RESEARCH

Falls prevention interventions for community-dwelling older adults: systematic review and meta-analysis of benefits, harms, and patient values and preferences

Jennifer Pillay[1*†], Lindsay A. Gaudet[1†], Sabrina Saba[1], Ben Vandermeer[1], Ashiqur Rahman Ashiq[1], Aireen Wingert[1] and Lisa Hartling[1]

Abstract

Background About 20–30% of older adults (≥ 65 years old) experience one or more falls each year, and falls are associated with substantial burden to the health care system, individuals, and families from resulting injuries, fractures, and reduced functioning and quality of life. Many interventions for preventing falls have been studied, and their effectiveness, factors relevant to their implementation, and patient preferences may determine which interventions to use in primary care. The aim of this set of reviews was to inform recommendations by the Canadian Task Force on Preventive Health Care (task force) on fall prevention interventions. We undertook three systematic reviews to address questions about the following: (i) the benefits and harms of interventions, (ii) how patients weigh the potential outcomes (outcome valuation), and (iii) patient preferences for different types of interventions, and their attributes, shown to offer benefit (intervention preferences).

Methods We searched four databases for benefits and harms (MEDLINE, Embase, AgeLine, CENTRAL, to August 25, 2023) and three for outcome valuation and intervention preferences (MEDLINE, PsycINFO, CINAHL, to June 9, 2023). For benefits and harms, we relied heavily on a previous review for studies published until 2016. We also searched trial registries, references of included studies, and recent reviews. Two reviewers independently screened studies.

The population of interest was community-dwelling adults ≥ 65 years old. We did not limit eligibility by participant fall history. The task force rated several outcomes, decided on their eligibility, and provided input on the effect thresholds to apply for each outcome (fallers, falls, injurious fallers, fractures, hip fractures, functional status, health-related quality of life, long-term care admissions, adverse effects, serious adverse effects). For benefits and harms, we included a broad range of non-pharmacological interventions relevant to primary care. Although usual care was the main comparator of interest, we included studies comparing interventions head-to-head and conducted a network meta-analysis (NMAs) for each outcome, enabling analysis of interventions lacking direct comparisons to usual care. For benefits and harms, we included randomized controlled trials with a minimum 3-month follow-up and reporting on one of our fall outcomes (fallers, falls, injurious fallers); for the other questions, we preferred quantitative

---

†Jennifer Pillay and Lindsay A. Gaudet contributed equally to this work.

*Correspondence:
Jennifer Pillay
jpillay@ualberta.ca
Full list of author information is available at the end of the article

data but considered qualitative findings to fill gaps in evidence. No date limits were applied for benefits and harms, whereas for outcome valuation and intervention preferences we included studies published in 2000 or later.

All data were extracted by one trained reviewer and verified for accuracy and completeness. For benefits and harms, we relied on the previous review team's risk-of-bias assessments for benefit outcomes, but otherwise, two reviewers independently assessed the risk of bias (within and across study). For the other questions, one reviewer verified another's assessments. Consensus was used, with adjudication by a lead author when necessary. A coding framework, modified from the ProFANE taxonomy, classified interventions and their attributes (e.g., supervision, delivery format, duration/intensity).

For benefit outcomes, we employed random-effects NMA using a frequentist approach and a consistency model. Transitivity and coherence were assessed using meta-regressions and global and local coherence tests, as well as through graphical display and descriptive data on the composition of the nodes with respect to major pre-planned effect modifiers. We assessed heterogeneity using prediction intervals. For intervention-related adverse effects, we pooled proportions except for vitamin D for which we considered data in the control groups and undertook random-effects pairwise meta-analysis using a relative risk (any adverse effects) or risk difference (serious adverse effects).

For outcome valuation, we pooled disutilities (representing the impact of a negative event, e.g. fall, on one's usual quality of life, with 0 = no impact and 1 = death and ~ 0.05 indicating important disutility) from the EQ-5D utility measurement using the inverse variance method and a random-effects model and explored heterogeneity. When studies only reported other data, we compared the findings with our main analysis. For intervention preferences, we used a coding schema identifying whether there were strong, clear, no, or variable preferences within, and then across, studies.

We assessed the certainty of evidence for each outcome using CINeMA for benefit outcomes and GRADE for all other outcomes.

Results A total of 290 studies were included across the reviews, with two studies included in multiple questions. For benefits and harms, we included 219 trials reporting on 167,864 participants and created 59 interventions (nodes). Transitivity and coherence were assessed as adequate. Across eight NMAs, the number of contributing trials ranged between 19 and 173, and the number of interventions ranged from 19 to 57. Approximately, half of the interventions in each network had at least low certainty for benefit. The fallers outcome had the highest number of interventions with moderate certainty for benefit (18/57). For the non-fall outcomes (fractures, hip fracture, long-term care [LTC] admission, functional status, health-related quality of life), many interventions had very low certainty evidence, often from lack of data.

We prioritized findings from 21 interventions where there was moderate certainty for at least some benefit. Fourteen of these had a focus on exercise, the majority being supervised (for > 2 sessions) and of long duration (> 3 months), and with balance/resistance and group Tai Chi interventions generally having the most outcomes with at least low certainty for benefit. None of the interventions having moderate certainty evidence focused on walking. Whole-body vibration or home-hazard assessment (HHA) plus exercise provided to everyone showed moderate certainty for some benefit. No multifactorial intervention alone showed moderate certainty for any benefit.

Six interventions only had very-low certainty evidence for the benefit outcomes. Two interventions had moderate certainty of harmful effects for at least one benefit outcome, though the populations across studies were at high risk for falls. Vitamin D and most single-component exercise interventions are probably associated with minimal adverse effects. Some uncertainty exists about possible adverse effects from other interventions.

For outcome valuation, we included 44 studies of which 34 reported EQ-5D disutilities. Admission to long-term care had the highest disutility (1.0), but the evidence was rated as low certainty. Both fall-related hip (moderate certainty) and non-hip (low certainty) fracture may result in substantial disutility (0.53 and 0.57) in the first 3 months after injury. Disutility for both hip and non-hip fractures is probably lower 12 months after injury (0.16 and 0.19, with high and moderate certainty, respectively) compared to within the first 3 months. No study measured the disutility of an injurious fall. Fractures are probably more important than either falls (0.09 over 12 months) or functional status (0.12). Functional status may be somewhat more important than falls.

For intervention preferences, 29 studies (9 qualitative) reported on 17 comparisons among single-component interventions showing benefit. Exercise interventions focusing on balance and/or resistance training appear to be clearly preferred over Tai Chi and other forms of exercise (e.g., yoga, aerobic). For exercise programs in general, there

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 3 of 32

is probably variability among people in whether they prefer group or individual delivery, though there was high certainty that individual was preferred over group delivery of balance/resistance programs. Balance/resistance exercise may be preferred over education, though the evidence was low certainty. There was low certainty for a slight preference for education over cognitive-behavioral therapy, and group education may be preferred over individual education.

Conclusions To prevent falls among community-dwelling older adults, evidence is most certain for benefit, at least over 1–2 years, from supervised, long-duration balance/resistance and group Tai Chi interventions, whole-body vibration, high-intensity/dose education or cognitive-behavioral therapy, and interventions of comprehensive multifactorial assessment with targeted treatment plus HHA, HHA plus exercise, or education provided to everyone. Adding other interventions to exercise does not appear to substantially increase benefits. Overall, effects appear most applicable to those with elevated fall risk. Choice among effective interventions that are available may best depend on individual patient preferences, though when implementing new balance/resistance programs delivering individual over group sessions when feasible may be most acceptable. Data on more patient-important outcomes including fall-related fractures and adverse effects would be beneficial, as would studies focusing on equity-deserving populations and on programs delivered virtually.

Systematic review registration Not registered.

Keywords Fall prevention, Interventions, Patient preferences, Systematic review, Network meta-analysis, Guideline

Background

About 20–30% of older adults (≥ 65 years old) experience one or more fall each year [1]. In Canada, 5.8% of older adults report being injured in a fall (limiting normal activities) in the past 12 months, with higher rates among women than men (6.5% vs. 5.0%) and among those 80 or older (e.g., 9.6% for those ≥ 85 vs. 5.5% at 65–69 years) [2]. Of those injured, 70% sought medical care; 39% experienced a fracture, accounting for the majority of cost, morbidity, and mortality from fall-related injuries [2, 3]. Fall-related hospitalizations occur annually in about 15 per 1,000 older adults, and the average length of hospitalization from fall-related injuries is nearly twice as long as that for all causes of hospitalization (i.e., 21 vs. 12 days) [2]. In Canada, 35% of older adults hospitalized for a fall are discharged to long-term care (LTC) [4]. Less serious injuries from falls, such as bruising, lacerations, and sprains, can still lead to pain, affect function, and result in substantial costs to individuals and the healthcare system [5]. Additionally, many older people experience psychological difficulties after a fall, including fear of falling or loss of confidence, which may contribute to further falling [6]. Falls are also associated with reduced quality of life [7] and can impact independence and overall health through loss of function in activities of daily living (ADLs) [8].

Falls are often the result of inability to maintain or regain one's balance due to a complex combination of risk factors [1]. Risk factors may be classified as follows: (i) biological, related to disease(s), and the natural aging process (e.g., balance and gait deficiency, acute or chronic health conditions, cognitive impairment, low vision); (ii) behavioral, such as use of unsuitable/poorly maintained assistive devices including footwear, fear of falling, and use of certain medications (e.g., psychotropics, sedatives, hypnotics) [9, 10]; (iii) social and economic (e.g., social isolation, poverty, poor access to healthcare); and (iv) environmental factors in the community (e.g., building entrances, lack of hand rails), home (e.g., type of furniture, home clutter), and/or related to weather and climate (e.g., icy surfaces) [1].

Screening patients for their risk for falls is one strategy to identify individuals who may benefit the most from further assessment and/or an intervention to prevent falls. Various methods for screening rely on single- or multiple-item history questions, self-report measures (e.g., Falls Efficacy Scale [11]), or performance-based measures (e.g., 30-s chair stand [12], Berg Balance Scale [13], Timed Up and Go [14]). A 2018 systematic review of 26 assessment tools for falls risk in older adults found that no single tool was sufficiently predictive to differentiate between people at high versus low risk of falls [15]. The authors concluded that clinicians should consider using at least two assessment tools together to better evaluate their patients' risk. Likewise, when evaluating the predictive ability of single and combined use of medical history questions, authors of another review concluded that no single question emerged as a powerful predictive tool, but that querying several factors together (i.e., fall history, difficulty with ADLs, use of an ambulatory device, concern about falling, and use of psychoactive medication) could be useful [16]. However, the time burden of utilizing multiple tools or questions may lead to unacceptably low uptake by primary care providers due to competing demands in the care of older patients, who often have multiple comorbidities to manage during

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 4 of 32

a clinical encounter. Furthermore, falls risk status may not modify intervention effects. Several reviews on falls prevention interventions have found no significant difference in effects for exercise, multiple component, or multifactorial interventions regardless of whether above-average or high falls risk was used for study inclusion [17–19].

Many interventions for preventing falls have been studied. These are often based on known, modifiable risk factors for falling. Most fall prevention interventions can be classified according to the internationally accepted taxonomy developed by the Prevention of Falls Network Europe (ProFANE) Group [20]. A major feature of the taxonomy is the distinction between different categories and combinations of interventions. Interventions may comprise *single-component interventions*, involving one or a combination of two or more interventions (e.g., gait and balance training with strength/resistance exercises) from the same category (e.g., exercise); *multicomponent interventions*, in which interventions from different categories are offered to everyone (e.g., exercise plus education); or *multifactorial interventions*, comprised of a multifactorial risk assessment used to determine which intervention components are offered to each participant. Single and multicomponent interventions may target individuals with one or more particular risk factor(s), but, unlike multifactorial interventions, they do not tailor intervention components to each individual's risk. Some interventions within the taxonomy are more applicable than others to the general population of community-dwelling older adults and practice of primary care (i.e., first-contact, accessible, ongoing, comprehensive, and coordinated care [21]). These may be provided directly by primary care providers (e.g., vitamin D supplementation, nurse-led education on falls risk and prevention), by an interprofessional team of providers (e.g., exercise and cognitive behavioral therapy [CBT]), through referral to an allied health care provider (e.g., an occupational therapist to conduct home hazard assessment and modification [HHA]), or in the community (e.g., patient-initiated attendance at Tai Chi classes). Other interventions, such as management of urinary incontinence or cataract surgery, usually target populations with a specific diagnosis or condition and may not have the primary aim of falls prevention. Apart from this taxonomy related to type of interventions, other features related to their delivery (e.g., group vs. individual) or content (e.g., type of exercise, comprehensiveness of risk assessment) may impact their effectiveness or ease of implementation.

Given similar information on the anticipated benefits and harms of interventions, different guideline panels, and individuals, may make different decisions based on their values and preferences [22]. Preferences for or against an intervention are a consequence of the relative importance people place on the expected or experienced health outcomes it incurs [23]. These outcome valuations can make use of comparisons between different health-state utility values (HSUVs) or data from other utility-based stated and revealed preference studies including contingent valuation studies, such as discrete choice experiments, or simple ratings scales or trade-offs. HSUVs reflect preference-based health-related quality of life (HRQoL) and represent the strength of an individual's preference for the health outcome or health state under consideration [24]. They are typically measured on a scale of 0 (death) to 1 (perfect health): a more desirable health outcome will have a higher utility value. When an event is considered to negatively affect one's usual health (e.g., fall) the scale is inverted to indicate the decrement in utility (disutility) from experiencing the outcome. HSUVs can be measured using direct choice-based utility elicitation methods such as standard gamble, time trade-off, or indirect methods using generic multi-attribute utility instruments such as the EQ-5D which generates a health state's utility based on previous valuations by members of the general public. When considering that multiple types of interventions may have varying impacts on different outcomes, the relative importance placed by patients on the potential outcomes may influence which interventions are considered more or less effective. While patient acceptance rates across various types of fall prevention interventions appear quite high (e.g., approximately 70% of older adults agree to participate in studies regardless of eligibility [25]), evidence on patients' stated preferences for different types and/or formats of interventions (e.g., [26–28]) could inform decisions about which interventions to recommend in general and to specific populations. Findings about patient preferences are one source of patient input for guideline panels to consider alongside the empirical evidence on the clinical effectiveness of interventions.

Objective

The aim of this set of three systematic reviews was to inform recommendations by the Canadian Task Force on Preventive Health Care (task force) for primary care providers on fall prevention interventions. The findings of these reviews will be considered alongside other sources of information on feasibility, acceptability, costs/resources, and equity, to make the final recommendations. We answered the following questions (key questions [KQs]):

1. What are the benefits and harms of interventions compared with usual care to prevent falls in commu-

nity-dwelling adults aged 65 and older? (*Benefits and harms*)

2. How do community-dwelling adults aged 65 and older weigh the potential benefits and harms of interventions to prevent falls? (*Outcome valuation*)

3. What are the preferences of community-dwelling adults aged 65 and older regarding different interventions demonstrated to prevent falls? (*Intervention preferences*)

Methods

To avoid duplication of effort, we utilized a previous review published in 2017, having a broader scope than ours, for identification of studies (to 2016) related to the benefits and harms and (when suitable) for data extraction and assessments of risk of bias [29]. In line with the previous review [29], and to maximize the utility of available data, we conducted network meta-analyses (NMA) on the benefits [30–32]. From searches conducted during our preparation for these reviews, we did not identify an existing review fully answering the questions on outcome valuations or intervention preferences, respectively, and therefore conducted de novo reviews for these questions.

These reviews were carried out following a protocol [33] and according to the task force methods [34], guided by the Cochrane Handbook and Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group [35–38]. This report follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guidelines for systematic reviews and the PRISMA extension statement for NMA, as applicable to each question [39, 40]. Any changes to the protocol are described herein; no change was made post hoc, that is, based on the results of any studies or analysis. A working group consisting of task force members and external clinical experts (see the "Acknowledgements") assisted in developing the eligibility criteria, intervention taxonomy and coding schema, and thresholds for effects, as described herein, but was not involved in the selection and risk of bias assessments of studies, data extraction, or analysis. This manuscript was reviewed by 10 stakeholders (Supplementary file 1 contains the comments and responses), with all comments considered and minor revisions made mainly to enhance clarity of language.

Eligibility

Detailed eligibility criteria for all KQs are located in Supplementary file 1 Table S1.1.

Key question 1

The population of interest for KQ1 was adults ≥ 65 years old living who were community dwelling, i.e., living in the community/private retirement homes/assisted living with minimal supports. We did not limit eligibility by participant fall history. We included studies where < 20% of participants were < 65 years old or required more than minimal at-home support, including those homebound or living in LTC (people could require homecare); the latter narrows from our original exclusion criteria of studies recruiting based on LTC residency. We excluded studies recruiting participants based on having ≥ 1 specific medical diagnosis that may require different fall prevention and management strategies than the general population (e.g., stroke, Parkinson's disease, severe dementia, or visual impairment) or that recruited participants on the basis of known vitamin D deficiency [≤ 30 nmol/L]). Studies recruiting inpatients were eligible if the intervention was primarily delivered in an outpatient or community setting (including primary care).

We included studies with at least one intervention aiming to prevent falls in community-dwelling older adults and delivered in, or referable from, primary care. A revision was made to the protocol to require that all studies prospectively collect data and report on at least one of our fall outcomes (falls, fallers, or injurious fallers). The review and our search were not intended to capture all studies (e.g., on vitamin D) aiming to prevent non-fall outcomes and incidentally reported on falls but rather to capture studies of interventions specifically targeting fall prevention. Because the causality of falls is complex, we included studies of interventions targeting known fall risk factors (e.g., balance or gait disturbance, fear of falling) if the authors proposed a mechanism to prevent falls (e.g., improving balance can reduce the likelihood of a fall), and the study reported a falls outcome. We excluded all pharmacologic interventions apart from vitamin D (with or without calcium), protective aids to reduce fractures from falls (e.g., hip protectors), podiatry assessments, alarms to signal a fall, and hearing aids or visual aids unless they were part of a multicomponent intervention. Interventions only providing screening/assessment results to general practitioners, without a protocol for follow-up, or focused on cognitive/executive function or visual attention were also excluded. We included dietary counselling but excluded sole provision of dietary supplements/meal replacements or fluid therapy. We also excluded studies where the only difference between interventions was an "add-on" (e.g., clinician reminders) to improve uptake or implementation of interventions already in place. Supplementary file 1 Table S1.2 has details of intervention eligibility.

The main comparator of interest was usual care (UC), which we defined as typical health or medical care received in a primary care setting without specific intervention to prevent falls. If authors described UC that met

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 6 of 32

Table 1 Framework for coding used to develop interventions (nodes) for the network meta-analyses

Intervention type	Variables used for full networks[a]	Possible number of nodes	Nodes used in at least one NMA (node number)
Control groups			
Usual care (UC)	None	1	UC (main comparator for all NMAs) (1)
Information (pamphlet on falls risk)	None	1	Information (2)
Attention control (social visits/engagement, gentle stretching, session(s) on healthy lifestyle, etc.)	None	1	Attention control (3)
Single-component interventions			
Active education	• Individual vs. group/mixed delivery • Intensity (1-2 sessions or ≤ 4 h vs. more intense)	4	Ed + Indv + Low (9) Ed + Indv + High (10) Ed + Grp + Low (11) Ed + Grp + High (12)
Exercise—balance/resistance	• Supervised (for > 2 sessions) vs. not (≤ 2 sessions) • Individual vs. group/mixed delivery • Virtual vs. offline (in-person) • Duration (≤ 3 [brief] vs. > 3 mos [long]) • Intensity of supervision (≤ 1 [low] vs. > 1 contact/week [high] (mean across total intervention duration, among long duration programs)	22	Bal/Res + Sup + Ind + Offline[b] + Brief (13) Bal/Res + SupH + Ind + Offline + Long (15) Bal/Res + SupL + Ind + Offline + Long (16) Bal/Res + SupH + Grp + Virt + Long (17) Bal/Res + Sup + Grp + Offline + Brief (18) Bal/Res + SupH + Grp + Offline + Long (19) Bal/Res + SupL + Grp + Offline + Long (20) Bal/Res + Uns + Ind + Offline + Brief (21) Bal/Res + Uns + Ind + Offline + Long (22)
Exercise—walking program (may include basic treadmill exercise)	• Supervised (for > 2 sessions) vs. not (≤ 2 sessions) • Individual vs. group/mixed delivery • Virtual vs. offline (in-person) • Duration of intervention (≤ 3 vs. > 3 mos)	16	Walk + Sup + Grp + Offline + Brief (23) Walk + Uns + Ind + Offline + Long (24)
Exercise—Tai Chi	• Supervised (for > 2 sessions) vs. not (≤ 2 sessions) • Individual vs. group/mixed delivery • Virtual vs. offline (in-person) • Duration of intervention (≤ 3 vs. > 3 mos)	16	Tai chi + Sup + Ind + Offline + Long (25) Tai chi + Sup + Grp + Offline + Brief (26) Tai chi + Sup + Grp + Offline + Long (27)
Exercise—other (e.g., square stepping, yoga, dance, aquatic)	• Supervised (for > 2 sessions) vs. not (≤ 2 sessions) • Individual vs. group/mixed delivery • Virtual vs. offline (in-person) • Duration of intervention (≤ 3 vs. > 3 mos)	16	Other + Sup + Ind + Offline + Brief (28) Other + Sup + Ind + Offline + Long (29) Other + Sup + Grp + Offline + Brief (30) Other + Sup + Grp + Offline + Long (31) Other + Uns + Ind + Offline + Long (32)
Medication review	• All-comers/universal vs. restricted/targeted to those at high risk/polypharmacy • Comprehensive (multiple medications) vs. focused (1-2 medication classes)	4	Med Review + All + Comp (36) Med Review + Targeted + Comp (37)
Vitamin D	• None: All doses and/or frequency/delivery routes were combined	1	Vit. D (8)
Home hazard assessment (HHA)	• Comprehensive (≥3 assessment home features) vs. limited (1–2 features) • Modifications provided or self-purchased • Intensity/dose (1 [low] vs. ≥ 2 home visits [high])	8	HHA + Comp + Provider + High (33) HHA + Comp + Provider + Low (34) HHA + Comp + Self + Low (35)
EAT—mobility aid or device (EAT = environment/assistive technology)	• None	1	Mobility aid (4)
EAT—communication aids (alarm to prevent falls)	• None	1	None
EAT—vision assessment & treatment	• All-comers/universal vs. restricted to those with vision problems	2	VAT—all-comers (5) VAT—targeted (6)
Psychological interventions (e.g., cognitive behavioral therapy (CBT))	• Duration of intervention (≤ 3 vs. > 3 mos)	2	CBT—long (38) CBT—brief (29)
Nutritional counselling	• None	1	None
Other	• E.g., sunlight intervention, chiropractic, whole-body vibration (WBV)	1 per type	WBV (7)

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 7 of 32

Table 1 (continued)

Intervention type	Variables used for full networks[a]	Possible number of nodes	Nodes used in at least one NMA (node number)
Multicomponent (MC) interventions			
MC with exercise	• Walking vs. other exercise type • Supervised exercise (Y/N) • Complexity (only adding education [simple] vs. other(s)) • Requires ≥1 modifications (HHA or device) (Y/N)	12	Walk + Sup + Education (40) Exercise + Sup + Complex + HHA (41) Exercise + Sup + Complex + No mod (42) Exercise + Sup + Education (43) Exercise + Uns + Education (44) Exercise + Uns + Complex + HHA (45) Exercise + Uns + Complex + No mod (46)
MC without exercise	• Potentially requires ≥1 modifications (HHA or device) (Y/N)	2	HHA + Other (47) MC other (no HHA or exercise) (48)
Multifactorial (MF) interventions			
Comprehensive (4 + assessment domains)	• Major focus on exercise (Y/N) (coded as no if lack of information reported about duration, intensity, type, etc.) • Potentially requires ≥1 modifications (HHA or device) (Y/N)	4	MF + Comp + Focus on exc + HHA (49) MF + Comp + Ltd exc + HHA (50) MF + Comp + Ltd exc + No mod (51)
Limited	• Major focus on exercise (Y/N) • Potentially requires ≥1 modifications (HHA or device) (Y/N)	4	MF + Ltd + Focus on exc + No mod (52) MF + Ltd + Ltd exc + HHA (53)
Mixed interventions (MF assessment + additional component delivered to everyone)			
Mixed: MF comprehensive (≥4 intervention components) + other	• Adding education only • Adding exercise only • Adding HHA or device only • Adding exercise + HHA/device • Other	5	Mixed + MF Comp + Educ (54) Mixed + MF Comp + Exc (55) Mixed + MF Comp + HHA (56) Mixed + MF Comp + Exc + HHA (57) Mixed + MF Comp + Other (58)
Mixed: MF limited + other	• Adding education only • Adding exercise only • Adding HHA or device only • Adding exercise + HHA/device • Other	5	Mixed + MF Ltd + HHA (59) Mixed + MF Ltd + Exc + HHA (60)

Abbreviations: *CBT* cognitive behavioral therapy, *Comp* comprehensive, *HHA* Home Hazard Assessment, *Ltd* limited, *MC* multicomponent, *MF* multifactorial, *mos* months, *UC* usual care, *VAT* visual assessment and treatment, *WBV* whole-body vibration

[a] For the simplified meta-regressions, we used one node per intervention type (left column) with the following exceptions: (i) we used supervised vs. not for balance/resistance exercise interventions, (ii) we combined all multifactorial interventions into one node, and (iii) we combined all mixed interventions into one node

[b] Offline, in-person

one of our intervention descriptions, we classified it as an intervention for our purposes. We also included studies with a non-/minimally active intervention comparator, such as a pamphlet on fall risks, gentle stretching, or social engagement activities. To maximize the utility of available data and generate effect estimates versus UC for interventions not directly compared to UC, we included head-to-head comparisons of different interventions and conducted NMAs [30–32]. This form of analysis simultaneously evaluates a suite of comparisons using data from direct comparisons between treatments to inform comparisons for treatments that have not been directly compared (referred to as indirect evidence). We excluded head-to-head studies that only compared interventions that fell within the same category in our intervention taxonomy and coding framework (e.g., different doses of vitamin D, intensities of strength training, different delivery providers) (see the section "Intervention coding and development of NMA nodes" and Table 1).

The outcomes of interest were chosen and rated by the task force for their importance to decision-making, using GRADE guidance [33, 38]. Eight outcomes were considered critical for decision-making (i.e., rated 7 or above on a scale of 1–9): number of fallers, number of falls, number of injurious fallers, number of fractures (any fracture; refined to exclude non-symptomatic vertebral fractures found via x-ray), number of hip fractures, residential status/institutionalization, HRQoL, and functional status. One result was selected for each outcome in each study using a hierarchy developed based on task force input either before (for injurious falls, using the most serious in each study) or after (for HRQoL and functional status) development of the protocol (Table S1.1). Intervention-related adverse effects (AEs) were rated as important (i.e., rated 4–6); we categorized AEs into four categories for extraction: any AE, serious AEs (both proportions), falls, and fractures (both if attributed to the intervention).

We included randomized controlled trials (RCTs) of any design and sample size with at least 3-month follow-up after randomization to adequately capture potential effects on the outcomes. We included studies only reported in conference proceedings or unpublished reports if the results were confirmed by the authors as final.

Key questions 2 and 3

Apart from the same main population as KQ1, for KQs 2 and 3, studies recruiting family members or caregivers on behalf of people with cognitive impairment were eligible (Table S1.1). For outcome valuation related to LTC admission in KQ2, studies of LTC residents were eligible if they measured the utility/disutility of admission to LTC within the first 6 months of residence. For KQ2, as per protocol, because we located studies reporting HSUV data for all of the critical outcomes (i.e., health states) in comparison with a "healthy" comparator group (i.e., without the health state being measured), we only included these comparative HSUV studies; retrospective pre-event measurement (i.e., via recall) was an acceptable comparator. This criterion allowed us to calculate a disutility from the outcome (i.e., utility of "healthy" state minus utility of health state from outcome such as fall; range 0 = no disutility to 1.0 = disutility equal to death and approximately ≥ 0.05 considered meaningful), which allows for easier interpretation across populations since it accounts for baseline health status. Participants could have experienced the outcome themselves or been provided with hypothetical scenarios about the experience. Apart from the above eligibility for comparative HSUV studies, we included studies using other utility-based stated and revealed preference methods including contingent valuation studies, such as discrete choice experiments, simple ratings scales, or trade-offs; if we had not found data from these study designs for a particular outcome, we would have included non-utility data from surveys or qualitative studies. For KQ3, we examined preferences between different interventions or between different attributes of interventions within one category. We selected studies after assessing the certainty of the evidence in KQ1 and only included studies measuring preferences among interventions for which we had moderate or high certainty of benefit for at least one critical outcome. Intervention attributes of interest included duration, cost, and format (group vs. individual and virtual vs. in-person delivery). Participants could have either participated in the relevant fall prevention intervention(s) or been provided with information about the interventions. We relied on quantitative studies using utility-based methods but considered data from surveys and, if necessary to fill gaps, qualitative studies.

For all KQs, we included reports in English or French, the official languages of Canada, as per task force methods. Language restrictions in systematic reviews do not appear to bias results from meta-analyses [41, 42]. No restrictions were placed on country or risk of bias. For KQ1, no limit on publication date was placed. However, for KQs 2 and 3, we limited inclusion to studies published on or after 2000 because it is expected that preferences change over time.

Information sources and search strategy

For KQ1 on benefits and harms, we located full texts of all studies included in the previously published review [29]. To identify new evidence, our information specialist updated the previous review's peer-reviewed searches (Supplementary file 1 contains all search strategies), limited to records after January 1, 2016. The search contains Medical Subject Heading terms and key words combining the concepts of falls/fallers, adults, and randomized controlled trials. We searched Ovid MEDLINE (1946-), Ovid Embase (1996-), Wiley Cochrane Central Register of Controlled Trials (inception-), and AgeLine. Reference lists of all new included trials and recent systematic reviews were hand-searched by one reviewer. We also searched the World Health Organization Clinical Trials Search Portal [43], which searches multiple trial registries. First authors of potentially relevant studies reported in conference abstracts or trial registries were contacted by email (with two reminders over 1 month) to confirm the study's eligibility or to obtain full study reports and/or additional study or outcome data. Database searches were run November 15–16, 2020, and updated on August 25, 2023 (with the exception of AgeLine which was not a unique source for any of the included studies from the initial search).

The information specialist developed a search covering both KQs 2 and 3, combining Medical Subject Heading terms and key words for falls, fractures, and transition to residential care with those for patient preferences, quality of life, preference-based instrument/methodology terms (e.g., EQ-5D, conjoint analysis), decision-making, attitudes, and acceptability. This search was peer-reviewed by another librarian using the Peer Review of Electronic Search Strategies 2015 checklist [44]. We searched Ovid MEDLINE (1946-), Ovid PsycINFO (1987-), and CINAHL via EBSCOhost (1937-) databases, screened studies in KQ1 for KQ2 and 3 eligibility, and hand-searched reference lists of included studies and relevant systematic reviews. Database searches were initially run September 23–25, 2021 and updated on June 9, 2023.

Selection of studies

Database search results were exported to an EndNote library (version X7, Clarivate Analytics, Philadelphia,

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 9 of 32

USA, 2018) for record-keeping and deduplication. Records identified through hand-searching were manually added to the library. Unique records from the database searches were uploaded to DistillerSR (Evidence Partners Inc., Ottawa, Canada) for screening. Structured eligibility forms for screening and selection were developed and pilot tested by all reviewers involved in screening (J. P., L. A. G., S. S., A. W., A. R. A.) with samples of 100 abstracts and 20 full texts, respectively, until agreement was high (> 95%). Thereafter, all titles and abstracts were screened by two reviewers independently, using broad inclusion criteria. Efforts to retrieve the full texts of all potentially relevant records were made, with the exception of qualitative studies potentially relevant to KQ2 where we had sufficient evidence from other study designs. Two reviewers independently reviewed all full texts (including the studies from the previous review) [29], and disagreements were resolved by consensus or adjudication by a third reviewer (L. A. G. or J. P.). Records identified from reference lists, trial registries, and websites were screened by one experienced reviewer, with two reviewers reviewing full texts. We recorded reasons for all exclusions at full text.

Data extraction
Key question 1
We planned to utilize data extraction from the previous review [29], but differences in our classification of interventions and definitions of several outcomes (e.g., separating fallers and falls, using a hierarchy for injurious falls, addition of functional status outcome) led to extracting all data anew. All data were extracted by one trained reviewer and verified for accuracy and completeness by a second trained reviewer. Data were extracted into purpose-built forms in Microsoft Excel that were piloted on a sample of 10 studies by all reviewers involved in data extraction (L. A. G., A. W., S. S., A. R. A.).

We collected data on the following: study characteristics (i.e., year and country of conduct, funding source [industry vs. not], enrolled sample size, recruitment setting, trial design); intervention components (including, as applicable, delivery method, location, provider, degree of supervision, details of any assessments made, provision of any modifications, duration in weeks, and number of sessions/hours); description of UC or other control; participant characteristics (sex, age, proportion with previous falls, any enrolment criteria related to a clinical condition); outcome measurement tools and ascertainment; and results data (with sample size) at longest follow-up. Studies with populations requiring contextual or equity considerations (e.g., Indigenous peoples, newcomers, low income, Black Canadians) by the task force were also noted. Data were collected to allow examination of

the homogeneity and transitivity assumptions for the NMAs and for assessment of risk of bias.

For binary data, we extracted crude events in each arm, when reported, or risk, rate, or odds ratios if necessary or suitable (i.e., adjusted for clustering). If falls were reported as incident rates (e.g., over person-years), we used the mean follow-up duration of the study to calculate the total number of events. We assumed injurious fall events only occurred once per person during follow-up if the number of injured fallers was not reported; in a few cases where only data on any injurious falls were available, we extracted data for events knowing this may overestimate the risk, but likely not the relative effects. When data for multiple fracture types (e.g., wrist, tibia, distal femur) were reported, we extracted the findings for all. Unless specified, we assumed that a participant had only one fracture, hip fracture, or serious AE during follow-up. If studies only reported serious AEs, we used this data for both the any AE and serious AE outcomes and performed sensitivity analysis. When authors reported zero AEs, we extracted this as zero for any AE and serious AEs outcomes but not for falls or fractures, which we did not assume authors considered to be intervention-related AEs.

For continuous outcomes measures, we extracted (by arm) the mean baseline and endpoint or change scores, standard deviations (SDs) or other measure of variability, and number analyzed. Means were approximated from medians, when necessary. When SDs were not provided, they were computed or estimated using established imputation methods [35]. Results were extracted from figures using WebPlotDigitizer version 4.6 [45] when no numerical values were provided. For crossover trials, we limited data extraction to the first period of the study to eliminate the possibility of carry-over effects. When two or more interventions in a multi-arm trial were classified as the same intervention in our taxonomy (e.g., different intensities/difficulty levels of a balance/resistance training intervention), we combined the results from those groups to avoid loss of information [35].

Key questions 2 and 3
For these questions, we collected data on the population as in KQ1, as well as exposure to any of the related outcomes and/or to fall prevention interventions. We collected details about instruments used, including development and composition of scenarios of health states; choice tasks, including definitions of all attributes; and survey questions. For KQ2 data on HSUVs, we extracted sample sizes, means, and 95% CIs, or medians and interquartile ranges (IQRs), when necessary; when other (e.g., standard deviations) measures of variance were reported, we extracted this information to later calculate or impute

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 10 of 32

95% CIs for analysis [35]. When HSUVs for multiple fracture types (e.g., wrist, tibia, distal femur) were reported, we extracted findings for all. When multiple HSUV measurement tools were used by a study for a single health state, we preferentially extracted EQ-5D over other utility measures as this is the most commonly used tool. If a study without EQ-5D used more than one utility tool, we extracted data from each tool, standardized them and pooled them. Our protocol did not specify a specific timepoint for measurement; rather, specifying that time since event (≤ 12 vs. > 12 months) would be a subgroup variable of interest. Few studies reported on fracture outcomes in the long term (> 12 months; although many at 12 months), and this variable was judged as a major potential confounder; therefore, we defined two measurement time points for utilities for our primary synthesis: time closest to the event (if ≤ 3 months) and time closest to 12 months (but > 3 and < 24 months). All relevant results data from studies not employing HSUV methods in KQ2 (e.g., rankings of outcomes by importance, coefficients from discrete choice experiments) and reported in quantitative studies for KQ3 were extracted. Qualitative findings for KQ3 were copied and pasted from the relevant findings or results sections of the report into a Microsoft Excel table for analysis. All data were extracted by one trained reviewer and verified for accuracy and completeness by a second trained reviewer. Data were extracted into purpose-built forms in Microsoft Excel that were piloted by all reviewers involved in data extraction (L. A. G., A. W., S. S.).

Within-study risk-of-bias assessments

To align with the previous review [29], we used the Cochrane Effective Practice and Organisation of Care Group's risk-of-bias tool for KQ1 [46]. Briefly, studies were evaluated on the following domains: random sequence generation, allocation concealment, contamination, similarity of baseline characteristics, similarity of outcome measures at baseline, incomplete outcome data, blinding of outcome assessors, and selective outcome reporting. In each study, each outcome was rated as being at low, unclear, or high risk of bias in each domain. Studies with an active comparator but no blinding were not considered at high risk of bias for contamination. Outcomes with no ratings of high risk in any domain were considered at overall low risk of bias. Outcomes with a ≥ 1 ratings of high risk in any domain other than incomplete outcome data were considered at moderate risk of bias. Outcomes rated at high risk of bias for incomplete outcome data and ≥ 1 other domain were considered at overall high risk of bias.

For the KQ1 outcomes of AEs attributed to the intervention, we revised the protocol to assess risk of bias using an adapted McMaster Quality Assessment Scale of Harms (McHarms) tool (Supplementary file 1) [47]. McHarms was developed to compare AEs between intervention and control groups in trials of pharmacologic interventions. Because we were not interested in the AEs in UC/control groups, we removed some questions and focused on whether ascertainment of AEs was active (rather than passive/spontaneous), used a valid measure (e.g., events that could vary in severity were classified as such), and if attrition was acceptable. If there were serious concerns for ≥ 1 of the seven questions in our form, the study was rated as high risk of bias.

There was no validated or commonly accepted tool to assess risk of bias in studies measuring patient values or preferences. Therefore, for KQs 2 and 3, we developed a critical appraisal tool adapted from the within-study risk-of-bias domains in the GRADE guidance for preference-based studies (Supplementary file 1) [36]. Our assessment is comprised of signalling questions addressing four main concepts: (1) selection of representative participants into the study, (2) completeness of data, (3) properties of the measurement instrument, and (4) data analysis. If there were serious concerns in ≥ 1 domain, the study was rated at high risk of bias. Qualitative studies in KQ3 were assessed using the qualitative critical appraisal tool from the Critical Appraisal Skills Programme [48] and rated as having few, some, or significant concerns.

For trials from the prior review included in KQ1, we relied on assessments by the previous review team for use for the benefit outcomes [29]; we assessed harms anew using our modified tool. For newly identified studies and all harm outcomes, two reviewers independently assessed the risk of bias by using the previous team's reviewer instructions and came to consensus on the final rating for each question/domain, with adjudication by a third reviewer when necessary. In KQs 2 and 3, we obtained approval from the task force for a methods deviation for one reviewer to assess risk of bias with a verification by a second reviewer (instead of using two independent reviewers), and disagreements were resolved by consensus or a review lead. Each risk of bias tool was piloted with a sample of at least five studies until agreement on all elements was high; each tool went through two rounds of pilot testing. The risk of bias/quality evaluations were incorporated into our assessment of the certainty of the evidence for each outcome (see section on "Assessing the certainty of evidence").

Intervention coding and development of NMA nodes (KQ 1)

In consultation with the working group, we adapted the ProFANE taxonomy [20] (to align with the most relevant primary care interventions) and developed a coding

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 11 of 32

framework to create meaningful delineations between interventions ("nodes" when referring to the NMA). Interventions were first categorized as follows: single component (e.g., one or more interventions from a single category), multicomponent (more than one intervention from different categories offered to all people), multifactorial (one or more interventions from different categories offered based on individual risk assessment, not all patients receive all interventions), or mixed (multifactorial intervention combined with ≥ 1 intervention provided for all independent of needs assessment). For control groups, it was decided to separate UC from information controls (e.g., pamphlet) and attention controls (e.g., social engagement, sessions on healthy living, gentle stretching). After data extraction had begun, but before analysis, we decided on various intervention components (e.g., four main types of exercise) and attributes, with their relative importance for implementation (e.g., supervision, delivery [individual vs. group /mixed], duration, comprehensiveness of assessment, varying by type of intervention) (Table 1). For some single-component interventions, we did not code additional attributes because there were usually very few studies examining these types of interventions. The working group decided against comparing different doses or means of administering vitamin D. Multicomponent, multifactorial, and mixed interventions were coded based on the broad types of interventions within them. For example, multicomponent interventions were coded primarily based on whether they included exercise (defined as walking vs. other and by level of supervision) and then by whether the other component(s) were limited to education and, if not, included HHA or a mobility device.

Intervention coding was conducted for each arm in every study and followed a similar process to data extraction. A form based on the coding schema was piloted between two project leads (L. A. G., J. P.). Once sufficient agreement was reached, one reviewer coded all intervention arms in a study, and the other reviewer verified the coding (SS and ARA assisted for studies from our search update). Trials in which all arms coded into the same node ($n = 20$, see Supplementary file 1 excluded studies list) were excluded, since at least one valid comparison is required to include a study in an NMA. Only one RCT was required to create a node for the NMA.

Data synthesis for KQ1 (benefits and harms)

For all critical (benefit) dichotomous outcomes, an odds ratio (OR) was the measure of effect, with the exception of falls for which we used a rate ratio (RR). For studies reporting multiple types of fractures, we calculated an average effect across fracture types. Studies reporting zero events in all arms were not analyzed.

For continuous outcomes (e.g., HRQoL), we scaled all study findings to 0–100 (most measurement tools used this or similar ranges) and then used a mean difference (MD) as the effect measure [49]. A total of 95% confidence intervals (CIs) were used for measuring variance. We employed random-effects NMA in Stata (version 17.0) using a frequentist approach, which accounts for correlations between effect sizes from multi-arm studies [50]. In an NMA, a web (network) of different comparisons is constructed to incorporate both direct evidence from comparisons of interest and indirect evidence from comparisons with only one intervention in common. For example, comparisons between treatments A vs. C and treatments B vs. C will contribute to the effect estimate for treatments A vs. B.

Pooled proportions were calculated for AE outcomes reported by two or more studies per node, with 95% CIs calculated using an exact method. One exception was for vitamin D where there are several possible AEs/serious AEs and attribution becomes difficult; for these outcomes, we undertook random-effects pairwise meta-analysis in Review Manager (version 5.4.1) to generate the average effect (relative risk for any AE and risk difference for serious AEs where there were zero-event studies) for vitamin D versus UC/information controls. For these analyses, we performed sensitivity analysis removing the studies only reporting on serious AEs (but used for the any AE analysis) or only reporting on a single serious AE.

Assessment of transitivity

The validity of indirect evidence in an NMA relies on the principle of transitivity, which assumes that covariates that act as relative treatment effect modifiers are similar across different interventions, or adjusted for in the analysis, so the effects of all treatments included in the model are generalizable across all included studies [51]. We assumed that there would be transitivity based on our restrictions in eligibility to community-dwelling older adults living independently. To further evaluate whether the transitivity assumption was valid, we first investigated the distribution of clinical and methodological covariates across the studies in each node. Potential effect modifiers investigated were as follows: age of study population (mean age ≥ 80 vs. < 80 years), fall history (100% vs. 30–99% vs. < 30% participants with previous falls), recruitment setting (primary care/community vs. emergency department vs. in-patient), health context (country similar to Canada [high-income Organisation for Economic Co-operation and Development member [52] or United States] vs. otherwise), and duration of follow-up after randomization (< 12 vs. ≥ 12 months). For the health context, with working group input, we considered the USA as having a similar health context to Canada, as

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 12 of 32

older adult in the USA are eligible for government-sponsored healthcare insurance (Medicare). Further, within the context of research studies that usually provide interventions without cost to participants, and many delivered in community settings, results from studies in countries where healthcare access for older adults may be inferior to Canada likely apply well to Canada. Studies were coded for these variables, and we assessed for each node the proportion of the total sample size that came from studies having each level of the variable (e.g., 30% of node sample came from studies with a mean age > 80 years). Graphs were created, and we visually inspected the distributions to see if there were large variations across nodes.

Transitivity was further checked through network meta-regressions for the falls, fallers, and injurious falls networks (other networks were too sparse). Due to concerns that meta-regressions in the full network may be underpowered, we also carried out meta-regressions on a simplified network, in which interventions of the same type or with similar characteristics were collapsed in order to increase statistical power to detect potential effect modifiers (Table 1 footnotes). A meta-regression was carried out for each effect modifier and for each network (full and simplified) for the three outcomes, in comparison with UC. The effect modifiers included age, falls history, recruitment setting, and health context. Though not used for transitivity, we also ran meta-regressions for risk of bias (high vs. moderate/low) to inform our certainty assessments. Our protocol specified we would examine fall history in three categories (0–30% vs > 30–99% vs. 100%), but this was revised (0–99% vs. 100% [usually over the past year]) because the findings of categorical regressions are difficult to interpret and a substantial proportion (32%) of studies did not report fall history in this manner (e.g., recruited instead on presence of falls risk factors) but could be assumed to enrol a sample with fewer than 100% having previous falls. We did not examine duration of follow-up because of our limit to a minimum of 3 months and a subgroup analysis in the previous review [29] did not find any evidence of an impact. Results of the meta-regressions (relying mostly on the full network regressions) were also used in our assessments of evidence certainty for the domains of indirectness and within-study risk of bias, together with data on the contribution of the relevant variables across studies.

Assessment of coherence

Coherence (often termed consistency) refers to the agreement between direct and indirect effect estimates contributing to the overall network estimate for each comparison. We assessed incoherence within comparisons using the Separate Indirect from Direct Evidence (SIDE or node splitting) approach [53] and globally across the entire network using the design-by-treatment interaction model [50]. These methods provide p-values, and $p < 0.10$ and $p < 0.01$ indicate some or major incoherence, respectively [54]. We also imported our data into the Confidence in Network Meta-analysis (CINeMA) web application [55] to generate contribution matrices of the proportion of participants in each comparison that came from direct (UC/control) and indirect comparisons. All results were used when assessing our NMA model and certainty of the evidence.

Assessment of heterogeneity

The concordance between CIs, which do not capture heterogeneity, and prediction intervals (PrI)—showing where the true effect of a new study similar to the existing studies is expected to lie—can be used to assess the importance of heterogeneity. We calculated PrIs for all network estimates (for comparisons with UC) and used these findings in our assessments of the certainty of evidence.

Reporting (across-study) bias assessments

To inform our certainty assessments, we assessed potential across-study bias in our effect estimates (i.e., network estimates, pooled AE estimates) from outcome data unable to contribute to our analyses because they were not reported even though the study should have collected them (i.e., missing outcome data) or because of publication bias (assessed via small study bias). To identify missing outcome data, we looked for trial registrations or protocols for all included studies. For falls or fallers, when it was clear that participants were asked to record all falls during follow-up (e.g., in diaries), we expected the study to report both falls and fallers despite what was reported in their protocol and considered this data missing at bias when they did not. For other outcomes, we considered outcome data to be missing at bias if a protocol, methods, or registration reported intention to collect it but we were unable to locate the results data for that outcome in any trial publication. For intervention-related AEs, clinical input suggested that all harm outcomes as defined for this review should have been measured. Therefore, we assessed data as missing at bias for all AEs outcomes not reported in each study. We analyzed small study effects visually and quantitatively using Egger's test [56] when ≥ 8 RCTs compared an intervention with UC using random-effects pair-wise meta-analysis in Stata. Other available methods for investigating small study effect in NMAs (e.g., comparison-adjusted funnel plots) [32] were not considered suitable as interventions could not be ordered in a meaningful way.

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 13 of 32

Presentation of NMA results and consideration of absolute effects

We generated network plots for each outcome, with the size of the nodes corresponding to the number of participants randomized to each treatment and the thickness of the edges (lines between nodes) weighted by the number of trials evaluating the comparison. Although we originally planned to present results for all pairwise comparisons in league tables, the large number of nodes included in our networks makes this format impractical; therefore, we present the network estimates of ORs or RRs and 95% CIs for all nodes compared to UC in forest plots (ranked by effect size).

With input from the working group, decision thresholds around effect size were created to delineate small (or greater) and moderate (or greater) benefit or harmful effect (i.e., unintended direction of effect for benefit outcomes). For fallers, these were set at OR=0.8 and OR=0.7 or OR=1.25 and $OR=1.4$, respectively, corresponding to approximately 5 and 8 fewer/more fallers per 100 people than the control event rates (i.e., the average rate for the UC group) in the studies. For each outcome, the control event rate across all nodes was calculated using the variance-stabilizing Freeman-Tukey double arcsine approach. To determine the thresholds for the other dichotomous outcomes, we used the pooled control event rate for fallers (41%) to convert the OR thresholds above into corresponding RRs (i.e., *RR*=0.88 or 1.13 for small and *RR*=0.80 or 1.20 for small and moderate benefit or harmful effect, respectively) which were then used as the thresholds across outcomes (with back-conversion to the corresponding OR using the respective control event rates) (Supplementary file 1 Table S1.5 contains the thresholds across outcomes). For harm outcomes, the thresholds of effect for any AE and falls (attributed to the intervention) were similar to those for the benefit outcome falls during follow-up (5 and 8 per 100 for small and moderate effects), while those used for fractures (attributed to the intervention) and serious AEs were 0.5 and 0.8 per 100 for small and moderate effects, respectively. For patient-reported outcome measures (i.e., functional status and HRQoL), which were all scaled to 0–100, we considered a 5-point difference to be at least a small effect and a 7-point difference to be at least a moderate effect, based on previous evaluations of meaningful change in patient-reported outcomes [57, 58]. Though the applicable relative effects of each network estimate were used for our certainty assessments, we present these together with calculated absolute effects for each network estimate.

Data synthesis for KQ 2 (outcome valuation)

A large majority (39/44; 89%) of the included studies measured HSUVs, and our synthesis focused on these findings. In each study, we estimated the disutility for each evaluated health state by subtracting the reported utility of the health state (e.g., fracture) from the utility without (when a control group was used) or before (if pre-event utilities were measured, including by recall, among those with the event) the heath state. When a study reported utilities for multiple types of fracture, we took a mean across types weighted by number of individuals reporting on each fracture type; this differed from our protocol that specified we would use ranges of values but allowed us to conduct a quantitative pooled analysis.

Disutilities for each health state, separated by utility measurement tool, were pooled in Review Manager version 5.4.1 by the inverse variance method using a random-effects model. For each analysis, pre-planned subgroup analyses were conducted based on study-level characteristics including mean age (<80 vs. ≥ 80 years), proportion female ($<80\%$ vs. $\geq 80\%$), and proportion with a history of previous fracture ($<30\%$ vs. $\geq 30\%$) [59]. We also added subgroup analysis for whether the non-fall outcomes were attributed (by the study authors) to a fall and (for fractures) for time since event among the primary analysis for ≤ 3 months (≤ 1 week vs. >1 week). We intended to use the subgroup estimates for our conclusions and certainty assessments if subgroup findings had been judged to be credible (e.g., differing significantly [$p \leq 0.10$] and by an important magnitude, based on several studies, reduced heterogeneity within subgroups). We also considered any data related to populations that may require equity or other contextual considerations (e.g., Indigenous peoples, newcomers, low income, Black Canadians). Apart from the estimated disutility of each health state, we revised our methods to also rank the health states primarily based on their EQ-5D disutility, though findings from other utility and non-utility measures were also considered. Study methods sections and protocols, if available, were used to assess for risk of bias from missing outcome data (e.g., study measured but did not report a useable result for a health state).

Data synthesis for KQ 3 (intervention preferences)

Eligible interventions for which preferences were sought (i.e., having at least moderate certainty of benefit from KQ1) included the following: cognitive behavioral therapy (CBT), education, three of four exercise types (balance and resistance, tai chi, other, but not walking), multicomponent (with exercise), and mixed interventions. We were unable to code interventions evaluated in preference studies with the same level of detail as in KQ1 and so relied on comparisons between categories of interventions and intervention attributes instead. Intervention attributes of interest included duration, cost, and format (group vs. individual and virtual vs. in-person

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 14 of 32

delivery). Preference between interventions (or intervention attributes) within included quantitative studies was categorized into four levels: "no clear preference," "slight preference," "clear preference," and "strong preference." Level of preference was determined by the ratio of responses favoring each intervention being compared: we considered a ratio > 2 to indicate a strong preference (e.g., 70% preferred exercise vs. 30% preferred education), 1.25 to < 2.0 was considered a clear preference, 1.12 to < 1.25 was considered a slight preference, and a ratio < 1.12 was considered to not indicate a clear preference either way (i.e., variable preferences). In qualitative studies, used when no quantitative findings were located for a given comparison, we did not quantify a strength of preference. Instead, findings were categorized as preference, possible preference (e.g., participants made negative statements about intervention A but neutral/mixed statements about intervention B), or no clear preference (e.g., positive or negative attitudes were described for both interventions).

To synthesize results across studies, preferences between pairs of interventions were informally tabulated across studies with stronger preferences given a higher weight and checked for agreement by a second reviewer. For comparisons with similarly strong but contradicting preferences, it was not possible to determine whether the disagreement was due to imprecision, population, or methodologic factors; therefore, these were rated as "variable preferences" rather than "no clear preference."

We examined any data related to populations that may require equity or other contextual considerations (e.g., Indigenous peoples, newcomers, low income, Black Canadians). Study methods sections and protocols, if available, were used to assess for risk of bias from missing outcome data (e.g., study measured a preference but did not report a useable result).

Assessing the certainty of the evidence

For KQ1 benefit outcomes, we assessed the certainty of evidence for each node compared to UC for all critical (benefit) outcomes guided by the CINeMA approach for NMAs. CINeMA is based on the GRADE framework, although it has conceptual and semantic differences [33, 60] and covers six domains: within-study bias, across-studies bias (i.e., publication and other reporting biases), indirectness, imprecision, heterogeneity (i.e., variation between studies within a comparison), and incoherence (i.e., disagreement between direct and indirect evidence for a comparison). For each outcome as follows: (i) data used for assessing transitivity (i.e., meta-regressions and contribution statistics, see section "Data synthesis for KQ1 (benefits and harms)") were used together for assessments of within-study bias and indirectness, (ii) assessments of missing outcome data and small study

effects were used for across-study bias, (iii) 95% CIs were compared with the thresholds for each outcome to assess imprecision without relying on statistical significance (see section Presentation of NMA results and consideration of absolute effects), (iv) concordance between CIs and PrIs was used for heterogeneity (e.g., rated down only if PrIs gave further concerns after assessment of imprecision), and (v) global and local coherence statistics were considered with information on the amount of direct/indirect evidence to assess incoherence. Similar to GRADE, each outcome from RCTs started at high certainty and was rated down to moderate, low, or very low based on the number of domains with serious concerns. The six CINeMA domains are interconnected and are considered jointly rather than in isolation [60]; if concerns in two domains were likely to be highly related, we only rated down once. Further, because we made assessments across six domains (vs. five domains for GRADE), when we had serious concerns about a domain, we only rated down by a half (0.5) instead of entire level (rounding up as applicable), as done previously [61]. For each domain, we rated down by a maximum of two levels: the main example being for imprecision, when the 95% CI crossed all four of our thresholds (indicating there could be moderate benefit or harm), and the network estimate showed little-to-no difference. When a moderate effect size (based on the network estimate exceeding this threshold) was rated down more for imprecision than it would be for a small effect, we report the higher certainty of a small effect. Details of our approach to assessing each domain are provided in Supplementary file 1.

Certainty in the evidence for harm outcomes followed the traditional GRADE domains of imprecision, inconsistency, reporting bias, within-study risk of bias, and indirectness. The thresholds of effect were used primarily to assess imprecision (if adequate sample sizes were used) and inconsistency (e.g., findings were inconsistent when study estimates were on either side of the threshold). We also rated down for inconsistency (i.e., lack of consistency) when > 90% of a node's sample size came from a single study. The evidence was rated down for reporting bias if there was evidence of small-study effects or if studies contributing > 50% of participants for the relevant node did not report on the outcome; a similar approach was used for within-study risk of bias, although we also compared the results between high and low/moderate risk-of-bias studies and considered whether conclusions would likely change before rating down. For indirectness, we rated down for the outcomes of any AE or fracture/serious AE when > 50% study participants of the studies contributing to an arm reported only on serious AEs or one type of fracture/serious AE, respectively. We also rated down for indirectness when ≥ 2 effect modifiers

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 15 of 32

(e.g., mean age > 80 years and 100% previous fallers) were present in ≥ 80% of the total sample.

For KQ2 and 3, our assessments in the certainty of the evidence were guided by the GRADE guidance for patient preferences [36, 37], which uses the standard domains of risk of bias, indirectness, imprecision, inconsistency, and reporting bias. All findings started at high certainty for this assessment. We rated down findings when > 50% of the participants came from studies at high risk of bias. We rated down findings for inconsistency if there was only one study for both KQs and for KQ3 if similar numbers of studies showed preferences in opposite directions or there was substantial disagreement compared to similar interventions. We assessed indirectness based on population characteristics (e.g., rating down comparisons when studies with a mean age > 80 contributed > 50% of the participants) and in KQ3 also on intervention specificity (e.g., if group/individual format or exercise type was not specified). For assessing imprecision, we relied on 95% CIs for KQ2; we had concerns when these were wider than 0.1 unless it could be explained by substantial inconsistency between studies' estimates. We relied on sample size to assess imprecision in KQ3 because CIs were not generated as part of the synthesis; based on GRADE guidance [36], comparisons with fewer than 380 participants were rated down. In KQ3, we also rated down for imprecision comparisons based on qualitative data, since it is not possible to quantify the precision of this kind of data. Finally, we considered whether missing outcome data or the presence of industry funding may have impacted the findings in KQ 2 and 3.

Ranking interventions in KQ1

We did not, as per protocol, rely on surface under the cumulative ranking curves (SUCRA) and heat rank plots to rank the interventions. Though SUCRA values account both for the location and the variance (imprecision) of all relative treatment effects, we wanted to also account for other factors related to the certainty of the evidence for each outcome (e.g., risk of bias, indirectness) as well as factors related to the multiple outcomes that were assessed. After gaining input from the task force, it was determined that decision-making would rely on findings of moderate or high certainty for the critical outcomes; this was our primary criteria for ranking interventions. For additional interpretation, we also applied a hierarchy based on the number of outcomes for which benefit was shown and, within these sub-groupings, the number with moderate or high certainty and, if needed, the certainty across the fall outcomes.

Results

A total of 290 studies were included across all KQs, with two studies included in multiple questions. Supplementary files 2 (KQ1) and 3 (KQs 2 and 3) contain the PRISMA flow diagrams and lists of included. Excluded studies with reasons are within the Supplementary files.

Benefit and harms of fall prevention interventions (KQ1)

Study selection and characteristics

After screening 8278 records and 1480 full texts, we included 219 RCTs reporting on 167,864 participants (Figure S2.1). We included 125 of 283 studies included in the previous review [29]. The most common reasons for excluding studies from the previous review were as follows: wrong population (e.g., homebound, visually impaired; $n = 30$), wrong setting (hospital or LTC; $n = 59$), or ineligible intervention (e.g., hip protectors, cataract surgery; $n = 22$). From our database searches to update the literature, many studies did not report on one of our three fall outcomes ($n = 205$) or the intervention did not target falls ($n = 229$, e.g., comprehensive geriatric assessment). Thirty-three studies were excluded during intervention coding because all study arms received treatments classified as the same intervention for our purposes (i.e., fell into a single category within our taxonomy).

Table 2 contains a summary of the study characteristics, and Supplementary file 2 has the characteristics by trial (Table S2.1) and risk-of-bias ratings (Table S2.2) by outcome for each trial. Outcomes missing with bias are listed in Table S2.1. A large majority (93%) of studies were undertaken in countries considered to offer similar heath care to older adults as in Canada. Most studies enrolled populations with a mean age < 80 years (85%) and a majority (but not large majority) of females (80%). Almost 70% of studies enrolled participants with an elevated risk for falling based on previous fall history and/or other risk factors (e.g., fear of falling, poor balance, ≥ 1 fall-risk increasing medications). Very few studies ($n = 12$ with 1.5% of total review population) recruited inpatients. Follow-up duration after randomization ranged from 3 to 96 months but was usually > 6 months with a majority (66%) ≥ 12 months. Fourteen trials (6.4%) followed up participants for > 24 months. In total, our coding defined 59 nodes among at least 130 possible. Of note, although nutrition therapy (i.e., dietary counselling) and aids for communication and signalling (i.e., alarm systems to prevent a fall) were eligible interventions, no eligible studies of these interventions were identified. Only one intervention was delivered virtually (node 17 in Table 1).

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 16 of 32

Table 2 Summary of study characteristics on the benefits and harms of fall prevention interventions (KQ1)

Characteristics	n RCTs (219 total); *N* participants (% of 167,864)
Age, mean, years	
• 65–72	42; 48,001 (28.6)
• > 72–79	128; 94,083 (56.0)
• ≥ 80	45; 24,386 (14.5)
• Not reported	4; 1394 (0.8)
% Female	
• 0–49	19; 8292 (4.9)
• ≥ 50–74	117; 132,706 (79.1)
• ≥ 75	81; 26,065 (15.5)
• Not reported	2; 801 (0.5)
History of falls	
• 0-29%	47; 55,187 (32.9)
• 30–99%	84; 42,578 (25.4)
• 100%	44; 16,108 (9.6)
• Not reported	44; 53,991 (32.2)
OECD countries or USA	
• Yes (including US)	191; 155,941 (92.9) (US studies: 47; 62,502 (37.2))
• No	28; 11,923 (7.1)
Recruitment setting	
• Primary care	28; 37,356 (22.3)
• Community	99; 88,566 (52.8)
• Mixed clinical and community	26; 13,411 (8.0)
• Emergency department	21; 6708 (4.0)
• Inpatient	12; 2465 (1.5)
• Not reported	7; 6365 (3.8)
Post-randomization follow-up duration (months)	
• 3–6	56; 14,399 (8.6)
• >6–12	116 (98 at 12 months); 39,751 (23.7)
• > 12	46; 113,601 (67.7)
• Not reported	1; 113 (0.1)

Abbreviations: *OECD* Organisation for Economic Co-operation and Development

Assessments of transitivity and incoherence

Transitivity was assessed as adequate for our analyses. We did not have serious concerns from visual inspection of the graphs showing contributions of effect modifiers across nodes, where there were small numbers of nodes containing entire populations with effect modifiers (e.g., out of 57 nodes for fallers, 1 had entirely inpatient recruitment, 3 had mean age > 80, 7 had patients with a previous fall history). Further, few meta-regressions were statistically significant ($p < 0.1$) in either the full or simplified networks. For the full networks, across all nodes and our 4 effect modifiers (i.e., > 200 tests), there were 10, 12, and 7 significant results for falls, fallers, and injurious fallers, respectively (results were similar for the simplified networks).

The global tests for coherence did not show major incoherence ($p < 0.01$), though there was some incoherence for the fracture and functional status outcomes ($p = 0.05$ and 0.09, respectively). Few local incoherence statistics showed major ($n = 1$ for fallers) or some ($n = 10$ for fallers) incoherence. We did not adjust our NMA geometry or analysis (i.e., consistency model) based on the transitivity and coherence assessments, but the results were used in our assessments of the evidence certainty.

NMA findings by outcome

Table 3 includes details about the network structure and findings of each NMA, including the pooled UC event rates used to calculate thresholds and absolute effects. Across all networks, the number of contributing RCTs ranged between 19 and 196, and the number of nodes ranged from 19 to 57. A small proportion (~ 5–10%) of the total sample for each outcome's network came from trials at high risk of bias. The larger

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 17 of 32

Table 3 Details on structure and findings for the NMAs, by outcome

	Falls	Fallers	Injurious fallers	Fracture	Hip fracture	LTC	Functional status	HRQoL
RCTs	169	173	93	67	19	33	55	58
Total sample measuring outcome (% of total review sample = 167,864)	77,483 (46.2)	83,306 (49.6)	84,644 (50.4)	61,190 (36.5)	28,855 (17.2)	9640 (5.7)	12,185 (7.3)	25,442 (15.2)
RCTs with UC (N)	78 (45,369)	93 (57,911)	45 (67,741)	41 (47,291)	12 (20,405)	20 (7316)	26 (6077)	32 (9458)
RCTs with high risk of bias, (N)	42; 8297	35; 8402	16; 4175	16; 4175	1; 430	8; 2364	14; 2381	21; 4325
Total number of interventions/ nodes Disconnected nodes[a], n (N)	57; 0	57; 0	49; 0	38; 0	19; 4 (930)	21; 2 (83)	33; 3 (812)	36; 6 (2130)
Average UC event rate (proportion)	1.15 per person	0.41	0.23	0.05	0.01	0.02	Not applicable	Not applicable
Certainty of evidence for benefit, n	High: 0 Moderate: 4 Low: 32	High: 0 Moderate: 18 Low: 9	High: 1 Moderate: 2 Low: 18	High: 1 Moderate: 1 Low: 17	High: 0 Moderate: 0 Low: 8	High: 0 Moderate: 0 Low: 10	High: 0 Moderate: 0 Low: 4	High: 0 Moderate: 0 Low: 3
Certainty of evidence for little-to-no difference, n	High: 0 Moderate: 0 Low: 11	High: 0 Moderate: 0 Low: 19	High: 0 Moderate: 0 Low: 15	High: 0 Moderate: 3 Low:1	High: 0 Moderate: 0 Low: 2	High: 0 Moderate: 0 Low:1	High: 0 Moderate: 0 Low: 14	High: 0 Moderate: 0 Low: 22
Certainty of evidence for harm (for benefit outcomes), n	High: 0 Moderate: 0 Low: 5	High: 0 Moderate: 2 Low: 3	High: 0 Moderate: 0 Low: 4	High: 0 Moderate: 1 Low: 6	High: 0 Moderate: 0 Low: 4	High: 0 Moderate: 0 Low: 8	High: 0 Moderate: 0 Low: 1	High: 0 Moderate: 0 Low: 2
Very low certainty (n due to no evidence)	6 (2)	7 (2)	18 (10)	28 (21)	44 (40)	39 (38)	39 (26)	31 (23)

Abbreviations: *HRQoL* health-related quality of life, *LTC* long-term care, *n* number of nodes, *N* sample size, *UC* usual care, *RCT* randomized controlled trial

[a] For disconnected notes, the total number of studies by outcome are 2 (hip fracture), 1 (LTC), 5 (HRQoL), and 4 (functional status)

networks were well connected with no disconnected nodes (where the node is not connected to the rest of the network through either direct or indirect comparisons), whereas the four smaller networks had two to six disconnected nodes. The sample sizes of the missing data due to disconnection were small relative to the total sample sizes of the respective networks.

Approximately, half of the nodes in each network had at least low certainty for at least one benefit outcome. The fallers outcome had the highest number of nodes (18 of 57 nodes) with moderate certainty for benefit. For the non-fall outcomes, there were many nodes with very low certainty evidence, often from lack of data. Imprecision and/or heterogeneity was the major concerns leading to low or very low certainty of evidence. Of five tests for small study bias, fallers and falls were significant ($p < 0.10$) for vitamin D, which we rated down. Missing outcome data was usually not a serious concern, with few

nodes rated down for this (i.e., 5 for fallers, 6 for falls, 8 for injurious falls, and 0–1 for other outcomes). We did not locate any data for the functional status outcome of new/increased homecare assistance.

Figures 1 and 2 present the network diagram and forest plot of network estimates, versus usual care, for fallers. Supplementary file 2 contains all network diagrams, network forest plots (ranked vs. UC), and evidence profiles (with network estimates and sample sizes for each node) for each outcome. Although we created more exercise nodes than for other intervention types, with the exception of vitamin D ($n = 9796$ for fallers) and a mixed intervention with comprehensive multifactorial assessment and HHA (provided to all) (node 56; $n = 7140$ for falls), there were no major differences on average in the sample sizes of the nodes (e.g., approximately 300–1500 participants for those showing moderate certainty evidence).
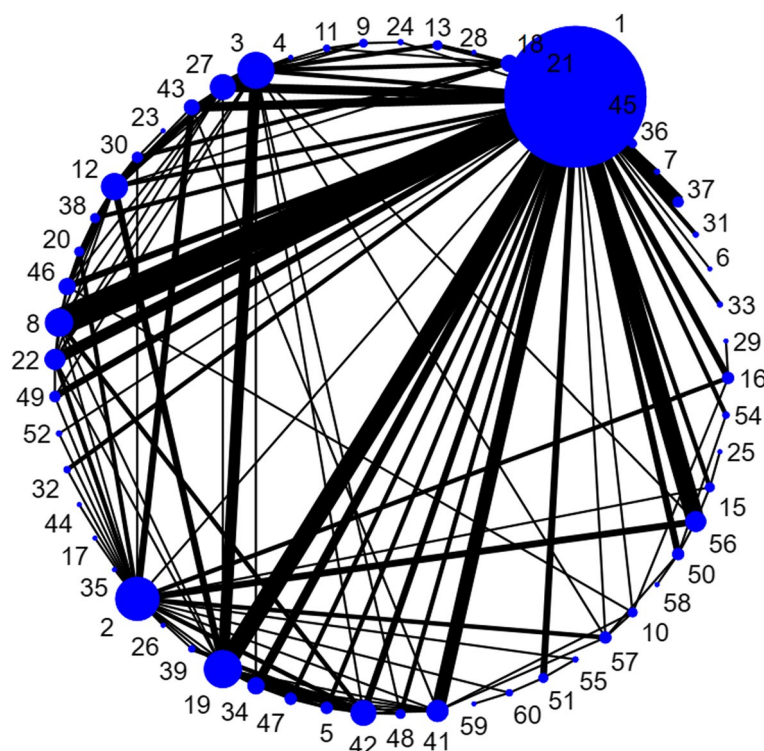
Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 18 of 32

Fig. 1 Network diagram for fallers network. Numbering of nodes corresponds to Table 1. The size of circles represents the number of participants (for the usual care node, the size was reduced), and the width of the lines represents the number of trials in the comparison

Summary of findings across all outcomes

Table 4 highlights the 21 interventions shown to have moderate or high certainty evidence for a beneficial effect for at least one benefit outcome and presents the summary of findings across all outcomes. The interventions are ordered by category; Supplementary file 2 Table S2.3 ranks these interventions based on the number of outcomes showing benefit and, among those, their certainty. Tables S2.4 and S2.5 contain the summary of findings for the following: (i) interventions having low or very low certainty across all benefit outcomes and (ii) interventions having moderate or high certainty evidence showing harm for one or more benefit outcomes without any evidence of benefit for other outcomes. When considering the effects in absolute terms, the findings for benefit or harm are best interpreted in terms of meeting/exceeding the thresholds applied rather than the network point estimate. For the benefit outcomes, the findings shown in light (small effect) or dark (moderate effect) green in Table 4 (based on a 12% or 20%, respectively, change in relation to the average UC event rate for each outcome) reflect approximately 5 or 8 fewer fallers, 14 or 23 fewer falls, 2 or 3 fewer injured fallers, 0.6 or 1 fewer fracture per 100 people and 1 or 2 fewer hip fracture for 1000 people. Light and darker red shading indicate the same absolute effects but in the opposite direction.

Fourteen of the 21 (67%) interventions with some moderate certainty evidence for benefit had a focus on exercise, either as a single component intervention ($n=11$) or within a multicomponent or mixed intervention. A majority of the exercise interventions, or components, were supervised (for >two sessions, not including initial instruction) and of long duration (>3 months). The supervised, long-duration balance/resistance (group or individual) and tai chi (group) interventions had at least low certainty evidence for benefit across more outcomes than the other exercise and single-component interventions (Table S2.3). Both of the long-duration tai chi interventions had some moderate certainty evidence, though there was more evidence of benefit across more outcomes (and more outcomes meeting the larger effect threshold) when delivered in a group setting. The one short-duration tai chi node had low certainty evidence across outcomes (Table S2.4). Six of nine balance/resistance interventions had some moderate certainty of benefit; the three with low certainty generally reported few outcomes and had smaller samples. None of the interventions with moderate certainty evidence focused on walking.
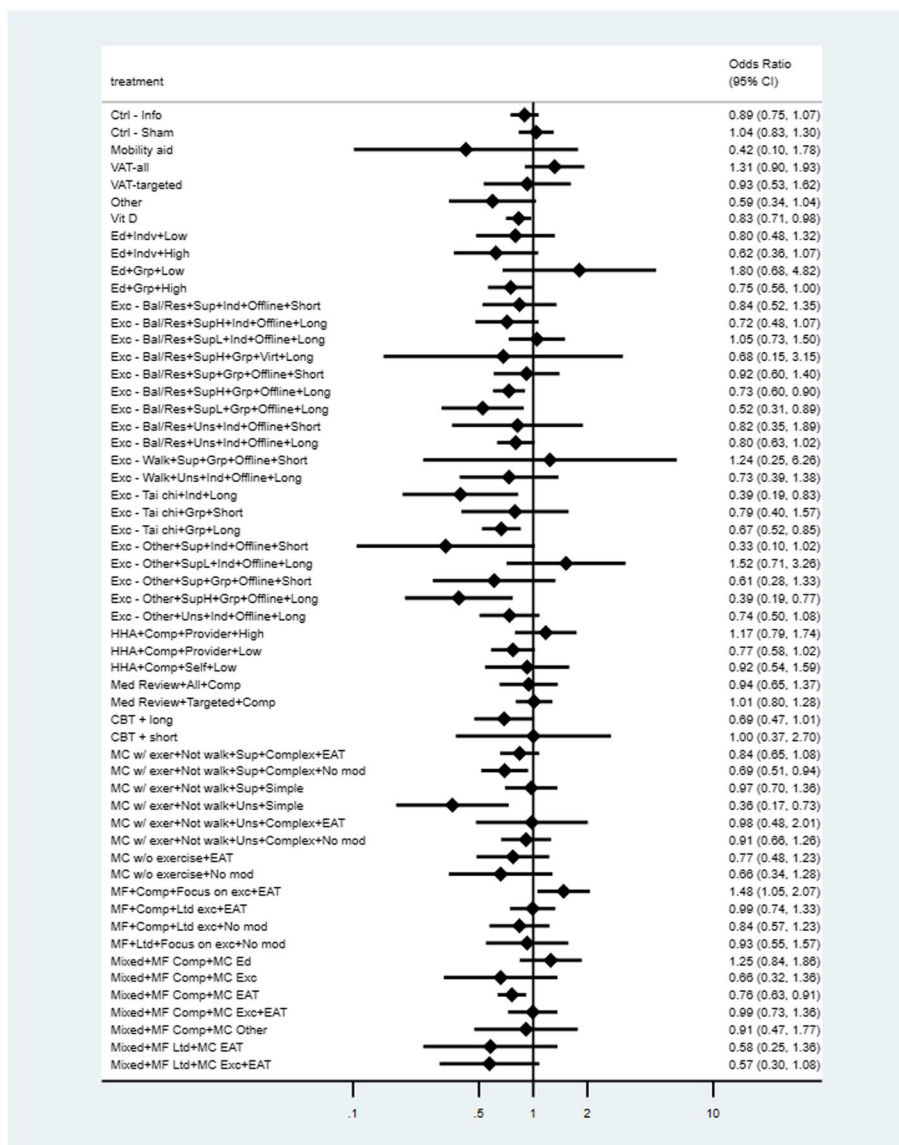
Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 19 of 32

treatment		Odds Ratio (95% CI)
Ctrl - Info		0.89 (0.75, 1.07)
Ctrl - Sham		1.04 (0.83, 1.30)
Mobility aid		0.42 (0.10, 1.78)
VAT-all		1.31 (0.90, 1.93)
VAT-targeted		0.93 (0.53, 1.62)
Other		0.59 (0.34, 1.04)
Vit D		0.83 (0.71, 0.98)
Ed+Indv+Low		0.80 (0.48, 1.32)
Ed+Indv+High		0.62 (0.36, 1.07)
Ed+Grp+Low		1.80 (0.68, 4.82)
Ed+Grp+High		0.75 (0.56, 1.00)
Exc - Bal/Res+Sup+Ind+Offline+Short		0.84 (0.52, 1.35)
Exc - Bal/Res+SupH+Ind+Offline+Long		0.72 (0.48, 1.07)
Exc - Bal/Res+SupL+Ind+Offline+Long		1.05 (0.73, 1.50)
Exc - Bal/Res+SupH+Grp+Virt+Long		0.68 (0.15, 3.15)
Exc - Bal/Res+Sup+Grp+Offline+Short		0.92 (0.60, 1.40)
Exc - Bal/Res+SupH+Grp+Offline+Long		0.73 (0.60, 0.90)
Exc - Bal/Res+SupL+Grp+Offline+Long		0.52 (0.31, 0.89)
Exc - Bal/Res+Uns+Ind+Offline+Short		0.82 (0.35, 1.89)
Exc - Bal/Res+Uns+Ind+Offline+Long		0.80 (0.63, 1.02)
Exc - Walk+Sup+Grp+Offline+Short		1.24 (0.25, 6.26)
Exc - Walk+Uns+Ind+Offline+Long		0.73 (0.39, 1.38)
Exc - Tai chi+Ind+Long		0.39 (0.19, 0.83)
Exc - Tai chi+Grp+Short		0.79 (0.40, 1.57)
Exc - Tai chi+Grp+Long		0.67 (0.52, 0.85)
Exc - Other+Sup+Ind+Offline+Short		0.33 (0.10, 1.02)
Exc - Other+SupL+Ind+Offline+Long		1.52 (0.71, 3.26)
Exc - Other+Sup+Grp+Offline+Short		0.61 (0.28, 1.33)
Exc - Other+SupH+Grp+Offline+Long		0.39 (0.19, 0.77)
Exc - Other+Uns+Ind+Offline+Long		0.74 (0.50, 1.08)
HHA+Comp+Provider+High		1.17 (0.79, 1.74)
HHA+Comp+Provider+Low		0.77 (0.58, 1.02)
HHA+Comp+Self+Low		0.92 (0.54, 1.59)
Med Review+All+Comp		0.94 (0.65, 1.37)
Med Review+Targeted+Comp		1.01 (0.80, 1.28)
CBT + long		0.69 (0.47, 1.01)
CBT + short		1.00 (0.37, 2.70)
MC w/ exer+Not walk+Sup+Complex+EAT		0.84 (0.65, 1.08)
MC w/ exer+Not walk+Sup+Complex+No mod		0.69 (0.51, 0.94)
MC w/ exer+Not walk+Sup+Simple		0.97 (0.70, 1.36)
MC w/ exer+Not walk+Uns+Simple		0.36 (0.17, 0.73)
MC w/ exer+Not walk+Uns+Complex+EAT		0.98 (0.48, 2.01)
MC w/ exer+Not walk+Uns+Complex+No mod		0.91 (0.66, 1.26)
MC w/o exercise+EAT		0.77 (0.48, 1.23)
MC w/o exercise+No mod		0.66 (0.34, 1.28)
MF+Comp+Focus on exc+EAT		1.48 (1.05, 2.07)
MF+Comp+Ltd exc+EAT		0.99 (0.74, 1.33)
MF+Comp+Ltd exc+No mod		0.84 (0.57, 1.23)
MF+Ltd+Focus on exc+No mod		0.93 (0.55, 1.57)
Mixed+MF Comp+MC Ed		1.25 (0.84, 1.86)
Mixed+MF Comp+MC Exc		0.66 (0.32, 1.36)
Mixed+MF Comp+MC EAT		0.76 (0.63, 0.91)
Mixed+MF Comp+MC Exc+EAT		0.99 (0.73, 1.36)
Mixed+MF Comp+MC Other		0.91 (0.47, 1.77)
Mixed+MF Ltd+MC EAT		0.58 (0.25, 1.36)
Mixed+MF Ltd+MC Exc+EAT		0.57 (0.30, 1.08)

Fig. 2 Forest plot of network estimates for fallers outcome for all interventions versus usual care

Whole-body vibration, which involves patients standing on a platform emitting low magnitude vibrations for a specified amount of time (e.g., 20 min/day, 5 days/week for 18 months [62]), is probably beneficial for reducing fallers. Moderate certainty for some benefit was also found for the two interventions offering high-intensity/dose (> 2 sessions/4 h) education (in groups or individually) but not for the two offering low-intensity education. Likewise, information and attention controls did not have any moderate certainty evidence for benefit (Table S2.4). One of three single-component HHA interventions showed moderate certainty of benefit. All of the HHA interventions used comprehensive assessments covering

multiple risk factors; the node with multiple home visits and study provision of modifications had low certainty evidence (Table S2.4 node 33). Despite contribution from a large negative trial [63], the mixed intervention combining comprehensive multifactorial assessment plus HHA showed moderate certainty of benefit, as did two other mixed interventions with education or HHA and exercise provided to all. The other four mixed interventions showed beneficial effects with low certainty for two or more benefit outcomes. CBT of long duration showed moderate certainty for reducing fallers; these interventions focused on fear of falling and usually contained multiple sessions (e.g., 7 to 8 weekly sessions with

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 20 of 32

Table 4 Summary of findings across all outcomes for interventions with moderate certainty evidence for some benefit (KQ1)

Node	Intervention	Fallers	Falls	Injurious fallers	Fractures (any)	Hip fractures	Long-term care admissions	Functional status	Health-related quality of life	Any adverse event	Serious adverse events*	Fractures from intervention	Falls from intervention
		CINeMA Certainty # prevented per 100 people treated (95% CI)						*GRADE Certainty* Mean difference (95% CI)		*Grade Certainty* # of events per 100 people (95% CI)			
7	WBV	⊕⊕⊕⊖ d, e 12 (22, -1)	⊕⊕⊕⊖ D, E 54 (82, 0)	No evidence	⊕⊕⊕⊖ D 3 (4, -4)	No evidence	No evidence	No evidence	No evidence	⊕⊕⊕⊖ d 4.7 (2.8 to 7.3)	⊕⊕⊕⊖ d 0 (0 to 0.9)	⊕⊕⊕⊖ d 0 (0 to 0.9)	No evidence
10	Education – I D^H	⊕⊕⊕⊖ d, e 11 (21, -2)	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ D -8 (-28, 7)	No evidence	No evidence	⊕⊕⊕⊖ D 3 (-14, 4)	No evidence	No evidence	⊕⊕⊕⊖ a, d 0 (0 to 1.4)		⊕⊕⊖⊖ a, D	No evidence
12	Education – G D^H	⊕⊕⊕⊖ d, e 7 (13, 0)	⊕⊕⊕⊖ D 32 (57, -2)	⊕⊕⊕⊖ D LTND	⊕⊕⊖⊖ D, f	⊕⊕⊖⊖ D, f	No evidence	No evidence	⊕⊕⊕⊖ d, E LTND	⊕⊕⊕⊖ b, d 2.1 (0.7 to 4.8)	⊕⊕⊖⊖ b, D	⊕⊕⊖⊖ b, D	⊕⊕⊖⊖ a, b, D, e
13	Balance/Resistance – SIB	⊕⊕⊕⊖ D, e LTND	⊕⊕⊕⊖ d, e 50 (75, 9)	⊕⊕⊕⊖ b, d, E 9 (-3, 16)	No evidence	No evidence	No evidence	⊕⊕⊖⊖ D, E LTND	⊕⊕⊕⊖ d, E 6 (-2, 14)	⊕⊕⊕⊖ e 7.5 (5.2 to 10.4)	⊕⊕⊕⊖ d 0 (0 to 0.9)	⊕⊕⊕⊖ d 0 (0 to 0.9)	⊕⊕⊕⊖ b, d 0.9 (0.1 to 2.7)
15	Balance/Resistance - S^HIL	⊕⊕⊕⊖ E 8 (16, -2)	⊕⊕⊕⊖ d, e 47 (64, 25)	⊕⊕⊕⊖ D, e -13 (-34, 4)	⊕⊕⊕⊖ D, f 1 (3, -1)	⊕⊕⊕⊖ D, f 0.5 (-14.5, 1)	⊕⊕⊕⊖ e -1 (-13, 3)	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ e 0.5 (0.2 to 0.8)	⊕⊕⊕⊖ e 0 (0 to 0.2)	⊕⊕⊕⊖ e 0 (0 to 0.2)	⊕⊕⊕⊖ e 0 (0 to 0.1)
19	Balance/Resistance - S^HGL	⊕⊕⊕⊖ d 7 (12, 3)	⊕⊕⊕⊖ E 31 (46, 14)	⊕⊕⊕⊖ c, d, E 7 (2, 10)	⊕⊕⊕⊖ D, f 1 (3, -1)	⊕⊕⊕⊖ D, f 0.3 (-17.5, 1)	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ E LTND	⊕⊕⊕⊖ b 3.4 (2.2 to 4.8)	⊕⊕⊕⊖ b, d 0 (0 to 0.4)	⊕⊕⊕⊖ b, d 0 (0 to 0.7)	⊕⊕⊕⊖ b 0.8 (0.3 to 1.8)
20	Balance/Resistance - S^LGL	⊕⊕⊕⊖ D 14 (23, 3)	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ D 4 (-10, 13)	⊕⊕⊕⊖ E 2 (5, -12)	No evidence	No evidence	⊕⊕⊕⊖ d, e 13 (6, 20)	⊕⊕⊕⊖ E 31 (23, 38)	⊕⊕⊖⊖ a, b, e	⊕⊕⊖⊖ a, b, D, e	⊕⊕⊖⊖ a, b, D, e	No evidence
21	Balance/Resistance – UIB	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ c 20 (11, 22)	⊕⊕⊖⊖ c, D	No evidence	No evidence	⊕⊕⊕⊖ D, e LTND	⊕⊕⊕⊖ E, f LTND	⊕⊕⊕⊖ e 4.7 (1.5 to 10.6)	⊕⊕⊕⊖ D 0 (0 to 3.4)	⊕⊕⊕⊖ D 0 (0 to 3.4)	⊕⊕⊖⊖ b, c, D, e
22	Balance/Resistance – UIL	⊕⊕⊕⊖ d, e 5 (11, 0)	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ D LTND	⊕⊕⊖⊖ D, f	⊕⊕⊖⊖ e 0.6 (-7.4, 1)	No evidence	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ E LTND	⊕⊕⊕⊖ e 2.7 (1.9 to 3.6)	⊕⊕⊕⊖ e 0.2 (0 to 0.5)	⊕⊕⊕⊖ e 0 (0 to 0.2)	⊕⊕⊕⊖ b 0.1 (0 to 0.5)
25	Tai Chi – SIL	⊕⊕⊕⊖ c, d 20 (29, 4)	⊕⊕⊕⊖ e 54 (82, 5)	⊕⊕⊕⊖ D -5 (-29, 10)	No evidence	No evidence	No evidence	⊕⊕⊖⊖ D, f LTND	⊕⊕⊖⊖ D LTND	⊕⊕⊕⊖ a, c, d, e	⊕⊕⊖⊖ a, c, D, e	⊕⊕⊖⊖ a, c, D e	No evidence
27	Tai Chi – SGL	⊕⊕⊕⊖ D 9 (14, 4)	⊕⊕⊕⊖ e 46 (61, 28)	⊕⊕⊕⊖ D, e 10 (2, 15)	⊕⊕⊕⊖ D, f 2 (5, -58)	No evidence	⊕⊕⊖⊖ e 3 (-23, 4)	⊕⊕⊖⊖ D, f	⊕⊕⊕⊖ d, E LTND	⊕⊕⊕⊖ b 6.8 (5.5 to 8.3)	⊕⊕⊕⊖ a, b 0.5 (0.2 to 1)	⊕⊕⊕⊖ b 0 (0 to 0.3)	⊕⊕⊕⊖ a, b 0 (0 to 0.7)
28	Other exercise – SIB	⊕⊕⊕⊖ e 22 (35, 0)	No evidence	⊕⊕⊕⊖ d, e 18 (4, 22)	No evidence	No evidence	No evidence	⊕⊕⊖⊖ D, f	No evidence	⊕⊕⊕⊖ d, e 0 (0 to 5.9)	⊕⊕⊕⊖ d	⊕⊕⊕⊖ D, e	⊕⊕⊕⊖ D, e
31	Other exercise - SGL	⊕⊕⊕⊖ b, d 20 (29, 6)	⊕⊕⊖⊖ e 25 (61, -33)	No evidence	No evidence	No evidence	⊕⊕⊕⊖ e 3 (-8, 4)	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ E LTND	⊕⊕⊕⊖ d 0 (0 to 1)	⊕⊕⊕⊖ d 0 (0 to 1)	⊕⊕⊕⊖ d 0 (0 to 1)	⊕⊕⊕⊕ 0 (0 to 1.2)
32	Other exercise – UIL	⊕⊕⊕⊖ d, e 7 (15, -2)	⊕⊕⊕⊖ D, e 22 (51, -21)	⊕⊕⊕⊖ D LTND	No evidence	No evidence	⊕⊕⊕⊖ e -1 (-11, 3)	⊕⊕⊕⊖ D, e, f LTND	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ d 0 (0 to 2.9)	⊕⊕⊕⊖ b, d	⊕⊕⊕⊖ b, D	⊕⊕⊕⊖ b, d 0 (0 to 4.6)
34	HHA – D^I M^P	⊕⊕⊕⊖ d, e 6 (12, 0)	⊕⊕⊕⊖ D, E 37 (55, 13)	⊕⊕⊕⊖ b, D, e -6 (-18, 4)	No evidence	No evidence	⊕⊕⊕⊖ e 2 (-1, 3)	⊕⊕⊕⊖ d, E LTND	⊕⊕⊕⊖ E LTND	⊕⊕⊖⊖ b, d, e	⊕⊕⊖⊖ b, D, e	⊕⊕⊖⊖ b, D, e	No evidence
38	CBT – L	⊕⊕⊕⊖ d, e 9 (16, 0)	⊕⊕⊕⊖ e 17 (47, -27)	⊕⊕⊕⊖ D LTND	⊕⊕⊕⊖ D, f 4 (5, -12)	No evidence	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ D, E LTND	No evidence	No evidence	No evidence	No evidence	No evidence
42	Exercise^S + Other (not HHA or education)	⊕⊕⊕⊖ e 9 (15, 1)	⊕⊕⊕⊖ E 45 (65, 20)	⊕⊕⊕⊖ D 5 (-2, 10)	⊕⊕⊕⊖ D 1 (3, -5)	No evidence	⊕⊕⊕⊖ e 3 (-1, 4)	No evidence	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ a, b 7.8 (5.5 to 10.7)	⊕⊕⊕⊖ b, d 0 (0 to 0.8)	⊕⊕⊕⊖ b, d 0.2 (0 to 1.2)	⊕⊕⊖⊖ a, b, d, e
44	Exercise^U + Education	⊕⊕⊕⊖ a, e 21 (30, 7)	⊕⊕⊕⊖ a, d, E 47 (77, -9)	No evidence	No evidence	No evidence	⊕⊕⊕⊖ a, D 2 (-34, 4)	⊕⊕⊖⊖ a, d, E, f	⊕⊕⊕⊖ a, E 29 (15, 44)	No evidence	No evidence	No evidence	No evidence
54	Mixed: Education + MF^Comp	⊕⊕⊖⊖ C, d, E -5 (4, -15)	⊕⊕⊕⊖ e 17 (51, -35)	⊕⊕⊕⊕ 15 (9, 19)	⊕⊕⊕⊕ 3 (4, 1)	⊕⊕⊖⊖ a, c, D	No evidence	No evidence	⊕⊕⊕⊖ a, c, D, E	⊕⊕⊖⊖ a, b, c, D, e	⊕⊕⊖⊖ a, b, c, D, e	⊕⊕⊖⊖ a, b, c, D, e	No evidence
56	Mixed: HHA + MF^Comp	⊕⊕⊕⊖ d, e 6 (11, 2)	⊕⊕⊕⊖ c, E 35 (47, 21)	⊕⊕⊕⊖ d, e 4 (0, 8)	⊕⊕⊕⊖ d, e 1 (2, -1)	⊕⊕⊕⊖ e 0.5 (-4.5, 0.9)	⊕⊕⊕⊖ e -1 (-6, 1)	⊕⊕⊕⊖ d, E 6 (2, 11)	⊕⊕⊕⊖ D, E LTND	⊕⊕⊕⊖ b, e 0.3 (0.2 to 0.5)	⊕⊕⊕⊖ b, e 0 (0 to 0.1)	⊕⊕⊕⊖ b, e 0 (0 to 0.1)	⊕⊕⊕⊖ b, e 0 (0 to 0.2)
60	Mixed: Exercise + HHA + MF^Ltd	⊕⊕⊕⊖ d, e 13 (24, -2)	⊕⊕⊕⊖ d 94 (114, -422)	⊕⊕⊕⊖ c, D, e 10 (-4, 17)	No evidence	No evidence	No evidence	No evidence	No evidence	No evidence	No evidence	No evidence	No evidence

Interventions legend: *B* brief (<14 week duration), *CBT* cognitive behavioral therapy, *D^H* high dose (≥2 sessions or visits), *D^L* low dose (≤2 sessions or visits), *G* Group, *HHA* home hazard assessment, *I* individual, *L* long (≥3 months or longer in duration), *M^P* home modifications initiated by intervention provider, *M^S* home modifications initiated by "self" (participant), *MF* multifactorial, *S^H* high-intensity supervision (>1 contacts per week on average), *S^L* low-intensity supervision (≤ contacts per week on average), *T* targeted, *U* universal, *VAT* vision assessment and treatment, *WBV* whole body vibration

Certainty legend: A = within study bias; B = across study bias; C = indirectness; D = imprecision; E = heterogeneity; F = incoherence. Lower case superscript letters indicate downrating by one level (0.5 step); uppercase indicates downrating a full level or more (red text signifying rating down 2 levels). ⊕⊕⊕⊕ = high certainty; ⊕⊕⊕⊖ = moderate certainty; ⊕⊕⊖⊖ = low certainty; ⊕⊖⊖⊖ = very low certainty

Explanation: Negative values indicate an increase (rather than reduction) in risk of outcome. Green shading indicates desirable effect, red shading indicates undesirable effect; no shading (white) indicates little-to-no difference (LTND). Intensity of shading reflects the magnitude of effect: light shading indicates that the certainty rating relates to the lower effect threshold (i.e., the effects meet or exceed this threshold) and darker shading reflects it relates to the moderate effect threshold (see text for absolute effects related to these thresholds for each outcome). Outcomes with very low certainty or no evidence are shaded in grey. To aid with interpretation, effect estimates showing LTND or with very low certainty are not reported here, but can be found in Supplementary file 2 evidence profiles

booster(s)), assigned skills-based homework, and had exercise demonstrations/instruction.

No multifactorial intervention, where intervention components were only delivered based on individual needs assessments, showed moderate certainty of benefit; among five interventions, three had low certainty for some benefit (*n*=1) or little-to-no difference (*n*=2) (Table S2.4), while one had very low certainty across outcomes, and another had moderate certainty of some harm (Table S2.5). Of nine multicomponent interventions, we only rated the certainty of evidence as moderate for some benefit for two, both including exercise.

Among those interventions only showing low or very low certainty evidence (*n*=35; Table S2.4), 27 had low certainty for at least a small benefit in at least one outcome. The following interventions offered low-certainty evidence for little-to-difference or some harm across all outcomes:

- Vision assessment and treatment (VAT)—targeted (node 6)
- Walking—supervised, group, and brief (node 23)
- Other exercise—supervised, individual, and long (node 29)
- CBT—brief (node 29)
- Multifactorial (limited assessment)—focus on exercise (node 52)

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 21 of 32

- Multifactorial (limited assessment)—focus on HHA (node 53)

For two interventions, there was limited evidence of any benefit but moderate certainty of harmful effect for one or more benefit outcomes (Table S2.5). VAT (universal, i.e., including a screening component and not limiting recruitment to those with visual acuity issues) had moderate certainty evidence that it increased the number of fallers and fractures [64, 65]. One comprehensive multifactorial intervention, offering exercise and HHA when indicated, had moderate certainty for increasing fallers. The four studies contributing direct evidence to this node targeted people with a high risk for falls, including the elderly (mean ages 83 [recruited from hospital] [66] and 85 years [67]) or those with previous falls [68] and in one also taking "fall-risk-increasing drugs" [69]. Both the VAT and multifactorial interventions also had low certainty for harm for falls and at least one fracture outcome.

Evidence on AEs was largely limited to studies reporting on exercise interventions or vitamin D. Most single-component exercise interventions are probably associated with minimal harm; across all exercise interventions in the review, two interventions (nodes 27 and 13; Table 4) had moderate certainty for small harms (possibly 5–8 per 100 people) from any AE, whereas there was low or moderate certainty for little-to-no harm for the more serious outcomes across multiple nodes. For vitamin D, there was moderate certainty for little-to-no difference for any AE and serious AEs (Table S2.4). Uncertainty exists about possible harm from other interventions. The proportion of nodes rated down for missing outcome data was higher than for the benefit outcomes (Table S2.4).

There were scarce data available for considering populations that may require equity or other contextual considerations. Nine trials, eight from the USA, exclusively enrolled participants experiencing low income or other factors related to low socioeconomic status. Two of the trials had UC control groups, enabling some indirect observations with the network estimates. One study (*n*=310) investigated low-contact HHA (node 34) and found a benefit for falls similar to our network estimate (Table 4) [70]. The other trial (*n*=1848) examined higher-contact HHA (node 33), reporting little-to-no difference in fallers, which was in line with the network estimate (Table S2.4) [71]. Neither study contributed a large majority of the sample in their respective node. None of the trials in the review exclusively recruited, or reported on subgroup effects for, Indigenous peoples, newcomers, or Black Canadians.

Outcome valuation (KQ2)
Study selection and characteristics
Our database searches for KQs 2 and 3 (*n*=9025) and other sources for KQ2 (*n*=17) identified 9042 citations; 797 full texts were retrieved, resulting in the inclusion of 44 studies described in 49 reports (Figure S3.1). Reports of four studies in three papers [72–74] in our initial search met the eligibility criteria but were excluded during our search update because they contained populations that were either the same or substantially overlapping with newer reports. Supplementary file 3 has the characteristics by study, including risk-of-bias ratings (Table S3.1), and lists of included and excluded studies.

The mean age across the studies was 79.6 years (range 69.4 to 85.7), with 59% having a mean age ≤ 80 years, and the average proportion of females was 74% (range 49 to 100%). Of 18 studies reporting on fracture history (i.e., before the one(s) assessed in the study, when applicable), the average proportion of participants with a prior fracture was 34%. Median sample size was 217 (*IQR* 140 to 396; range 53 to 6548). Of 40 studies from single countries, 14 came from Sweden (6 studies used 3 similar but not substantially overlapping populations); 4 each came from Australia, Germany, and Spain; 3 from Canada; 2 from the Netherlands and UK; and 1 each from Estonia, Iran, Ireland, Mexico, Portugal, the USA, and Turkey. There were four studies using data from multiple countries, three using similar but not substantially overlapping populations or reporting different data (e.g., timing after hip fracture) from the same sample. Overall, 5 (11%), 20 (46%), and 19 studies (43%) were rated at low, moderate, and high risk of bias, respectively. The main reasons for a rating of high risk of bias were incomplete data and/or inadequate use of the measurement instrument (e.g., use of a tariff for the HSUV measurement tool that did not come from the same country as the participants).

Thirty-four studies measured HSUVs using the EQ-5D index measure, 1 used the EQ-5D visual analogue scale, 2 used time trade-offs, and 2 used the Health Utilities Index Mark II (1 also used SF-6D and a standard gamble). In all studies measuring HSUVs but one which used a time trade-off [75], participants had experienced the event(s) being assessed rather than basing their answers on hypothetical scenarios. Five studies did not measure HSUVs but measured preferences in discrete choice (*n*=4) or another conjoint experiment using hypothetical scenarios.

Neither industry funding nor missing outcome data were found to be a serious concern for any outcome.

Findings
Table 5 contains the summary of findings for all outcomes including the results of the meta-analyses and our

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 22 of 32

Table 5 Summary of findings for outcome valuation (KQ2)

Number of included studies; sample size	Disutility (95% CI) 0 (no decrease in HRQoL) to 1 (HRQoL equal to death)	GRADE[1]	What does the evidence say?
Disutility from hip fracture			
16 studies; n = 7409	**EQ-5D, at time of fracture (< 3 months)** 0.53 (0.44 to 0.62)	⊕⊕⊕⊖ Moderate[a]	The disutility of a hip fracture is probably 0.53 immediately after injury
27 studies; n = 9399	**EQ-5D, 12-month post-fracture** 0.16 (0.12 to 0.20)	⊕⊕⊕⊕ High	The disutility of a hip fracture is 0.16 at 12 months after injury
Disutility from non-hip* fracture			
4 studies; n = 1792	**EQ-5D, at time of fracture (< 3 months)** 0.57 (0.43 to 0.71)	⊕⊕⊖⊖ Low[ab]	The disutility of a non-hip fracture may be 0.57 immediately after injury
4 studies; n = 1792	**EQ-5D, 12-month post-fracture** 0.19 (0.10 to 0.28)	⊕⊕⊕⊖ Moderate[b]	The disutility of a non-hip fracture is probably 0.19 at 12 months after injury
Disutility from any injurious fall			
0 study	**No evidence**	Not applicable	The disutility of an injurious fall is uncertain
Disutility from a fall (within last 12 months)			
6 studies; n = 4653	**EQ-5D** 0.09 (− 0.04 to 0.22)	⊕⊕⊖⊖ Low[A]	The disutility after a fall may be 0.09
Disutility from functional impairment (impairment in at least one ADL)**			
1 study; n = 123	**HUI Mark II** 0.12 (0.05 to 0.19)	⊕⊕⊖⊖ Low[cd]	The disutility from impairment in one or more ADLs may be 0.12
Disutility of LTC admission (compared to full health)			
1 study; n = 194	**TTO** Median (IQR): 1 (1, 1) "80% of participants said they would rather be dead"	⊕⊕⊖⊖ Low[cd]	The disutility from a long-term care admission (compared to full health) may be 1
Relative importance across health states			
EQ-5D unless otherwise specified Disutility, LTC admission: 1 (TTO) Disutility, non-hip fracture (< 3 mos): 0.57 Disutility, non-hip fracture (12 mos): 0.19 Disutility, hip fracture (< 3 mos): 0.53 Disutility, hip fracture (12 mos): 0.16 Disutility, ADL impairment: 0.12 (HUI Mark II) Disutility, fall: 0.09 Disutility, injurious fall: unknown Also see below rows for findings from other preference-based studies, used for comparison		⊕⊕⊖⊖ Low[e,f]	LTC admission may be more important than all other outcomes
		⊕⊕⊕⊖ Moderate[a]	Fracture (hip or non-hip) is probably more important than falls and functional impairment
		⊕⊕⊖⊖ Low[e,f]	Functional impairment may be somewhat more important than a fall

Findings of relative importance between health states

50% decrease in fracture risk:50% decrease in fall risk: ratio, SMD of coefficients: 2.43 (Milette 2013, Franco 2015; DCE)
50% improvement in daily functioning:50% decrease in fall risk: ratio, SMD of coefficients: 2.11 (Franco 2015; DCE)
Ability to manage domestic activities:HRQoL: ratio, relative importance score: 1.58 (Hilingsman 2020; CA)
Ability to manage domestic activities:fall frequency: ratio, relative importance score: 1.20 (Hilingsman 2020; CA)
LTC admission:falls risk: ratio, relative importance score: 1.18 (Robinson 2015; CA)
50% decrease in fracture risk:50% improvement in daily functioning (Milette 2013, Franco 2015; DCE): ratio, SMD of coefficients: 1.14

Data from other utility instruments, disutilities

Hip fracture; disutility at time closest to injury
Average of HUI Mark II, SF-6D, SG, & FT
1 study; n = 80
0.25 (0.20 to 0.30)
Hip fracture; disutility at time closest to 12 mos
TTO
2 studies; n = 471
0.57 (0.32 to 0.82)
Average of HUI Mark II, SF-6D, SG, & FT
1 study; n = 80
0.12 (0.07 to 0.17)
Falls: disutility any time after event
TTO
1 study; n = 203
0.33 (0.26 to 0.40)

Abbreviations: *ADL* activity of daily living, *CA* conjoint analysis, *DCE* discrete choice experiment, *FT* feelings thermometer, *SG* standard gamble, *TTO* time trade-off.
*Non-hip fracture refers to any fracture other than hip fracture; studies contributing to this health state reported on distal forearm, humerus, and clinical vertebral fractures (weighted average used for analysis). **This study measured utility using HUI Mark II in individuals with and without impairment in ≥ 1 ADL; participants were not asked how long they had been ADL impaired. *Explanation of GRADE ratings*: Lowercase superscript indicates rating down once, and an uppercase superscript indicates rating down twice. A, inconsistency in estimates across multiple studies. B, > 50% of sample size is from studies rated at high risk of bias. C, concerns about indirectness due to mean age > 80 years in the single study. D, concerns about lack of consistency due to evidence from single study. E, concerns about indirectness due to outcome measure other than EQ-5D. F, concerns about inconsistency due to lack of agreement between utility ranking and relative importance studies

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 23 of 32

certainty assessments. Supplementary file 3 contains the forest plots (Figures S3.2 and S3.3) and tables with findings from subgroup analyses (Tables S3.2 and S3.3).

Admission to LTC had the highest disutility (median=1 [equivalent to death]), indicating that it is the most important health state considered in this review. However, this health state utility was measured using time trade-off in a single study ($n=194$) in a sample of adults ≥ 75 years old (mean age 83 years), and the evidence was rated as low certainty. Both hip (moderate certainty) and non-hip (e.g., means across distal forearm, humerus, and clinical vertebral fractures; low certainty) fracture may result in substantial disutility (0.53 and 0.57) in the first 3 months after injury. Disutilities for both hip and non-hip fractures are probably lower at 12 months after injury (0.16 and 0.19, with high and moderate certainty, respectively) compared to within the first 3 months. No study measured the disutility of an injurious fall. Based on the much higher disutility, fracture (of any type) is probably more important than either falls (0.09 over 12 months) or functional status (0.12 for impairment in at least 1 ADL). Functional status may be somewhat more important than falls; this conclusion is also supported by findings from other HSUV measurement tools and utility-based experiments (Table 5).

Examination of subgroups by age (mean age < 80 years vs. ≥ 80 years), sex (< 80% vs. ≥ 80% female), timing of measurement (within 1 week vs. 1 week to 3 months after event), and proportion with previous fracture (30% vs. 30–99% vs. unknown) did not explain heterogeneity in either hip or non-hip fracture disutilities within the first 3 months from injury. Stratifying by whether the fracture was attributed to a fall led to a significant interaction ($p<0.0001$) and marginally reduced heterogeneity for attributed non-hip fracture, though the disutility among studies where the attribution was present was only slightly lower than the estimate across all studies (0.52 vs. 0.57), and this result did not change our conclusions.

At 12 months after injury, subgroup analyses by age and sex did not explain heterogeneity in either hip or non-hip fracture. For non-hip fractures, there was a significant difference ($p<0.0001$) in disutility between studies that attributed the fracture to a fall (0.22) and those that did not (0.09, one study), though the finding when there was attribution to a fall was close to our overall estimate (0.19). For hip fractures, there was a significant difference ($p=0.03$) between studies with > 30% previous fracture (0.12) versus 30–99% previous fracture (0.18) and previous fracture unknown (0.18); again, the overall estimate (0.16) was close to these subgroup values, and findings did not change our conclusions. Subgroup analyses by age, sex, and previous fracture did not reveal any subgroup differences for disutility of a fall within

the last 12 months. The large degree of variability due to between-study heterogeneity ($I^2>90\%$) in our analyses was likely due to precise study findings and was not considered serious because most findings agreed well with our pooled estimates and conclusions.

When examining data related to populations that may require equity or other contextual considerations, none of the studies reported relevant subgroup analysis; in three studies [76–78] where a large proportion (>60%) of participants reported low income, the disutilities of hip fracture were quite similar to the average across all studies. None of the trials exclusively recruited, or reported on subgroup effects, for Indigenous peoples, newcomers, or Black Canadians.

Preferences for interventions (KQ3)
Study selection and characteristics
Our database searches for KQs 2 and 3 ($n=9025$) and other sources for KQ3 ($n=154$) identified 9179 citations to screen, with 354 being retrieved for full-text screening and inclusion of 29 studies described in 30 reports (Figure S3.4). Three studies were also included in either KQ1 [79, 80] or KQ2 [27]. Supplementary file 3 has the characteristics and risk-of-bias assessments by study (Tables S3.4 and S3.5) and lists of included and excluded studies. The average mean age across studies was 76.6 years, and proportion of females was 67%. Of 18 studies reporting on fall history, 4 enrolled samples having > 50% fallers in past year, and 3 exclusively enrolled previous fallers. Of 20 studies reporting quantitative data, the median sample size was 177 (*IQR* 96 to 270); the median size of 9 studies reporting qualitative data was 69 (*IQR* 19 to 187). Multiple studies came from Australia ($n=7$), the USA ($n=7$), the UK ($n=4$), Canada ($n=2$), and the Netherlands ($n=2$); single studies came from each of Denmark, Germany, Sweden, Belgium, Malaysia, India, and one study was carried out across multiple European countries. Of the quantitative studies, 9 (45%) were rated at high, 11 (55%) at moderate, and none at low risk of bias. Most ratings of high risk came from sampling strategies judged as possibly or likely biased in the selection of participants and/or low response rates. For one (11%), five (56%), and three (33%) of the qualitative studies, we had significant, some, and few concerns about the quality. The significant concerns in one study came from unclear theoretical underpinnings and lack of specifying the qualitative approach.

Most (16/20) of the quantitative studies employed a survey approach with questions about attitudes (e.g., using Likert scales) or preferences/choices for different programs or their attributes; in four of these studies, some of the participants had experienced the program(s), and in another four, participants had been

Table 6 Summary of findings on preferences for interventions and their attributes (KQ3)

	CBT (group)	CBT (indiv)	CBT (NS)	Education (Group)	Education (Indiv)	Education (NS)	Exercise (B&R, Group)	Exercise (B&R, Indiv)	Exercise (B&R, NS)	Exercise (NS, Group)	Exercise (NS, Indiv)	Exercise (Other, NS)	Exercise (Other, Group)	Exercise (Tai Chi, Group)
CBT (group)	■	•												
CBT (indiv)	⊕⊕⊖⊖ Low ab	■												
CBT (NS)			■			▲				-				
Education (Group)				■	◀						-			
Education (Indiv)				⊕⊕⊖⊖ Low bc	■		-							
Education (NS)			⊕⊕⊖⊖ Low bc			■		▲▲				-		-
Exercise (B&R, Group)					⊕⊖⊖⊖ Very Low bcd		■	▲▲					◀◀	◀
Exercise (B&R, Indiv)		⊕⊖⊖⊖ Very Low abc				⊕⊕⊕⊕ High		■					◀◀◀	◀◀
Exercise (B&R, NS)						⊕⊕⊖⊖ Low cd			■			-		
Exercise (NS, Group)				⊕⊖⊖⊖ Very Low bcd						■	•			
Exercise (NS, Indiv)									⊕⊕⊖⊖ Low bd		■			
Exercise (Other, NS)						⊕⊖⊖⊖ Very Low bcd			⊕⊖⊖⊖ Very Low bcd			■		
Exercise (Other, Group)							⊕⊕⊖⊖ Low bc	⊕⊕⊖⊖ Low bc					■	•
Exercise (Tai Chi, Group)						⊕⊖⊖⊖ Very Low bcd	⊕⊕⊖⊖ Low bc	⊕⊕⊖⊖ Low bc					⊕⊕⊖⊖ Low bc	■

Abbreviations: *B&R* balance and resistance (exercise), *CBT* cognitive behavioral therapy, *Indiv* individual, *NS* not specified. Explanation for preference findings: ▲▲▲ strong preference to intervention in direction of arrow, ▲▲ clear preference (to intervention in direction of arrow), ▲ slight preference (to intervention in direction of arrow), ● variable preferences, orange symbols indicate evidence from qualitative studies. Cell shading reflects certainty in evidence: green, high; blue, moderate; yellow, low; gray, very low; white, no evidence. Explanation of GRADE ratings: lowercase superscript indicates rating down once, and an uppercase superscript indicates rating down twice for that reason/domain. A, > 50% of sample size from studies rated at high risk of bias. B, concerns due to inconsistency in direction of preference across studies. C, concerns of imprecision due to sample size < 380. D, indirectness due to lack of specificity of proposed intervention (e.g., not specified whether exercise program is group or individual) or because of the population (e.g., all >80 years)

offered the program, whereas in other studies, the program was hypothetical for mostly general populations of community-dwelling older adults. One study used a preference-based discrete choice experiment to calculate utility values for different types and formats of exercise among those with a previous fall or mobility-related disability. Three other studies enrolled participants of exercise programs, reporting results of questions about program acceptance. Of the nine qualitative studies, three conducted their interviews with participants (one also interviewed those who had withdrawn or refused the intervention) of a specific fall prevention program. In the other six studies, the program was hypothetical with participants among a general population ($n=2$), interested in a program ($n=1$), or with one or more risk factors for falls (e.g., walking aid, previous fall; $n=3$).

Industry funding was not found to be a concern in this KQ. One study included in KQ1 [81] was excluded for measuring but not reporting on preferences between interventions, but we did not consider this to be a serious concern for any of the KQ3 comparisons in our synthesis.

Findings

Table 6 contains the summary of findings for preferences among interventions and their attributes. Supplementary file 3 contains the individual study data and evidence profile (Table S3.6). Studies reported on 17 relevant comparisons; we did not identify any studies specifically evaluating preferences for virtual versus in-person delivery of interventions or between different multicomponent (with exercise) or mixed interventions. Above the diagonal in Table 6, the direction and number of arrowheads indicate the preferred option and its strength, respectively; a small circle indicates variable preference, and gray shading indicates very low certainty evidence with no conclusions made about the comparison. Below the diagonal are the certainty ratings for the corresponding findings above the diagonal with indication of the GRADE domains for which we rated down. Empty cells indicate no data were available.

Balance/resistance exercise, particularly when offered using an individual format, appears to be clearly preferred among exercise interventions including over tai chi and, slightly more so, other forms of exercise (e.g., yoga,

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 25 of 32

aerobic). For exercise programs in general, preferences for group or individual delivery probably vary, though individual balance/resistance programs were clearly preferred over group programs with high certainty. Strong preferences were shown favoring individual balance/resistance programs to group "other" (yoga in the study) [27] programs, though findings were of low certainty. Delivered in a group setting, there may be variable preferences between tai chi and other forms of exercise. Balance/resistance exercise may be preferred over education, though the evidence was low certainty and did not examine preferences for delivery format. Evidence from two other comparisons between education and exercise was rated as very low certainty. Evidence was uncertain about preference between balance/resistance exercise and CBT, though there was low certainty for a slight preference for education over CBT. Group education may be preferred over individual education. Data on the attribute of cost was rated as having very low certainty (Table S3.6).

Scarce data were available in relation to populations that may require equity or other contextual considerations. In a study among a rural US population ($n=369$) [82] and at high risk of bias, there was a strong preference for a free exercise program (55% willing to register) versus one of low cost (19% willing). A large UK study of primarily low-income participants ($n=5328$) [26] found a clear preference for individual versus group balance/resistance exercise (61% vs. 41% willing), which aligned with the findings from other studies in the review. Finally, a small study of a US Latino, Spanish-speaking population ($n=103$) [83] found more interest (78%) in an individual education session compared with a group balance/resistance program (65%); no other studies reported on this comparison. None of the trials exclusively recruited or reported on subgroup effects for Indigenous peoples, newcomers, or Black Canadians.

Discussion

Summary of principal findings within and across questions

Falls in older, community-dwelling adults lead to substantial burden to the health care system and individuals from resulting injuries, fractures, and reduced functioning and quality of life. Many interventions for preventing falls have been studied, and determining which interventions to use in primary care may be influenced by their effectiveness, factors relevant to their implementation, and by patient preferences. We systematically reviewed evidence on benefits and harms for various types of fall prevention interventions and examined the relative importance placed by patients on the potential outcomes and their preferences between interventions shown to offer benefit. Findings will inform national recommendations in Canada and other decision-makers, such as primary care providers and policy makers, on fall prevention interventions.

Benefits and harms

For benefit outcomes, we defined 59 interventions interventions, including the main comparator of usual care, and ran 8 network meta-analyses. Approximately, half of the interventions in each network had at least low certainty for benefit in one or more outcome. The fallers outcome had the highest number of interventions with moderate certainty for benefit (18/57). For the non-fall outcomes (fractures, hip fracture, long-term care admission, functional status, HRQoL), there were many interventions with very low certainty evidence, many from lack of data. Although this review did not directly examine effects by level of risk for falling, because almost 70% of studies enrolled participants with an elevated risk for falling (based on previous fall history and/or other risk factors, the evidence may be most applicable to older community-dwelling adults at elevated risk for falling.

When examining interventions with moderate or high certainty of some benefit ($n=21$), the majority included a structured exercise component, and the supervised (>2 sessions), long-duration balance/resistance and group tai chi interventions had at least low certainty evidence for benefit across more outcomes than any other exercise or single-component intervention. None of the interventions with moderate certainty evidence focused on walking. Other beneficial interventions included the following: whole-body vibration, high-intensity/dose education, cognitive-behavioral therapy (CBT), and mixed interventions combining comprehensive multifactorial assessment with home-hazard assessment (HHA), or with HHA plus exercise or education. Overall, adding other interventions to exercise does not appear to substantially increase benefits for falls prevention. No multifactorial intervention showed moderate certainty for any benefit.

Six interventions offered low-certainty evidence for little-to-no difference or some harm across all outcomes. Of note, these included multifactorial interventions with limited assessment as well as brief (<3 months) supervised group walking; no data was found for supervised longer-duration walking programs. To our knowledge, no comprehensive meta-analysis of benefits and harms of walking programs in older adults has been undertaken; the potential signals of harm for single-component walking interventions found in our review suggest caution may be advisable when considering these programs for individuals, especially those at elevated risk of falls.

Two interventions, universal visual assessment and treatment provided to a general population ("universal VAT") and multifactorial intervention with a focus on

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 26 of 32

exercise and HAA (when indicated), had moderate certainty for some harm, though the populations across studies of the multifactorial intervention were at high risk of falls. Vitamin D and most single-component exercise interventions are probably associated with minimal harm. Uncertainty exists about possible harm from other interventions.

Outcome valuation

For outcome valuation, admission to long-term care had the highest disutility (1.0), but evidence was rated as low certainty. Fractures (hip and non-hip fracture disutilities 0.53 and 0.57 in the first 3 months but substantially lower at 12 months) are probably more important than either falls (0.09 over 12 months) or functional status (0.12). The findings on fracture disutility are applicable to fractures resulting from falls. Functional status may be somewhat more important than falls. Considering this data alongside the findings on the benefits of interventions, findings for fractures may be important. Among the interventions with moderate certainty for one or more benefit, interventions where there was some certainty for reducing fractures (at least 0.6 fewer per 100 treated) included, with moderate and high certainty evidence, mixed interventions adding HHA or education to a multifactorial intervention and, with low certainty, supervised, long-duration balance/resistance or group tai chi interventions, long-duration CBT, and whole-body vibration.

Intervention preferences

For intervention preferences, exercise focusing on balance/resistance, particularly when offered using an individual format, appears to be clearly preferred over tai chi and, slightly more so, other forms of exercise (e.g., yoga, aerobic). Preferences for group versus individual delivery of exercise programs in general probably vary among people, though there was high certainty that individual balance/resistance programs were clearly preferred over group-based programs. Balance/resistance exercise may be preferred over education. There was low certainty for a slight preference for education over CBT, and group education may be preferred over individual education. Findings from our analysis on intervention benefits showed that both group and individual balance/resistance and education interventions (all of long duration/high intensity) are probably beneficial; all of these options could be offered with allowance for patient preference. Some people may prefer tai chi over balance/resistance programs, and offering this choice (in group sessions) seems suitable, if available. Evidence was lacking for preferences between exercise and mixed or multicomponent interventions; however, mixed interventions with HHA may be a suitable alternative for patients not motivated or able to enrol in exercise, education, or CBT classes. As mentioned above, multicomponent interventions including exercise and/or education may not offer added benefit to these single-component interventions.

Comparison with other reviews and guidelines

Although we used another review [29] to locate studies before 2016 and avoid duplication of work in certain areas (mainly risk-of-bias assessment), that review's scope, eligible population, and method for classifying interventions differed substantially from ours, and comparing findings may not be very informative. We also considered factors other than effect size and statistical significance when making conclusions. Some interventions found effective in that review lacked data in our review. A 2020 subgroup analysis of the previous review focused on types of exercise components, and these findings may be useful for those implementing exercise programs [84]. Another review based on and in collaboration with the review's lead authors updated the search and limited eligibility to community-dwelling adults without specific conditions [85]. They used a very similar taxonomy to the original review, again making comparison with our findings difficult, but found that exercise reduced fallers, fall rates, and fractures. In sensitivity analysis, they combined all multifactorial interventions into one intervention, which was shown to reduced fall rates and not fallers. The findings of this review helped inform (with other reports and reviews including the two Cochrane reviews discussed below) the Task Force on Global Guidelines for Falls in Older Adults when developing recommendations for community-dwelling older adults within the World Guidelines for Falls [86]. The guidance recommends risk stratification—via opportunistic case finding using three questions about falls over past year, balance, and fear/worry about falling—and then (i) for individuals at low fall risk, provision of advice on how to maintain safe mobility and optimize physical functioning; (ii) if at intermediate risk (e.g., a single non-severe fall and have gait and/or balance problems), exercise programs (supervised, at least 3 sessions per week for at least 12 weeks) focusing on functional balance and resistance, or tai chi; and (iii) for those at high risk, multi-domain (multicomponent) interventions, informed by a multi-professional, multifactorial falls risk assessment (several recommendations for the assessment including HHA with provision of modifications). Other single-component interventions (e.g., medication review, VAT, HHA, CBT) are not recommended unless part of a multi-domain intervention. It is not clear what evidence was

used to differentiate between recommendations based on risk or how they determined the specific recommendations about exercise.

Two Cochrane reviews about fall prevention in community-dwelling populations used a fairly similar intervention classification to ours. Authors of the review on multifactorial and multiple component interventions compared with usual care or exercise found among multifactorial interventions (44 trials) low certainty evidence for a reduction in fall rates but little-to-no difference for fallers and injurious falls (leading to hospitalization or medical attention), and that multicomponent interventions (18 trials, usually with exercise and education or HHA) reduced fallers and fall rates with moderate certainty but had very low certainty for fractures and injurious falls [18]. Very little evidence was examined for comparisons with exercise. Our search added 6 years of studies to the evidence base for these interventions. Further, we separated out interventions only offering multifactorial assessment with needs-based treatment from those adding one or more interventions provided to everyone (which we have termed "mixed"), which may have further impacted any differences found. The second Cochrane review, focusing on exercise as a single intervention, found evidence of high certainty for a reduction in fallers and fall rates and of low certainty for fractures and injurious falls [19]. Subgroup findings were significant based on type of exercise, identifying benefit from balance/resistance, balance/functioning, and tai chi, but uncertainty around interventions focused on resistance training or dance or walking. Our findings have hopefully clarified questions about intensity, delivery format, and need for supervision; however, uncertainty around the impact of single-component walking programs remains.

The United States Preventive Services Task Force recently released final recommendations for fall prevention in community-dwelling older adults [87]. With a B grade, they recommend exercise among those at increased risk for falls, and with a C grade, they recommend clinicians individualize the decision to offer multifactorial interventions to prevent falls to community-dwelling adults 65 years or older who are at increased risk for falls. The C grade is clarified with a statement that existing evidence indicates that the overall net benefit of routinely offering multifactorial interventions to prevent falls is small. The commissioned systematic review sought trials of multiple interventions in comparison with non-active control groups and undertook pair-wise meta-analysis where suitable [88]. Assistive technology and vitamin D and other supplement interventions were excluded. The authors approached

injurious falls differently than we did, using the most inclusive outcome rather than most serious. For multifactorial interventions, the authors reported moderate certainty for a reduction in falls and low certainty for no reduction in fallers, injurious falls/injured fallers, and fall-related fractures. They comment on the potential challenges of implementing this type of complex multi-step intervention in any setting and hypothesize that the limited benefit may be due to the contemporary standard of care (fall risk assessment, medication review, exercise referrals) modifying risk in the control group. The two major differences between their review and ours are as follows: (i) their reliance on statistical significance when making conclusions about direction of effect and (ii) the handling of multifactorial interventions. They classified several interventions as multifactorial that we included in our mixed intervention category, stating that several but not all of their multifactorial interventions provided intervention components (e.g., HHA, education, exercise) to all participants. We also distinguished between comprehensive and limited risk assessment. Exercise interventions were given moderate certainty for reducing falls and fallers and low certainty for reducing injurious falls and for not offering benefit for fractures. For exercise interventions, the authors examined potential effect modifiers including intervention components, the presence of a behavior change component, the presence of cognitive task exercises, and delivery format. They concluded that no single "best" exercise/physical therapy program protocol appeared, but that nearly all examined interventions included gait/balance/functional training and strength/resistance, and that adherence to exercise classes may be variable in real-world settings. Our review examined the presence of supervision and duration of exercise interventions and allowed us to conclude that supervised and long-duration interventions may be most beneficial. Their conclusions of low and often very low ("insufficient") strength of evidence about exercise combined with education, education alone ($n=1$ of group, high-intensity education), exercise and environmental interventions, environmental interventions alone, medication review, and psychological interventions differ in some instances from ours. It is unclear to what extent this difference comes from review methodology (i.e., intervention classifications, use of network meta-analysis) versus study findings. During our CINeMA assessments, we did not rate down for single studies (either for heterogeneity or indirectness), and we rated down a maximum of two levels for imprecision when at times there were very imprecise findings (hence why we focus on findings of moderate or high certainty). This review,

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 28 of 32

as well as others, used longest follow-up duration (as did we) for their main analysis.

Limitations of the evidence, research gaps, and implications for practice

While there are many trials of interventions to prevent falls, evidence from trials for interventions, such as long-duration walking, medication review, multicomponent interventions without exercise, and mobility aids, was either lacking or came from studies with small samples leading to high imprecision. Limited comparisons of virtual interventions were available, and the effects of virtual interventions remain uncertain. One large, ongoing cluster RCT ($n = 12{,}238$ enrolled, to be completed in late 2025) in the USA, Electronic Strategies for Tailored Exercise to Prevent FallS (NCT04993781), will provide evidence about whether exercise programs can be successfully tailored and supervised using a web application. Intervention-related adverse effects were uncommonly reported; trials should actively ascertain and report on these outcomes. More data on fractures, long-term care admissions, and from patient-reported outcome measures would also strengthen the evidence base.

Although many trials measured outcomes over 12 months, the durability of effects over several years is less certain. Further, not all interventions will work for everyone. Large trials of the most promising interventions, i.e., balance/resistance exercise, should compare differing lengths and approaches, for example, whether "instructional booster sessions" are as effective as continual supervision, and follow participants over several years to better determine effect durability. Frequent assessment of a patient's fall status and other risk factors may be prudent to provide high-quality patient care. It is unclear if the findings apply well to people at average/low risk of falling, as are many older community-dwelling adults. Strategies determining for whom to offer interventions should likely be considered. There were few trials focused on, or reporting effects specific to, populations where equity factors should be considered, such as among Indigenous peoples, Black Canadians or newcomers. Future trials should purposively recruit equity-deserving populations and investigate possible differences in effects.

We did not locate preference data for several comparisons of interest, including exercise or education versus mixed interventions or multicomponent interventions (e.g., with exercise and education or medication review). The high-certainty preference for individual over group balance/resistance programs may need to be considered alongside the availability of and resources required for individual program. It is uncertain what people think about whole-body vibration, particularly at the intensity of the protocol studied in this review; attending sessions 5 days per week for 18 months may not be feasible or acceptable for many older adults. Likewise, if mixed or multicomponent interventions are not available and/or difficult to implement, seeking local preferences for these may be useful.

Limitations of the reviews

We implemented rigorous searches to locate all potentially relevant studies, but may have missed some studies, particularly if unpublished. Some decisions for eligibility, data extraction, and analysis were revised after the protocol was developed, but these are described herein and were made without consideration of any study findings or analyses. We performed meta-regressions and subgroup analyses to investigate potential effect modifiers, though not all potential variables were considered. Further, meta-regressions for the non-fall outcomes were not conducted, and our full networks were quite large, with some interventions reported by one or few studies, such that power to detect differences may have been insufficient. Because of this, we ran meta-regressions on simplified networks to further inform our investigations of transitivity and certainty assessments. Heterogeneity in the analyses for disutilities was not considered serious. We used verification, rather than duplicate independent data extraction, for results data and, in the reviews on outcome valuation and intervention preferences, for risk-of-bias assessments. Our team is highly trained in systematic reviews and uses pilot exercises to ensure understanding to prevent errors and biases; nevertheless, some errors may have occurred.

As the last search was run August 25, 2023, the evidence used in these reviews may not be fully up to date. For complex reviews such as these, updates are time and resource intensive. Especially for reviews with a large evidence base for which new findings are not anticipated to substantially change the review conclusions, extended timelines are justifiable, particularly when the risk of not publishing outweighs the need to provide up-to-date evidence [89].

Not all decision-makers would choose the same coding framework as ours, and this could impact how findings are interpreted. In general, findings of the most benefit for supervised balance/resistance and group tai chi exercise, mixed interventions, and multicomponent interventions with exercise align with the findings from the simplified networks where there were larger effects and more precise findings for these versus other interventions (Supplementary file 2), though we did not rate the certainty of those findings. Effectiveness may vary based on variables we did not investigate, such as delivery

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 29 of 32

provider or intensity of strength training, though the task force considered the variables examined to be highly relevant to implementation of programs and national recommendations. Topics related to fall risk assessment, engagement of patients, and policy or organizational factors that may impact implementation were not examined. Some guidance exists on these important aspects that may be useful to consult [90]. Our review findings may not be applicable for populations with specific conditions placing them at heightened risk for falls or subsequent sequelae of falls (e.g., severe dementia or visually impaired, stroke) and where the interventions may need to be highly tailored to the condition. Findings on intervention preferences by age, gender, or fall risk level would have been interesting to explore, though very few studies reported subgroup findings for these variables.

Conclusions

To prevent falls among community-dwelling older adults, evidence is most certain for benefit, at least over 1–2 years, from supervised, long-duration balance/resistance and group tai chi interventions, whole-body vibration, high-intensity/dose education or cognitive-behavioral therapy, and interventions combining comprehensive multifactorial assessment and targeted treatment with HHA, HHA and exercise, or education provided to everyone. Adding other interventions to exercise does not appear to substantially increase benefits, and effects overall appear most applicable to those with elevated falls risk. Choice among available effective interventions should consider patient preferences. If implementing new balance/resistance programs, offering individual rather than group sessions when feasible may be most acceptable. Data on more patient-important outcomes including fall-related fractures would be beneficial, as would studies on virtual delivery of programs and differing approaches to maintain protection.

Abbreviations

LTC	Long-term care
ADL	Activity of daily living
ProFANE	Prevention of Falls Network Europe
CBT	Cognitive behavioral therapy
HHA	Home hazard assessment and modification
HSUVs	Health-state utility values
HRQoL	Health-related quality of life
KQ	Key question
UC	Usual care
NMA	Network meta-analysis
GRADE	Grading of Recommendations Assessment, Development and Evaluation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
AE	Adverse effect
RCT	Randomized controlled trial
SD	Standard deviation
CI	Confidence interval
IQR	Interquartile range
OR	Odds ratio
RR	Rate ratio
MD	Mean difference
SIDE	Separate indirect from direct evidence
CINeMA	Confidence in Network Meta-Analysis
PrI	Prediction interval
VAT	Vision assessment and treatment

Supplementary Information

Supplementary Material 1. Additional details on methods and responses to stakeholder reviews. Table S1.1 Eligibility criteria for all key questions. Table S1.2 Intervention eligibility. Search strategies for KQ1 and KQs 2/3. Table S1.3 Risk of bias tool for adverse events. Table S1.4 Risk of bias tool for KQ2/3 studies with quantitative data. Details of certainty assessments for KQ1 benefits and harms including thresholds. Table S1.5 Effect thresholds.

Supplementary Material 2. Additional results for benefits and harms (KQ1). Figure S2.1 Literature flow diagram (KQ1): Table S2.1 Characteristics of studies included in KQ1. Table S2.2 Risk of bias ratings, by outcome (KQ1). Table S2.3 Summary of findings for interventions with moderate certainty evidence for some benefit, ranked. Table S2.4. Summary of findings for interventions with low certainty evidence. Table S2.5. Summary of findings for interventions with moderate certainty evidence for harm. Network diagrams. Forest plots of network estimates (versus UC). Evidence profiles for benefits and harms. Included studies lists (KQ1).

Supplementary Material 3. Additional results for outcome valuation and intervention preferences (KQs 2 and 3). KQ2: Figure S3.1 Literature flow diagram (KQ2). Table S3.1 Characteristics of studies included in KQ2. Included studies list for KQ2. Figure S3.2 Disutility of hip fracture at >3 mos and 12 mos, EQ5D. Figure S3.3 Disutility of non-hip fracture (at <3 mos and 12mos) and falls (within 12 mos), EQ5D. Table S3.2 Subgroup analysis for outcomes measured ≤3 months after event. Table S3.3 Subgroup analysis for outcomes measured closest to 12 months after event. KQ3: Figure S3.4 Literature flow diagram (KQ3). Table S3.4 Characteristics of studies included in Kq3, Quantitative. Table S3.5 Characteristics of studies included in KQ3, qualitative. Included studies list for kq3. Table S3.6 Evidence profile for KQ3.

Supplementary Material 4. Excluded studies lists.

Authors' contributions

JP and LAG contributed to the conception and design of the work, provided methodological expertise at all phases, participated in certainty of evidence appraisals, and drafted the manuscript with assistance from SS. JP is the guarantor of the reviews. SS, LAG, AW, and ARA screened studies for inclusion, extracted and verified data, and participated in risk-of-bias assessments. LAG, JP, and SS coded all interventions in KQ1. BV assisted with data extraction, data conversions, and ran all analyses for the NMAs. LAG and SS performed the analyses for AEs (SS) and KQs 2 and 3 (LAG). LH contributed to the conception

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 30 of 32

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Alberta Research Centre for Health Evidence, Faculty of Medicine and Dentistry, University of Alberta, 11405 87 Avenue NW, Edmonton, AB T6G 1C9, Canada.

References

1. Public Health Agency of Canada. Seniors' falls in canada: second report. PHAC. 2014. Available from: https://www.canada.ca/en/public-health/services/health-promotion/aging-seniors/publications/publications-general-public/seniors-falls-canada-second-report.html. Accessed 27 Aug 2024.
2. Public Health Agency of Canada. Surveillance report on falls among older adults in Canada. PHAC. 2022. Available from: https://www.canada.ca/content/dam/hc-sc/documents/reserch/surveillance/senior-falls-in-Canada-en.pdf. Accessed 27 Aug 2024.
3. Scuffham P, Chaplin S, Legood R. Incidence and costs of unintentional falls in older people in the United Kingdom. J Epidemiol Community Health. 2003;57(9):740–4.
4. Scott V, Wagar L, Elliott S. Falls & related injuries among older Canadians: fall-related hospitalizations & intervention initiatives. Victoria Scott Consulting. 2010. Available from: https://www.academia.edu/download/86566448/research_cemfia_phac_epi_and_inventor_20100610.pdf. Accessed 27 Aug 2024.
5. Burns ER, Stevens JA, Lee R. The direct costs of fatal and non-fatal falls among older adults - United States. J Safety Res. 2016;58:99–103.
6. Jorstad EC, Hauer K, Becker C, Lamb SE. Measuring the psychological outcomes of falling: a systematic review. J Am Geriatr Soc. 2005;53(3):501–10.
7. Stenhagen M, Ekstrom H, Nordell E, Elmstahl S. Accidental falls, health-related quality of life and life satisfaction: a prospective study of the general elderly population. Arch Gerontol Geriatr. 2014;58(1):95–100.
8. Taylor-Piliae RE, Peterson R, Mohler MJ. Clinical and community strategies to prevent falls and fall-related injuries among community-dwelling older adults. Nurs Clin North Am. 2017;52(3):489–97.
9. Girgis CM, Clifton-Bligh RJ, Hamrick MW, Holick MF, Gunton JE. The roles of vitamin D in skeletal muscle: form, function, and metabolism. Endocr Rev. 2013;34(1):33–83.
10. Gunton JE, Girgis CM. Vitamin D and muscle Bone Rep. 2018;8:163–7.
11. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). Age Ageing. 2005;34(6):614–9.
12. Rikli RE, Jones CJ. Development and validation of criterion-referenced clinically relevant fitness standards for maintaining physical independence in later years. Gerontologist. 2013;53(2):255–67.
13. Berg KO, Maki BE, Williams JI, Holliday PJ, Wood-Dauphinee SL. Clinical and laboratory measures of postural balance in an elderly population. Arch Phys Med Rehabil. 1992;73(11):1073–80.
14. Podsiadlo D, Richardson S. The timed "up & go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142–8.
15. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging Clin Exp Res. 2018;30(1):1–16.
16. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E, et al. Determining risk of falls in community dwelling older adults: a systematic review and meta-analysis using posttest probability. J Geriatr Phys Ther. 2017;40(1):1–36.
17. Guirguis-Blake JM, Michael YL, Perdue LA, Coppola EL, Beil TL. Interventions to prevent falls in older adults: updated evidence report and systematic review for the US Preventive Services Task Force. JAMA. 2018;319(16):1705–16.
18. Hopewell S, Adedire O, Copsey BJ, Boniface GJ, Sherrington C, Clemson L, et al. Multifactorial and multiple component interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev. 2018;7(7):Cd012221.
19. Sherrington C, Fairhall NJ, Wallbank GK, Tiedemann A, Michaleff ZA, Howard K, et al. Exercise for preventing falls in older people living in the community. Cochrane Database Syst Rev. 2019;1(1):Cd012424.
20. Lamb SE, Becker C, Gillespie LD, Smith JL, Finnegan S, Potter R, et al. Reporting of complex interventions in clinical trials: development of a taxonomy to classify and describe fall-prevention interventions. Trials. 2011;12:125.
21. Atun R. What are the advantages and disadvantages of restructuring a health care system to be more focused on primary care services? WHO Regional Office for Europe. 2004. Available from: https://apps.who.int/iris/bitstream/handle/10665/363801/9789289057356-eng.pdf. Accessed 27 Aug 2024.
22. Schunemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, Mustafa R, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2014;186(3):E123–42.
23. Zhang Y, Coello PA, Brozek J, Wiercioch W, Etxeandia-Ikobaltzeta I, Akl EA, et al. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. Health Qual Life Outcomes. 2017;15(1):52.
24. Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care. 1989;5(4):559–75.
25. Nyman SR, Victor CR. Older people's participation in and engagement with falls prevention interventions in community settings: an augment to the Cochrane systematic review. Age Ageing. 2012;41(1):16–23.
26. Yardley L, Kirby S, Ben-Shlomo Y, Gilbert R, Whitehead S, Todd C. How likely are older people to take up different falls prevention activities? Prev Med. 2008;47(5):554–8.
27. Franco MR, Howard K, Sherrington C, Ferreira PH, Rose J, Gomes JL, et al. Eliciting older people's preferences for exercise programs: a best-worst scaling choice experiment. J Physiother. 2015;61(1):34–41.
28. Dorresteijn TA, Rixt Zijlstra GA, Van Eijs YJ, Vlaeyen JW, Kempen GI. Older people's preferences regarding programme formats for managing concerns about falls. Age Ageing. 2012;41(4):474–81.
29. Tricco AC, Thomas SM, Veroniki AA, Hamid JS, Cogo E, Strifler L, et al. Comparisons of interventions for preventing falls in older adults: a systematic review and meta-analysis. JAMA. 2017;318(17):1687–99.
30. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50(6):683–91.
31. Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods. 2012;3(2):80–97.

Pillay *et al. Systematic Reviews*   (2024) 13:289

Page 31 of 32

32. Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PLoS One. 2013;8(10):e76654.

33. Pillay J, Riva JJ, Tessier LA, Colquhoun H, Lang E, Moore AE, et al. Fall prevention interventions for older community-dwelling adults: systematic reviews on benefits, harms, and patient values and preferences. Syst Rev. 2021;10(1):18.

34. Canadian Task Force on Preventive Health Care. CTFPHC Methods Manual. Canadian Task Force on Preventive Health Care. 2022. Available from: https://canadiantaskforce.ca/methods/. Accessed 27 Aug 2024.

35. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane. 2019. Available from: https://training.cochrane.org/handbook. Accessed 27 Aug 2024.

36. Zhang Y, Alonso-Coello P, Guyatt GH, Yepes-Nunez JJ, Akl EA, Hazlewood G, et al. GRADE guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-risk of bias and indirectness. J Clin Epidemiol. 2019;111:94–104.

37. Zhang Y, Coello PA, Guyatt GH, Yepes-Nunez JJ, Akl EA, Hazlewood G, et al. GRADE guidelines: 20. Assessing the certainty of evidence in the importance of outcomes or values and preferences-inconsistency, imprecision, and other domains. J Clin Epidemiol. 2019;111:83–93.

38. Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. 2013. Available from: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. Accessed 27 Aug 2024.

39. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009;62(10):e1–34.

40. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med. 2015;162(11):777–84.

41. Moher D, Pham B, Klassen TP, Schulz KF, Berlin JA, Jadad AR, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol. 2000;53(9):964–72.

42. Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. Int J Technol Assess Health Care. 2012;28(2):138–44.

43. . World Health Organization Clinical Trials Search Portal World Health Organization. http://apps.who.int/trialsearch/. Accessed 25 Aug 2023.

44. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol. 2016;75:40–6.

45. Rohatgi A. Web Plot Digitizer version 4.6. 2022. https://automeris.io/. Accessed 26 July 2023.

46. Cochrane Effective Practice and Organisation of Care. Suggested risk of bias criteria for EPOC reviews. Cochrane. 2017. Available from: https://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Resources-for-authors2017/suggested_risk_of_bias_criteria_for_epoc_reviews.pdf. Accessed 27 Aug 2024.

47. Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, Assessing harms when comparing medical interventions., et al. In: Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; 2008. Available from: https://www.ncbi.nlm.nih.gov/books/NBK47098/. Accessed 27 Aug 2024.

48. Critical Appraisal Skills Programme. CASP Qualitative Checklist. 2018. Available from: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/. Accessed 24 Aug 2024.

49. Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. Res Synth Methods. 2011;2(3):188–203.

50. White IR, Barrett JK, Jackson D, Higgins JP. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. Res Synth Methods. 2012;3(2):111–25.

51. Donegan S, Williamson P, D'Alessandro U, Tudur SC. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods. 2013;4(4):291–323.

52. Barua B, Jacques D. Comparing performance of universal health care countries, 2018. Fraser Institute 2018. Available from: https://www.frase rinstitute.org/sites/default/files/comparing-performance-of-universal-health-care-countries-2018.pdf. Accessed 27 Aug 2024.

53. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med. 2010;29(7–8):932–44.

54. Huhn M, Nikolakopoulou A, Schneider-Thoma J, Krause M, Samara M, Peter N, et al. Comparative efficacy and tolerability of 32 oral antipsychotics for the acute treatment of adults with multi-episode schizophrenia: a systematic review and network meta-analysis. Lancet. 2019;394(10202):939–51.

55. . CINeMA: Confidence in Network Meta-Analysis. Institute of Social and Preventive Medicine, University of Bern. 2017. https://cinema.ispm.unibe.ch/. Accessed 27 Aug 2024.

56. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629–34.

57. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. Value Health. 2017;20(4):644–50.

58. Tsiplova K, Pullenayegum E, Cooke T, Xie F. EQ-5D-derived health utilities and minimally important differences for chronic health conditions: 2011 Commonwealth Fund Survey of Sicker Adults in Canada. Qual Life Res. 2016;25(12):3009–16.

59. Si L, Winzenberg TM, de Graaff B, Palmer AJ. A systematic review and meta-analysis of utility-based quality of life for osteoporosis-related conditions. Osteoporos Int. 2014;25(8):1987–97.

60. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, et al. CINeMA: an approach for assessing confidence in the results of a network meta-analysis. PLoS Med. 2020;17(4):e1003082.

61. Elliott SA, Gaudet LA, Fernandes RM, Vandermeer B, Freedman SB, Johnson DW, et al. Comparative efficacy of bronchiolitis interventions in acute care: a network meta-analysis. Pediatrics. 2021;147(5):e2020040816.

62. Leung KS, Li CY, Tse YK, Choy TK, Leung PC, Hung VW, et al. Effects of 18-month low-magnitude high-frequency vibration on fall rate and fracture risks in 710 community elderly–a cluster-randomized controlled trial. Osteoporos Int. 2014;25(6):1785–95.

63. Lamb SE, Bruce J, Hossain A, Ji C, Longo R, Lall R, et al. Screening and intervention to prevent falls and fractures in older people. NJEM. 2020;383(19):1848–59.

64. Cumming RG, Ivers R, Clemson L, Cullen J, Hayes MF, Tanzer M, et al. Improving vision to prevent falls in frail older people: a randomized trial. J Am Geriatr Soc. 2007;55(2):175–81.

65. Fitzharris MP, Day L, Lord SR, Gordon I, Fildes B. The Whitehorse NoFalls Trial: effects on fall rates and injurious fall rates. Age Ageing. 2010;39(6):728–33.

66. Fairhall N, Sherrington C, Lord SR, Kurrle SE, Langron C, Lockwood K, et al. Effect of a multifactorial, interdisciplinary intervention on risk factors for falls and fall rate in frail older people: a randomised controlled trial. Age Ageing. 2014;43(5):616–22.

67. Ferrer A, Formiga F, Sanz H, de Vries OJ, Badia T, Pujol R. Multifactorial assessment and targeted intervention to reduce falls among the oldest-old: a randomized controlled trial. Clin Interv Aging. 2014;9:383–93.

68. Whitehead C, Wundke R, Crotty M, Finucane P. Evidence-based clinical practice in falls prevention: a randomised controlled trial of a falls prevention service. Aust Health Rev. 2003;26(3):88–97.

69. Tan PJ, Khoo EM, Chinna K, Saedon NI, Zakaria MI, Ahmad Zahedi AZ, et al. Individually-tailored multifactorial intervention to reduce falls in the Malaysian Falls Assessment and Intervention Trial (MyFAIT): a randomized controlled trial. PLoS One. 2018;13(8):e0199219.

70. Stark S, Keglovits M, Somerville E, Hu Y-L, Barker A, Sykora D, et al. Home hazard removal to reduce falls among community-dwelling older adults: a randomized clinical trial. JAMA netw open. 2021;4(8):e2122044.

71. Keall MD, Pierse N, Howden-Chapman P, Cunningham C, Cunningham M, Guria J, et al. Home modifications to reduce injuries from falls in the home injury prevention intervention (HIPI) study: a cluster-randomised controlled trial. Lancet. 2015;385(9964):231–8.

72. Keene DJ, Vadher K, Willett K, Mistry D, Costa ML, Collins GS, et al. Predicting patient-reported and objectively measured functional outcome 6 months after ankle fracture in people aged 60 years or over in the UK: prognostic model development and internal validation. BMJ Open. 2019;9(7):e029813.

73. Svedbom A, Borgstom F, Hernlund E, Strom O, Alekna V, Bianchi ML, et al. Quality of life for up to 18 months after low-energy hip, vertebral,

Pillay *et al. Systematic Reviews*   *(2024) 13:289*

Page 32 of 32

and distal forearm fractures-results from the ICUROS. Osteoporos Int. 2018;29(3):557–66.

74. Tidermark J, Zethraeus N, Svensson O, Tornkvist H, Ponzer S. Quality of life related to fracture displacement among elderly patients with femoral neck fractures treated with internal fixation. J Orthop Trauma. 2002;16(1):34–8.

75. Salkeld G, Cameron ID, Cumming RG, Easter S, Seymour J, Kurrle SE, et al. Quality of life related to fear of falling and hip fracture in older women: a time trade off study. BMJ. 2000;320(7231):341–6.

76. Abimanyi-Ochom J, Watts JJ, Borgstrom F, Nicholson GC, Shore-Lorenti C, Stuart AL, et al. Changes in quality of life associated with fragility fractures: Australian arm of the International Cost and Utility Related to Osteoporotic Fractures Study (AusICUROS). Osteoporos Int. 2015;26(6):1781–90.

77. Guirant L, Carlos F, Curiel D, Kanis JA, Borgstrom F, Svedbom A, et al. Health-related quality of life during the first year after a hip fracture: results of the Mexican arm of the International Cost and Utility Related to Osteoporotic Fractures Study (MexICUROS). Osteoporos Int. 2018;29(5):1147–54.

78. Jurisson M, Pisarev H, Kanis J, Borgstrom F, Svedbom A, Kallikorm R, et al. Quality of life, resource use, and costs related to hip fracture in Estonia. Osteoporos Int. 2016;27(8):2555–66.

79. Wetherell JL, Bower ES, Johnson K, Chang DG, Ward SR, Petkus AJ. Integrated exposure therapy and exercise reduces fear of falling and avoidance in older adults: a randomized pilot study. Am J Geriatr Psychiatry. 2018;26(8):849–59.

80. Renehan E, Meyer C, Elliott RA, Batchelor F, Said C, Haines T, et al. Posthospital falls prevention intervention: a mixed-methods study. J Aging Phys Act. 2019;27(2):155–65.

81. Snooks HA, Anthony R, Chatters R, Dale J, Fothergill R, Gaze S, et al. Support and Assessment for Fall Emergency Referrals (SAFER) 2: a cluster randomised trial and systematic review of clinical effectiveness and cost-effectiveness of new protocols for emergency ambulance paramedics to assess older people following a fall with referral to community-based care when appropriate. Health Technol Assess. 2017;21(13):1–218.

82. Kiami SR, Sky R, Goodgold S. Facilitators and barriers to enrolling in falls prevention programming among community dwelling older adults. Arch Gerontol Geriatr. 2019;82:106–13.

83. Hanlin ER, Delgado-Rendon A, Lerner EB, Hargarten S, Farias R. Fall risk and prevention needs assessment in an older adult Latino population: a model community global health partnership. Prog Community Health Partnersh. 2013;7(2):191–9.

84. Sibley KM, Thomas SM, Veroniki AA, Rodrigues M, Hamid JS, Lachance CC, et al. Comparative effectiveness of exercise interventions for preventing falls in older adults: a secondary analysis of a systematic review with network meta-analysis. Exp Gerontol. 2021;143:111151.

85. Dautzenberg L, Beglinger S, Tsokani S, Zevgiti S, Raijmann R, Rodondi N, et al. Interventions for preventing falls and fall-related fractures in community-dwelling older adults: a systematic review and network meta-analysis. J Am Geriatr Soc. 2021;69(10):2973–84.

86. Montero-Odasso M, van der Velde N, Martin FC, Petrovic M, Tan MP, Ryg J, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing. 2022;51(9):afac205.

87. Nicholson WK, Silverstein M, Wong JB, Barry MJ, Chelmow D, Coker TR, et al. Interventions to prevent falls in community-dwelling older adults: US Preventive Services Task Force recommendation statement. JAMA. 2024;332(1):51–7.

88. Guirguis-Blake JM, Perdue LA, Coppola EL, Bean SI. Interventions to prevent falls in older adults: updated evidence report and systematic review for the Us Preventive Services Task Force. JAMA. 2024;332(1):58–69.

89. Stokes G, Sutcliffe K, Thomas J. Is a one-size-fits-all '12-month rule' appropriate when it comes to the last search date in systematic reviews? BMJ Evid Based Med. 2023;28(6):359–63.

90. Registered Nurses' Association of Canada. Registered Nurses' Association of Canada. 4th ed. Toronto; 2017.

Publisher's Note