97 12-mo: p=0.93 24-mo: p=0.56	
					Mean % change (baseline vs. 24-mo) Never user: 0.98% CHC A: -0.61% CHC B: 0.49%		

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	
				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	12
				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)
				FEMORAL NECK BMD (DXA) g/cm ²	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	
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)	LUMBAR SPINE BMD (DXA) g/cm ²	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	17
				BGP (serum)	NR	p<0.05	
				PYD (urine)	NR	p<0.05	
				D-PYD (urine)	NR	NS	
				LUMBAR SPINE BMD (dual-photon absorptiometry) g/cm ²	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/cm ²	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/cm ²	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				
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 Irregular group: 0.02 ± 0.35, p (0.96) Regular group: 0.77 ± 0.17, p < 0.0001	22			
				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)
					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	
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)	BAP (serum) CTX (serum) LUMBAR SPINE BMD (DXA) g/cm ² FEMORAL NECK BMD (DXA) g/cm ²	(baseline) NR NR NR	NS NS NR NR	12
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) D-PYD (urine) nmol/mmol	Mean (12-mo, adjusted for baseline) Never user: 40.4±1.03 New user: 35.7±1.03 Mean (12-mo, adjusted for baseline) Never user: 9.8 ±1.03 New user: 9.0 ±1.03	p=0.004 p=0.08	13
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	
				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)	
Vessey ,1999 (RC, UK)	n (NR (25-39)	Oral; estrogen (≥50µg) N/A (5 to ≥97)	Never Users (0 days)	Displaced cervicval 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)	10
				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)
						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)	
				Ankle Fracture (medical record/referrals)	NR		
				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)
Martin, 2021 (PC, UK)	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	8
				PINP (serum biomarker) ng·mL ⁻¹	Mean values across menstrual cycle/CHC cycle Never user: 64.9±21.9 Ongoing user: 62.9±22.1	p=0.81	
	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·L ⁻¹	Mean values across menstrual cycle/CHC cycle Never user: 560±180 Ongoing user: 500±200	p=0.37	
				Bone ALP (serum biomarker) U·L ⁻¹	Mean values across menstrual 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	
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)	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	
				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	
				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	10

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) µm ² /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) cm ²	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)
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)	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	23
				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)	
				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) mm ²	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) mm ²	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) mm ²	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/cm ²	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/cm ²	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/cm ²	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/cm ²	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/cm ²	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/cm ²	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/cm ²	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 Phosphatase), 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 (Deoxy pyridinoline), 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

Study Year	Reporting										External validity				Internal validity – Bias					Internal validity –confounding					Power	Total (0-32)		
	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
Hansen 1991	1	1	1	0	1	1	1	0	0	1	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	8	
Cooper 1993	1	1	0	0	0	1	1	0	1	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	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	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	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	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	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	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	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	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	1	0	1	0	0	1	1	0	0	1	1	0	14	
Rome 2004	1	1	1	0	1	1	0	0	0	1	1	1	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	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	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	0	0	4	24	
Hartard 2006	1	1	1	1	2	1	1	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	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	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	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	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	1	0	0	1	0	0	0	0	0	0	1	1	1	0	0	13	
Liederbach 2008	1	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	1	0	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	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	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	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	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	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	0	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	0	0	1	1	0	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	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	1	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
Brajić 2018	1	1	1	0	2	1	1	0	1	1	1	1	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	0	1	1	1	0	0	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	0	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	1	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*	

¹For 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 [†]	Inconsistency [‡]	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			unclear	
	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	Consistent	Indirect	unclear	Very Low
	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						
	Overall			4484			Mod Quality +3	Very Serious -2	0	-1	-1	-2		
	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	
Femoral Neck BMD	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	Consistent	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	
		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	
Whole Body BMD	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
	Elgan 2003	18-26	24-months	118	new vs. ongoing user	- 1.5 ± 0.49	RC	Very Serious				
	Nappi 2003	22-34	12-months	56	new vs. non-user	p<0.05	Quasi	Very Serious				
D-PYD	Rome 2004	12-18	12-months	317	new vs. non-user	p=0.08	PC	Very Serious	Unclear	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	NS	Quasi	Very Serious				
	Overall			668			Mod Quality +3	Very Serious -2	-1	0	-1	-1
	Reiger 2016	28-25	Baseline	23	ongoing vs. non-user	NS	PC	Very Serious				
	Almstedt 2020	18-20	12-months	62	ongoing vs. non-user	NS	PC	Very Serious				
CTX	Martin 2021	?	One pill cycle	28	ongoing vs. non-user	p=0.37	PC	Very Serious	Consistent	Indirect	Unclear	Very Low
	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
	Procter-Gray 2008	18-26	26.6-months	101	new vs non-user	0.77 ± 0.17	RCT	Very Serious				
	Almstedt 2020	18-20	12-months	62	ongoing vs non-user	NR	PC	Very Serious				
LBM	Bonny 2009	12-18	6-months	36	new vs non-user	p=0.07	Quasi	Very Serious	Unclear	Indirect	Unclear	Very Low
	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
	Biason 2015	12-19	12-months	61	new vs non-user	0.153	Quasi	Very Serious				Unclear
	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
Lumbar Spine BMC	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	Inconsistent	Indirect	Imprecise	Very Low
	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	
	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	Unclear	Direct	Precise	Very Low	
	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				precise	
	Vessey 1998	25-39	26 years	310000 person years	ongoing vs. non-user	aRR: 1.2 (1.1, 1.4)	RC	Very Serious	Consistent	direct	precise	Low	
	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					
	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	Consistent	Indirect	Unclear	Very Low	
	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					
	Nappi 2005	22-34	12-months	67	new vs. non-user	p<0.05	Quasi	Very Serious	Consistent	direct	Unclear	Very Low	
	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					
	Martin 2021	18-35	One pill cycle	28	ongoing vs non-user	p=0.47	PC	Very Serious	Unclear	Indirect	Unclear	Very Low	
	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				Unclear
	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	Consistent	Direct		Imprecise Very Low
	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				Imprecise
	Cobb 2002	18-30	7-years	476	past vs non-user	Beta -0.000012 ± 0.0002	RC	Not Serious	Inconsistent	Indirect		Imprecise Very Low
	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				precise
	Hellevik 2017	≥20	8.3 ± 4.5 years	18126	Previous user vs. non-user	aHR: 1.36 (1.00, 1.86)	PC	Serious	Consistent	direct		precise Low
	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)

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

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