Supplemental Table S3: GRADE – Physical Activity (overall)

GRADE evidence rating: Physical activity interventions compared to usual care for older adults living with frailty or pre-frailty

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Physical activity int	usual care	Relative / Absolute (95% CI)		
1. Mobility ***(follow up: range 8 weeks to 12 months; assessed with: Performance measures (Gait speed, Timed up & go, chair sit & stand, balance, short physical performance battery))***											
19 [a]	randomised trials	serious [b]	not serious [c]	not serious	not serious [d]	none	946	778	SMD **0.6 SD higher** (0.37 higher to 0.83 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
2. Activities of daily living ***(follow up: range 8 weeks to 12 months; assessed with: ADL / IADL instruments)***											
9 [e]	randomised trials	serious [f]	not serious [c]	not serious	not serious [d]	none	495	415	SMD **0.5 SD higher** (0.15 higher to 0.84 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
3. Cognitive function ***(follow up: range 8 weeks to 6 months; assessed with: MMSE, LOTCA-G, rey memory score, RBANS z-score)***											
5 [g]	randomised trials	serious [h]	not serious [c]	not serious	not serious [i]	none	186	191	SMD **0.35 SD higher** (0.09 higher to 0.61 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
4. Quality of life ***(follow up: range 12 weeks to 9 months; assessed with: : SF-36 Physical & Mental component, EQ5D-VAS, SSWO score)***											
6 [j]	randomised trials	serious [k]	not serious [c]	not serious	not serious [i]	none	260	240	SMD **0.6 SD higher** (0.13 higher to 1.07 higher)	⊕⊕⊕◯ MODERATE	CRITICAL

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Physical activity int	usual care	Relative / Absolute (95% CI)		
5. Frailty ***(follow up: range 6 weeks to 6 months; assessed with: Cardiovascular Health Study, Edmonton frailty, Modified Fried criteria)***											
4 [l]	randomised trials	serious [m]	not serious [c]	not serious	not serious [i]	none	120	124	SMD **1.29 SD lower** (2.22 lower to 0.36 lower)	⊕⊕⊕◯ MODERATE	CRITICAL
6. Prevalence of Frailty ***(follow up: range 8 weeks to 24 months; assessed with: Number frail at post-intervention)***											
4 [n]	randomised trials	serious [o]	not serious [c]	not serious	not serious [p]	none	166/763 (21.8%)	246/775 (31.7%)	**RR 0.58** (0.36 to 0.93)	**133 fewer per 1,000** (from 203 fewer to 22 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
7. Incidence of Fall ***(follow up: range 20 weeks to 12 months; assessed with: Number of events at post-intervention)***											
4 [q]	randomised trials	serious [r]	serious [s]	not serious	serious [t]	none	156/420 (37.1%)	137/304 (45.1%)	**RR 0.80** (0.51 to 1.26)	**90 fewer per 1,000** (from 221 fewer to 117 more)	⊕◯◯◯ VERY LOW	CRITICAL

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Physical activity int	usual care	Relative / Absolute (95% CI)		
8. Fatigue level *(follow up: range 8 weeks to 6 months; assessed with: VAS Fatigue intensity, Chinese fatigue inventory, SF-12 subscale)*											
3 [u]	randomised trials	serious [v]	not serious [w]	not serious	serious [x]	none	92	92	SMD **0.27 SD lower** (0.65 lower to 0.12 higher)	⨁⨁◯◯ LOW	CRITICAL

CI: Confidence interval; **SMD:** Standardised mean difference; **RR:** Risk ratio

Note: There was only data from one included study, and therefore no GRADE, for the following outcomes; Health Services Use

Explanations

a. Kuo, 2018; Gill, 2002; Brown, 2000; Tsang, 2013; Chen, 2019; Clegg, 2014; Yoon, 2018; Takatori, 2016; Tieland, 2015; Faber, 2006; Losa-Reyna, 2019; Liu, 2017; Santabalbina, 2016; Pin Ng, 2015; Kwon, 2015; Daniel, 2012; Gine-Garriga, 2010; Binder, 2002; de Jong, 2000

b. 11 out of 19 studies were rated as unclear risk (9 studies) and high risk (2 studies) with concerns regarding randomization, allocation concealment, blinding, incomplete and selective outcome reporting, and other risk of bias (such as baseline imbalance across groups).

c. High statistical heterogeneity observed, however, the direction of effect is consistent across most studies with overlapping confidence intervals and statistical heterogeneity is likely due to small versus large effects observed across studies.

d. The sample size is adequate (=>300) in both intervention and control arms and effect estimate is precise (Confidence intervals do not include the no effect value "0").

e. Kuo, 2018; Gill, 2002; Clegg, 2014; Faber, 2006; Santabalbina, 2016; Daniel, 2012; Gine-Garriga, 2010; Binder, 2002; de Jong, 2000

f. 4 out of 9 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

g. Kuo, 2018; Tsang, 2013; Yoon, 2018; Santabalbina, 2016; Pin Ng, 2015

h. 3 out of 5 studies were rated as unclear risk with concerns regarding incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

i. The sample size is not adequate (<300) in each arm, however, effect estimate is precise with confidence intervals not including the no effect value of "0". j.

Clegg, 2014; Santabalbina, 2016; Kwon, 2015; Gine-Garriga, 2013; Binder, 2002; de Jong, 2000

k. 4 out of 6 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

l. Yoon, 2018; Losa-Reyna, 2019; Santabalbina, 2016; Pin Ng, 2015

m. 3 out of 4 studies were rated as unclear risk with concerns regarding incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

n. Chen, 2019; Liu, 2018; Santabalbina, 2016; Pin Ng, 2015

o. 2 out of 4 studies were rated as unclear risk with concerns regarding other risk of bias (such as baseline imbalance across groups).

p. The sample size is adequate (=>300) in both intervention and control arms and effect estimate is precise (Confidence intervals do not include the no effect value "1").

q. Gill, 2002; Faber, 2006; Takatori, 2016; Pin Ng, 2015

r. 1 out of 4 studies rated as high risk with concerns regarding randomization, allocation concealment, blinding and other risk of bias (such as baseline imbalance across groups).

s. The direction of effect is not consistent and confidence intervals do not overlap with substantial level of statistical heterogeneity observed across studies.

t. The sample size is adequate (=>300) in each arm, however, the number of events are low and effect estimate is imprecise with confidence intervals including the no effect value of "1".

u. Kuo, 2018; Liu, 2017; Pin Ng, 2015

v. 1 out of 3 studies rated as high risk with concerns regarding randomization, allocation concealment, blinding and other risk of bias (such as baseline imbalance across groups).

w. The confidence intervals overlap with moderate level of statistical heterogeneity observed across studies.

x. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "0".

Supplemental Table S4: GRADE – Aerobic Physical Activity

GRADE evidence rating: **Aerobic physical activity interventions compared to usual care for older adults living with frailty or pre-frailty**

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Aerobic interventions	usual care	Relative / Absolute (95% CI)		
1. Mobility ***(follow up: mean 8 weeks; assessed with: Performance measures (Gait speed, Timed up & go test))***											
1 [a]	randomised trials	serious [b]	not serious	not serious	serious [c]	none	15	21	SMD **0.71 SD higher** (0.23 higher to 1.2 higher)	⊕⊕◯◯ LOW	CRITICAL
2. Activities of daily living ***(follow up: mean 8 weeks; assessed with: ADL / IADL instruments)***											
1 [a]	randomised trials	serious [b]	not serious	not serious	very serious [d]	none	15	21	SMD **0.46 SD higher** (0.03 lower to 0.94 higher)	⊕◯◯◯ VERY LOW	CRITICAL
3. Cognitive function ***(follow up: mean 8 weeks; assessed with: MMSE score)***											
1 [a]	randomised trials	serious [b]	not serious	not serious	very serious [d]	none	15	21	SMD **0.15 SD higher** (0.5 lower to 0.8 higher)	⊕◯◯◯ VERY LOW	CRITICAL
4. Fatigue level ***(follow up: mean 8 weeks; assessed with: VAS fatigue intensity)***											
1 [a]	randomised trials	serious [b]	not serious	not serious	very serious [d]	none	15	21	SMD **0.39 SD lower** (0.87 lower to 0.09 higher)	⊕◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **SMD:** Standardised mean difference

Note: There was no data in the included studies for the following outcomes; Quality of Life, Frailty, Falls, Health Services Use

Explanations

a. Kuo, 2018

b. The study had concerns regarding allocation concealment and other risk of bias (such as baseline imbalance across groups).

c. The sample size is small (<30) in each arm, and effect estimate is imprecise with wide confidence intervals.

d. The sample size is small (<30) in each arm, and effect estimate is imprecise with confidence intervals including the no effect value of "0"

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Supplemental Table S5: GRADE – Muscle Strengthening Physical Activity

GRADE evidence rating: Muscle strengthening physical activity interventions compared to usual care for older adults living with frailty or pre-frailty

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Muscle strengthening int	usual care	Relative / Absolute (95% CI)		
1. Mobility ***(follow up: range 8 weeks to 6 months; assessed with: Performance measures (Gait speed, Timed up & go, chair sit & stand, balance, short physical performance battery))***											
6 [a]	randomised trials	serious [b]	not serious [c]	not serious	not serious [d]	none	419	303	SMD **0.57 SD higher** (0.08 higher to 1.06 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
2. Activities of daily living ***(follow up: range 12 weeks to 20 weeks; assessed with: ADL / IADL instruments)***											
2 [e]	randomised trials	serious [f]	not serious	not serious	serious [g]	none	164	114	SMD **0.16 SD higher** (0.05 lower to 0.37 higher)	⊕⊕◯◯ LOW	CRITICAL
3. Cognitive function ***(follow up: mean 16 weeks; assessed with: Rey memory score, cognitive flexibility, processing speed (TMT-A), frontal assessment battery (FAB))***											
1 [h]	randomised trials	serious [i]	not serious	not serious	very serious [j]	none	22	23	SMD **0.45 SD higher** (0.19 higher to 0.72 higher)	⊕◯◯◯ VERY LOW	CRITICAL
4. Quality of life ***(follow up: mean 12 weeks; assessed with: EQ5D-VAS)***											
1 [k]	randomised trials	serious [l]	not serious	not serious	very serious [m]	none	40	30	SMD **0.15 SD higher** (0.33 lower to 0.63 higher)	⊕◯◯◯ VERY LOW	CRITICAL

GRADE – Muscle Strengthening Physical Activity

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Muscle strengthening int	usual care	Relative / Absolute (95% CI)			
5. Frailty ***(follow up: mean 16 weeks; assessed with: Cardiovascular Health Study criteria)***												
1 [h]	randomised trials	serious [i]	not serious	not serious	very serious [m]	none	22	23	SMD **0.2 SD lower** (0.79 lower to 0.39 higher)		⊕◯◯◯ VERY LOW	CRITICAL
6. Prevalence of Frailty ***(assessed with: Number Frail at post-intervention)***												
1 [n]	randomised trials	not serious	not serious	not serious	serious [o]	none	6/33 (18.2%)	29/33 (87.9 %)	**RR 0.21** (0.10 to 0.43)	**694 fewer per 1,000** (from 791 fewer to 501 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
7. Incidence of Fall ***(follow up: range 20 weeks to 6 months; assessed with: Number of events at post-intervention)***												
2 [p]	randomised trials	serious [q]	serious [r]	not serious	serious [s]	none	102/280 (36.4%)	74/162 (45.7 %)	**RR 0.78** (0.37 to 1.65)	**100 fewer per 1,000** (from 288 fewer to 297 more)	⊕◯◯◯ VERY LOW	CRITICAL

CI: Confidence interval; **SMD:** Standardised mean difference; **RR:** Risk ratio

Explanations

a. Chen, 2019; Clegg, 2014; Yoon, 2018; Takatori, 2016; Tieland, 2015; Faber, 2006

b. 4 out of 6 studies were rated as unclear risk (3 studies) and high risk (1 study) with concerns regarding randomization, allocation concealment, blinding, incomplete and selective outcome reporting, and other risk of bias (such as baseline imbalance across groups).

c. High statistical heterogeneity observed, however, the direction of effect is consistent across most studies with overlapping confidence intervals and statistical heterogeneity is likely due to small versus large effects observed across studies.

d. The sample size is adequate (=>300) in both intervention and control arms and effect estimate is precise (Confidence intervals do not include the no effect value "0").

e. Clegg, 2014; Faber, 2006

f. 1 out of 2 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

g. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "0". h. Yoon, 2018

i. The study had concerns regarding randomization, allocation concealment, incomplete outcome reporting and other risk of bias (such as baseline imbalance across groups).

j. The sample size is small (<30) in each arm, and effect estimate is imprecise with wide confidence intervals.k. Clegg, 2014

l. The study had concerns regarding randomization, blinding, allocation concealment, incomplete outcome reporting and other risk of bias (such as baseline imbalance across groups).

m. The sample size is small (<30) in each arm, and effect estimate is imprecise with confidence intervals including the no effect value of "0" n. Chen, 2019

o. The sample size is not adequate (<300) in each arm, and effect estimate is imprecise with wide confidence intervals.

p. Faber, 2006; Takatori, 2016

q. 1 out of 2 studies rated as high risk with concerns regarding randomization, allocation concealment, blinding and other risk of bias (such as baseline imbalance across groups).

r. The direction of effect is not consistent and confidence intervals do not overlap with substantial level of statistical heterogeneity observed across studies.

s. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "1".

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Supplemental Table S6: GRADE – Mobility & Rehab Physical Activity

GRADE evidence rating: Mobility & Rehab physical activity interventions compared to usual care for older adults living with frailty or pre-frailty

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mobility & Rehab int	usual care	Relative / Absolute (95% CI)		
1. Mobility ***(follow up: range 12 weeks to 12 months; assessed with: Performance measures (Gait speed, Timed up & go, chair sit & stand, balance, short physical performance battery))***											
3 [a]	randomised trials	serious [b]	not serious [c]	not serious	not serious [d]	none	175	155	SMD **0.29 SD higher** (0.17 higher to 0.42 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
2. Activities of daily living ***(follow up: mean 12 months; assessed with: ADL / IADL instruments)***											
1 [e]	randomised trials	serious [f]	not serious	not serious	not serious [d]	none	91	91	SMD **0.48 SD higher** (0.28 higher to 0.67 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
3. Cognitive function ***(follow up: mean 12 weeks; assessed with: LOTCA-G)***											
1 [g]	randomised trials	serious [h]	not serious	not serious	serious [i]	none	61	55	SMD **0.12 SD higher** (0.1 lower to 0.34 higher)	⊕⊕◯◯ LOW	CRITICAL

GRADE – Mobility & Rehab Physical Activity

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mobility & Rehab int	usual care	Relative / Absolute (95% CI)			
4. Incidence of Fall ***(follow up: mean 12 months; assessed with: Number of events at post-intervention)***												
1 [e]	randomised trials	serious [f]	not serious	not serious	serious [j]	none	51/92 (55.4%)	58/92 (63.0 %)	**RR 0.88** (0.69 to 1.12)	**76 fewer per 1,000** (from 195 fewer to 76 more)	⨁⨁◯◯ LOW	CRITICAL

CI: Confidence interval; **SMD:** Standardised mean difference; **RR:** Risk ratio

Note: There was no data in the included studies for the following outcomes; Quality of Life, Frailty, Fatigue, Health Services Use

Explanations

a. Gill, 2002; Brown, 2000; Tsang, 2013

b. 2 out of 3 studies were rated as unclear risk with concerns regarding other risk of bias (such as baseline imbalance across groups).

c. The confidence intervals overlap with low statistical heterogeneity observed across studies.

d. The sample size is not adequate (<300) in each arm, however, effect estimate is precise with confidence intervals not including the no effect value of "0". e. Gill, 2002

f. The study had concerns regarding allocation concealment and other risk of bias (such as baseline imbalance across groups).

g. Tsang, 2013

h. The study had concerns regarding randomization, blinding, allocation concealment and other risk of bias (such as baseline imbalance across groups).

i. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "0".

j. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "1".

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Supplemental Table S7: GRADE – Mixed Physical Activity

GRADE evidence rating: Mixed physical activity interventions compared to usual care for older adults living with frailty or pre-frailty

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed physical activity int	usual care	Relative / Absolute (95% CI)		
1. Mobility ***(follow up: range 6 weeks to 9 months; assessed with: Performance measures (Gait speed, Timed up & go, chair sit & stand, balance, short physical performance battery))***											
9 [a]	randomised trials	serious [b]	not serious [c]	not serious	not serious [d]	none	337	299	SMD **0.75 SD higher** (0.4 higher to 1.1 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
2. Activities of daily living ***(follow up: range 12 weeks to 9 months; assessed with: ADL / IADL instruments)***											
5 [e]	randomised trials	serious [f]	serious [g]	not serious	not serious [h]	none	225	189	SMD **0.64 SD higher** (0.004 higher to 1.27 higher)	⊕⊕◯◯ LOW	CRITICAL
3. Cognitive function ***(follow up: mean 24 weeks; assessed with: MMSE, RBANS z-score)***											
2 [i]	randomised trials	serious [j]	not serious [c]	not serious	not serious [h]	none	88	92	SMD **0.62 SD higher** (0.12 higher to 1.11 higher)	⊕⊕⊕◯ MODERATE	CRITICAL
4. Quality of life ***(follow up: range 12 weeks to 9 months; assessed with: SF-36 Physical & Mental component, EQ5D-VAS, SSWO score)***											
5 [k]	randomised trials	serious [l]	not serious [c]	not serious	not serious [h]	none	220	210	SMD **0.68 SD higher** (0.16 higher to 1.21 higher)	⊕⊕⊕◯ MODERATE	CRITICAL

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed physical activity int	usual care	Relative / Absolute (95% CI)		
5. Frailty ***(follow up: range 6 weeks to 6 months; assessed with: Cardiovascular Health Study, Edmonton frailty, Modified Fried criteria)***											
3 [m]	randomised trials	serious [n]	not serious [c]	not serious	not serious [h]	none	98	101	SMD **1.57 SD lower** (2.57 lower to 0.57 lower)	⊕⊕⊕◯ MODERATE	CRITICAL
6. Prevalence of Frailty ***(follow up: range 24 weeks to 24 months; assessed with: Number frail at post-intervention)***											
3 [o]	randomised trials	serious [n]	not serious [p]	not serious	not serious [q]	none	160/730 (21.9%)	217/742 (29.2 %)	**RR 0.72** (0.63 to 0.83)	**82 fewer per 1,000** (from 108 fewer to 50 fewer)	⊕⊕⊕◯ MODERATE	CRITICAL
7. Falls ***(follow up: mean 24 weeks; assessed with: Mean number)***											
1 [r]	randomised trials	serious [s]	not serious	not serious	serious [t]	none	40	42	SMD **0.37 SD lower** (0.81 lower to 0.07 higher)	⊕⊕◯◯ LOW	CRITICAL

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Certainty assessment							№ of patients		Effect		Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed physical activity int	usual care	Relative / Absolute (95% CI)			
8. Incidence of Fall ***(follow up: mean 6 months; assessed with: Number of events at post-intervention)***												
1 [u]	randomised trials	not serious	not serious	not serious	very serious [v]	none	3/48 (6.3%)	5/50 (10.0 %)	**RR 0.62** (0.16 to 2.47)	**38 fewer per 1,000** (from 84 fewer to 147 more)	⊕⊕◯◯ LOW	CRITICAL
9. Fatigue level ***(follow up: range 16 weeks to 6 months; assessed with: Chinese fatigue inventory, SF-12 subscale)***												
2 [w]	randomised trials	serious [x]	serious [g]	not serious	serious [t]	none	77	71	SMD **0.23 SD lower** (0.85 lower to 0.39 higher)		⊕◯◯◯ VERY LOW	CRITICAL
10. Health services use ***(follow up: mean 24 weeks; assessed with: Mean number of Emergency visits)***												
1 [r]	randomised trials	serious [s]	not serious	not serious	serious [t]	none	40	42	SMD **0.21 SD lower** (0.65 lower to 0.23 higher)		⊕⊕◯◯ LOW	CRITICAL

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

Certainty assessment							№ of patients		Effect	Certainty	Importance
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Mixed physical activity int	usual care	Relative / Absolute (95% CI)		
11. Health services use ***(follow up: mean 6 months; assessed with: Number hospitalized at post-intervention)***											
1 [u]	randomised trials	not serious	not serious	not serious	very serious [v]	none	1/48 (2.1%)	2/50 (4.0%)	**RR 0.52** (0.05 to 5.56)	**19 fewer per 1,000** (from 38 fewer to 182 more)	⨁⨁◯◯ LOW	CRITICAL

CI: Confidence interval; **SMD:** Standardised mean difference; **RR:** Risk ratio

Explanations

a. Losa-Reyna, 2019; Liu, 2017; Santabalbina, 2016; Pin Ng, 2015; Kwon, 2015; Daniel, 2012; Gine-Garriga, 2010; F. Binder, 2002; de Jong, 2000

b. 5 out of 9 studies were rated as unclear risk (4 studies) and high risk (1 study) with concerns regarding randomization, allocation concealment, blinding, incomplete and selective outcome reporting, and other risk of bias (such as baseline imbalance across groups).

c. High statistical heterogeneity observed, however, the direction of effect is consistent across most studies with overlapping confidence intervals and statistical heterogeneity is likely due to small versus large effects observed across studies.

d. The sample size is adequate (=>300) in both intervention and control arms and effect estimate is precise (Confidence intervals do not include the no effect value "0").

e. Santabalbina, 2016; Daniel, 2012; Gine-Garriga, 2010; Binder, 2002; de Jong, 2000

f. 3 out of 5 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

g. The direction of effect is not consistent and confidence intervals do not overlap with substantial level of statistical heterogeneity observed across studies.

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.

h. The sample size is not adequate (<300) in each arm, however, effect estimate is precise with confidence intervals not including the no effect value of "0". i. Santabalbina, 2016; Pin Ng, 2015

j. 1 out of 2 studies were rated as unclear risk with concerns regarding randomization, allocation concealment and other risk of bias (such as baseline imbalance across groups).

k. Santabalbina, 2016; Kwon, 2015; Gine-Garriga, 2013; Binder, 2002; de Jong, 2000

l. 3 out of 5 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

m. Losa-Reyna, 2019; Santabalbina, 2016; Pin Ng, 2015

n. 2 out of 3 studies were rated as unclear risk with concerns regarding blinding, incomplete outcome reporting, and other risk of bias (such as baseline imbalance across groups).

o. Liu, 2018; Santabalbina, 2016; Pin Ng, 2015

p. The confidence intervals overlap with low statistical heterogeneity observed across studies.

q. The sample size is adequate (=>300) in both intervention and control arms and effect estimate is precise (Confidence intervals do not include the no effect value "1").

r. Santabalbina, 2016

s. The study had concerns regarding randomization, allocation concealment and other risk of bias (such as baseline imbalance across groups).

t. The sample size is not adequate (<300) in each arm and effect estimate is imprecise with confidence intervals including the no effect value of "0". u. Pin Ng, 2015

v. The sample size is not adequate (<300) in each arm and observed number of events are very low with imprecise effect estimate (wide confidence intervals including the no effect value of "1"). w. Liu, 2017; Pin Ng, 2015

x. 1 out of 2 studies rated as high risk with concerns regarding randomization, allocation concealment, blinding and other risk of bias (such as baseline imbalance across groups).

Appendix 5, as supplied by the authors. Appendix to: Racey M, Ali MU, Sherifali D, et al. Effectiveness of physical activity interventions in older adults with frailty or prefrailty: a systematic review and meta-analysis. *CMAJ Open* 2021. DOI:10.9778/cmajo.20200222.