Supplementary File 1

A

Table 1. Overview of Evidence Level

Hierarchy	EL Criteria
1	Systematic reviews of randomized trials or n-of-1 trials
2	Randomized trials or observational studies with dramatic effect
3	Nonrandomized controlled cohort or follow-up studies
4	Case-series, case–control studies, or historically controlled studies
5	Mechanism-based reasoning

EL = Evidence Level

B

Table 2. Quality Assessment

Item	Description
1	Specification of eligibility criteria
2	Random allocation of subjects
3	Concealed allocation of subjects
4	Similar groups at baseline
5	Blinding of all subjects
6	Blinding of all therapists
7	Blinding of all assessors
8	Key outcome measures obtained from 85% of subjects
9	Treatment received as allocated or “intention to treat”
10	Between-group statistical comparisons reported for one key outcome
11	Point measures/variability measures provided for one key outcome

C

Table 3. Adapted version of a quality assessment tool used for quantitative research (Barry et al., 2014)

Question	Scoring
1. Are the research objectives clearly stated?	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
2. Is the study design clearly stated? Type of trial, number of arms.	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
3. Are participant characteristics described in detail? (Mean Age, Male/Female)	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
4. Are inclusion and exclusion criteria stated? (Health Status, ambulatory ability)	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
5. Is sample size justified? (Why this number?)	1 – Yes, 0 – No
6. Is randomization of groups explained? (not just stated)	1 – Yes, 0 – No
7. Are the location/settings described? (Clinical Laboratory, community based)	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
8. Is instrumentation clearly described? (Consoles, Games, Game duration, levels, scoring)	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
9. Are outcome measures described in detail? (Primary, secondary, tertiary)	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No

10. Is duration and intensity of intervention explained?	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
11. Are the exergaming and if used, other exercise stated in detail?	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No
12. Is post intervention follow up used? If so, is it described?	1 – Yes, 0.5 – Yes, lacking detail or clarity, 0 – No

D

Table 4. Articles included in this review

Author	Year	Title
Pluchino et al	2012	Pilot Study Comparing Changes in Postural Control After Training Using a Video Game Balance Board Program and 2 Standard Activity-Based Balance Intervention Programs
Ray et al	2012	The Effects of a 15-Week Exercise Intervention on Fitness and Postural Control in Older Adults
Toulotte et al	2012	Wii Fit® training vs. Adapted Physical Activities: which one is the most appropriate to improve the balance of independent senior subjects? A randomized controlled study
Lai et al	2013	Effects of interactive video-game based system exercise on the balance of the elderly
Singh et al	2013	Effects of balance-focused interactive games compared to therapeutic balance classes for older women
Chow and Mann	2015	Effect of Cyber-Golfing on Balance Amongst the Elderly in Hong Kong: A Pilot Randomised Trial
Merriman et al	2015	Successful balance training is associated with improved multisensory function in fall-prone older adults
Sato et al	2015	Improving Walking, Muscle Strength, and Balance in the Elderly with an Exergame Using Kinect: A Randomized Controlled Trial
Whyatt et al	2015	A Wii Bit of Fun: A Novel Platform to Deliver Effective Balance Training to Older Adults
Nicholson et al	2015	Six weeks of unsupervised Nintendo Wii Fit gaming is effective at improving balance in independent older adults
Park et al	2015	The effects of virtual reality game exercise on balance and gait of the elderly
Tange et al	2012	A pilot with Exergames in Elderly Homes

E

Table 5 Results of Evidence Level of included articles

Author and Date	Hiearchy	Evidence Level
Pluchino et al., 2012	RCT (PS)	2
Ray et al., 2012	RCT	2
Toulotte et al., 2012	RCT	2
Merriman et al., 2015	RCT	2
Sato et al., 2015	RCT	2
Whyatt et al., 2015	RCT	2
Lai et al., 2013	RCT	2
Singh et al., 2013	RCT	2
Chow and Mann., 2015	RCT (PS)	2
Nicholson et al., 2015	Non - RCT	3
Park et al., 2015	Non - RCT	3
Tange et al., 2012	Non – RCT (PS)	3

F

Table 6 Outcomes from custom designed quality assessment tool

Author and Date	Research objectives clearly stated	Study Design Clearly Stated	Participant Characteristics Detailed	Inclusion/ Exclusion Criteria Stated	Outcome measures Described	Sample size justified
	1= yes, 0.5= yes lacking detail, 0=no	1= yes, 0.5= yes lacking detail, 0=no	Number, Age, Sex	1= yes, 0.5= yes lacking detail, 0=no	1= yes 0 =no	1= yes 0 =no
(Pluchino et al., 2012)	1	1	1,1,1	1	1	1
(Ray et al., 2012)	1	0.5	1,1,1	0.5	0.5	0
(Toulotte et al., 2012)	1	0.5	1,1,1	1	1	0
(Merriman et al., 2015)	1	1	1,1,1	0.5	1	0
(Sato et al., 2015)	0.5	0.5	1,1,0.5	0.5	1	1
(Whyatt et al., 2015)	1	0.5	1,1,1	0.5	1	0
(Lai et al., 2013)	0.5	1	1,1,1	0.5	1	0
(Singh et al., 2013)	1	0	1,1,1	0.5	1	0
(Chow and Mann,2015*)*	0.5	0	1,1,1	0	1	Convenience sample
(Nicholson et al., 2015)	1	1	1,1,1	0.5	1	Convenience sample
(Park et al., 2015)	1	0	1,1,1	0.5	1	0
(Tange et al., 2012)	0.5	1	1,1,0	0.5	1	0

F

Table 6. Continued…

Author and Date	Baseline and Post test data presented	Randomization of groups explained	location/settings described	Exergames instrumentation explained (console used, games used	Is duration and intensity of intervention explained?	Are the exergaming and if used, other exercise stated in detail?	Was post intervention follow up used?
	1= yes, 0.5= yes lacking detail, 0=no	1= yes, 0.5= yes lacking detail, 0=no	1= yes 0 =no	1= yes, 0.5= yes lacking detail, 0=no	1= yes, 0.5= yes lacking detail, 0=no	1= yes, 0.5= yes lacking detail, 0=no	1= yes 0 =no
(Pluchino et al., 2012)	1	1	1	1	1	1	0
(Ray et al., 2012)	0.5	0	1	0.5	1	0.5	0
(Toulotte et al., 2012)	0.5	0.5	1	1	1	1	0
(Merriman et al., 2015)	1	0.5	1	1	1	1	0
(Sato et al., 2015)	1	1	0	1	1	1	0
(Whyatt et al., 2015)	1	0	0	1	1	1	0
(Lai et al., 2013)	1	0.5	0	0.5	1	0.5	0
(Singh et al., 2013)	1	0	1	0.5	1	1	0
(Chow and Mann,2015)	0	0	0	0.5	1	1	0
(Nicholson et al., 2015)	1	0	1	1	1	0.5	0
(Park et al., 2015)	1	0	0	1	1	0.5	0
(Tange et al., 2012)	0.5	0	0	0.5	1	0	0

G

Table 7 Standardised Mean Difference of included studies

Study	Std. Mean Difference (Hedges g)	95% CI		Weight (%)	I^2 (%)
Rating Scales					
Lai et al. (BBS)	-0.67	-1.19	-0.15	14.2	
Merriman et al. (BBS)	0.57	0.11	1.03	14.7	
Pluchino et al. (Tinetti Dynamic)	0.54	-0.25	1.33	11.9	
Pluchino et al. (Tinetti Static)	0.40	-0.38	1.18	11.9	
Sato et al. (BBS)	-0.01	-0.53	0.51	14.2	
Toulotte et al. (Tinetti Dynamic)	-2.14	-3.35	-0.92	8.5	
Toulotte et al. (Tinetti Static)	-1.24	-2.28	-0.21	9.8	
Whyatt et al. (BBS)	-0.46	-0.90	-0.02	14.8	
Overall	*-0.27*	*-0.78*	*0.23*	*100.0*	*80*
Reaching Tasks					
Chow and Mann. (FRT)	-0.20	-1.08	0.68	13.0	
Nicholson et al. (FRT)*	-0.22	-0.84	0.39	18.2	
Nicholson et al. (LRT-L)*	-0.98	-1.63	-0.33	17.4	
Nicholson et al. (LRT-R)*	-0.55	-1.18	0.08	18.0	
Pluchino et al. (FRT)	-0.22	-0.99	0.56	14.9	
Ray et al. (Sit and Reach Test)	0.49	-0.11	1.09	18.6	
Overall	*-0.28*	*-0.70*	*0.15*	*100.0*	*57*
Timed Tasks					
Chow and Mann (OLSEO)	-0.22	-0.10	0.66	6.1	
Chow and Mann (TUG)	0.16	-0.72	1.04	6.1	
Lai et al. (OLSEO)	-0.27	-0.78	0.24	10.7	
Lai et al. (TUG)	0.47	-0.04	0.99	10.6	
Nicholson et al. (OLSEO-L)*	-0.82	-1.46	-0.18	8.7	
Nicholson et al. (OLSEO-R)*	-0.40	-1.02	0.22	9.0	
Nicholson et al. (TUG)*	0.55	-0.08	1.18	8.9	
Park et al. (TUG)*	0.25	-0.47	0.97	7.8	
Pluchino et al. (OLSEO)	0.49	-0.29	1.28	7.0	
Pluchino et al. (TUG)	-0.19	-0.97	0.58	7.1	
Ray et al. (8ftTUG)	0.21	-0.38	0.80	9.5	
Singh et al. (TUG)	-0.65	-1.32	0.02	8.3	
Overall	*-0.03*	*-0.30*	*0.24*	*100.0*	*50*
Self-Report Measures					
Lai et al. (MFES)	-0.36	-0.87	0.15	15.5	
Merriman et al. (ABC)	-0.34	-0.79	0.11	19.6	
Merriman et al. (FES)	-0.07	-0.52	0.38	19.9	
Nicholson et al. (I-FES)	0.00	-0.61	0.61	10.7	
Pluchino et al. (FES)	0.33	-0.45	1.11	6.7	
Pluchino et al. (FROP-COM)	-0.56	-1.34	0.23	6.5	
Whyatt et al. (ABC)	-0.40	-0.84	0.04	21.1	
Overall	*-0.23*			*100.0*	*0*

BBS = Berg Balance Scale; FRT = Functional Reach Test; OLSEO = One Leg Stance Eyes Open, L = Left & R = Right; TUG = Timed Up and Go, FT = Foot; MFES = Modified Falls Efficacy Scale; ABC = Activities Specific Balance Confidence Scale; FES = Falls Efficacy Scale; I-FES = Iconographic Falls Efficacy Scale; FROP-COM = Falls Risk for Older People living in the Community; Hedges g (random effects); CI = confidence Interval; * = Non-randomisd study.

H

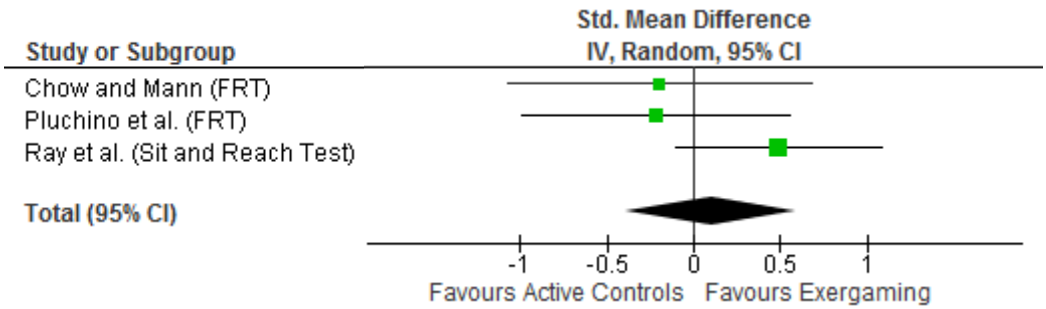

Figure 1. Outcome measures using reaching tasks for Exergaming vs. active controls excluding non-RCTs. FRT = Functional Reach Test; Std. = standardised; IV = inverse variance; CI = confidence interval.

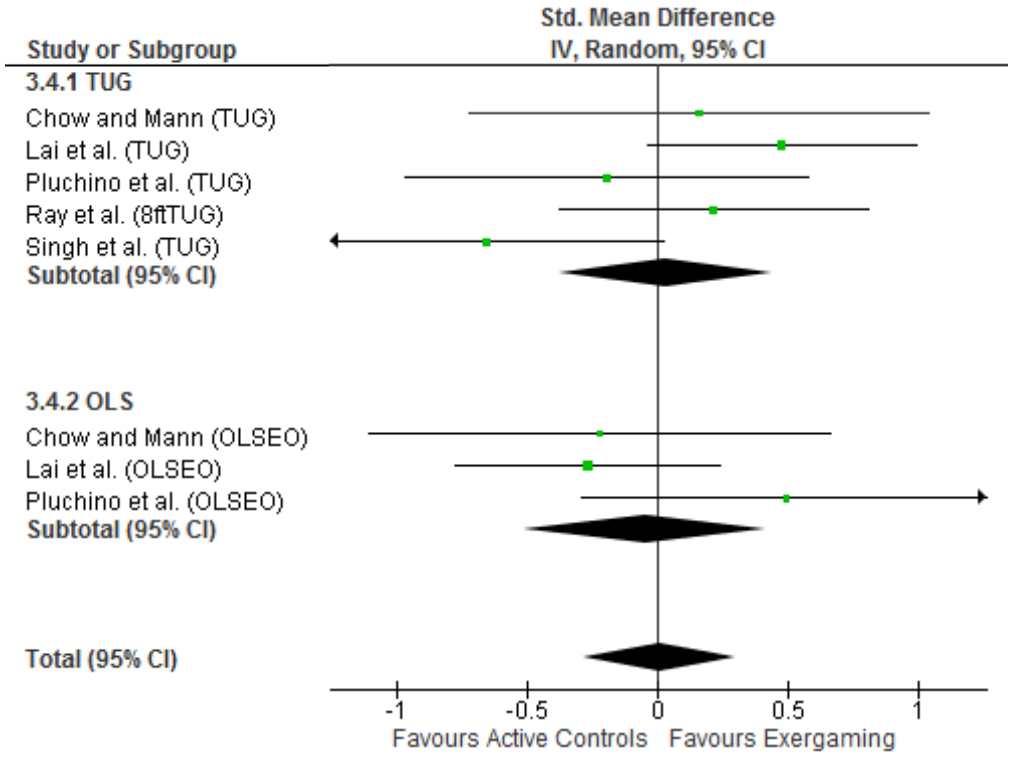

Figure 2. Outcome measures using timed tasks for Exergaming vs. active controls excluding non-RCTs. TUG = Timed Up and Go; OLSEO = One Leg Stance Eyes Open; Std. = standardised; IV = inverse variance; CI = confidence interval.

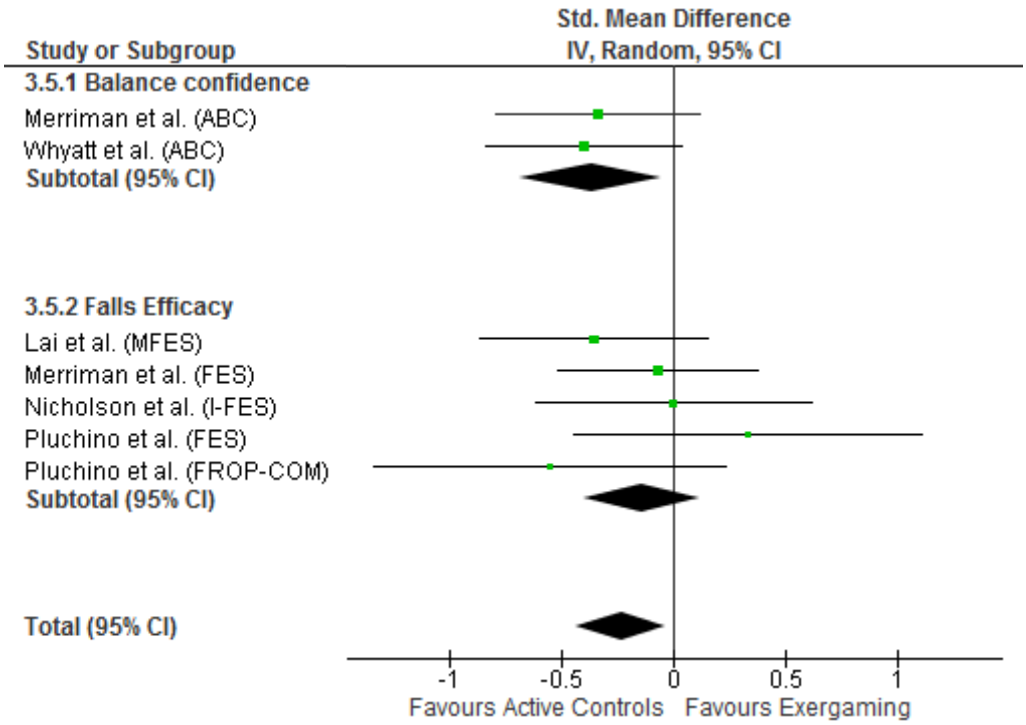

Figure 3. Self-Report Measures of balance confidence and fear of falling for Exergaming vs. active controls excluding non-RCTs. FES = Falls Efficacy Scale; ABC = Activities-specific Balance Confidence Scale; FROP-COM = Falls Risk for Older People living in the Community; I = Iconographical and M = Modified; Std. = standardised; IV = inverse variance; CI = confidence interval.

I. Findings from primary and secondary OMs with insufficient data to pool into meta-analysis.

One trial used a version of the single leg stance test and counted the number of times an individual's foot touched the floor for both eyes open and closed conditions [44], one trial [40] assessed gait speed and found a significant ($P < 0.01$) improvement of 0.04 m/s^{-1} in the exergaming group. Another trial performed a distance-based 6-minute walk test, finding no pre–post differences ($p = 0.455$) or training group differences ($p = 0.705$) [41]. Three studies performed chair stands by number of repetitions in 30 seconds. Ray et al. [53] found improvements from pre to post test scores for both active groups, however, no differences were found between an exergaming and group fitness training group ($p = 0.320$). Sato et al. [58] found significant differences in chair stand repetitions in an exergame group versus controls performing everyday activities (MD: 6.50, 95% CI 4.46 to 8.54, $p < 0.01$). Nicholson et al. [52] found no significant difference between exergame and everyday activities control groups.

J. Details of Intervention effect from Tertiary Outcome Measures.

Pluchino et al. [39] found no significant group x time interactions for sany COP measures, and effect sizes were small ranging from $n^2 = .005$ to .183. Lai et al. [47] found a significant improvement in bipedal sway velocity (SV) of the COP with eyes open and closed ($p < 0.05$) and a trend of decrease in bipedal sway area (SA) of the COP for one of the groups performing the exergame phase of the intervention when compared to a 'no exercise' phase ($p < 0.05$). Singh et al. [45] found no significant differences were observed between exergame and therapeutic balance exercise groups in overall performance index (OPI) of the COP ($F = 0.66$, $p = 0.42$). Park et al. [48] found significant reductions in SV and sway length (SL) for both exergame and ball exercise groups ($p < 0.05$) and SL for the exergame group had a greater reduction than the ball exercise group ($p < 0.05$). Two trials used the Wii Balance Board with a custom designed game targeting static and dynamic balance measures based on time spent in specific target areas (static) and number of target areas obtained within a given time frame (dynamic) [42,43]. The COP was used to control the position visible on-screen, in real time. Merriman et al. [42] found a significant interaction between test group and assessment (F (1, 72) = 13.44, $p < 0.001$) as following the exergame/control period, the exergame group performed significantly better than the control group ($p = 0.003$). Whyatt et al. [43] found a pronounced improvement in the exergaming group for performance in levels of COP excursion. For dynamic

balance, Merriman et al. [42] found the exergame group significantly out-performing controls, as did Whyatt et al. [43] ($F(1, 80 = 39.54, p < 0.001)$, $n^2 = 0.331$). Toulotte et al. [44] calculated the percentage change of the COG on a Wii Balance Board™ and was pronounced for the exergaming group (61% change score) compared to an adapted physical activities (APA) group. It should also be noted that a third group (APA+ exergaming) also improved significantly (44% change score). Ray et al. [41] found no significant differences between exercise groups (Wii Fitness™ and Group fitness) for a Sensory Organisation Test (SOT). Pluchino et al. [39] used a perturbation platform to quantify dynamic balance and found no significant group by time interactions for any DP variables and effect sizes ranged from $n^2 = .003$ to .041 for exergaming versus formal class-based exercise (Tai Chi and Standardized balance training program) suggesting exergaming is as effective as alternative balance programs.