Supplementary Data 1

Example of Search Strategy for SportDiscus	TI "older adul*" OR AB "older adul*" OR TI "elderly" OR AB "elderly" OR TI "aged" OR AB "aged" OR TI "retired" OR AB "retired" OR TI "geriatric" OR AB "geriatric" OR TI "older peopl*" OR AB "older peopl*" OR TI "older populatio*" OR AB "older populatio*" OR TI "older perso*" OR AB "older perso*" OR TI "older wom?n" OR AB "older wom?*" OR TI "older men" OR AB "older men*" OR TI "mature" OR AB "mature" OR TI "senio*" OR AB "senio*" AND TI "age stereotyp*" OR AB "age stereotyp*" OR TI "aging stereotyp*" OR AB "aging stereotyp*" OR TI "ageing stereotyp*" OR AB "ageing stereotyp*" OR TI "ageism" OR AB "ageism" OR TI "agism" OR AB "agism" OR TI "priming" OR AB "priming" OR TI "positive priming" OR AB "positive priming" OR TI "negative priming" OR AB "negative priming" OR TI "implicit priming" OR AB "implicit priming" OR TI "explicit priming" OR AB "explicit priming" OR TI "positive stereotyp*" OR AB "positive stereotyp*" OR TI "negative stereotyp*" OR AB "negative stereotyp*" OR TI "implicit stereotyp*" OR AB "implicit stereotyp*" OR TI "counterstereotyp*" OR AB "counterstereotyp*" OR TI "counter stereotyp*" OR AB "counter stereotyp*" OR TI "counter-stereotyp*" OR AB "counter-stereotyp*" OR TI "self regula*" OR AB "self regula*" OR TI "self-regula*" OR AB "self-regula*" OR TI "views on aging" OR AB "views on aging" OR TI "views on ageing" OR AB "views on ageing" OR TI "stereotype boost" OR AB "stereotype boost" OR TI "stereotype threat" OR AB "stereotype threat" OR TI "stereotype embodiment" OR AB "stereotype embodiment" OR TI "stereotype internali?ation" OR AB "stereotype internali?ation" OR TI "positive vie*" OR AB "positive vie*" OR TI "negative vie*" OR AB "negative vie*" OR TI "imagery" OR AB "imagery" OR TI "imagined contact" OR AB "imagined contact" OR TI "intergenerational contact" OR AB "intergenerational contact" OR TI "intergroup contact" OR AB "intergroup contact" OR TI "self perceptio* of aging" OR AB "self perceptio* of aging" OR TI "self perceptio* of ageing" OR AB "self perceptio* of ageing" OR TI "self-perceptio* of aging" OR AB "self-perceptio* of aging" OR TI "self-perceptio* of ageing" OR AB "self-perceptio* of ageing" OR TI "subjective aging" OR AB "subjective aging" OR TI "subjective ageing" OR AB "subjective ageing" OR TI "stereotype prejudice" OR AB "stereotype prejudice" OR TI "stereotype discrimination" OR AB "stereotype discrimination" OR TI "social interactio*" OR AB "social interactio*" OR TI "social support" OR AB "social support" AND TI "physical activity" OR AB "physical activity" OR TI "exercise" OR AB "exercise" OR TI "physical function" OR AB "physical function" OR TI "physical performance" OR AB "physical performance" OR TI "motor performance" OR AB "motor performance" OR TI "motor learning" OR AB "motor learning" OR TI "well-being" OR AB "well-being" OR TI "wellbeing" OR AB "wellbeing" OR TI "well being" OR AB "well being" OR TI "quality of life" OR AB "quality of life" OR TI "cognition" OR AB "cognition" OR TI "cognitive functio*" OR AB "cognitive functio*" OR TI "memory" OR AB "memory" OR TI "functional capacity" OR AB "functional capacity" OR TI "health related variabl*" OR AB "health related variabl*" OR TI "health-related variabl*" OR AB "health-related variabl*" OR TI "self-worth" OR AB "self-worth" OR TI "self worth" OR AB "self worth" OR TI "anxiety" OR AB "anxiety" OR TI "physical recovery" OR AB "physical recovery" OR TI "self related health" OR AB "self related health" OR TI "self-related health" OR AB "self-related health" OR TI "self esteem" OR AB "self esteem" OR TI "self-esteem" OR AB "self-esteem" OR TI "self efficacy" OR AB "self efficacy" OR TI "self-efficacy" OR AB "self-efficacy" OR TI "motivation" OR AB "motivation" OR TI "obesity" OR AB "obesity" OR TI "subjective aging" OR AB "subjective aging" OR TI "subjective ageing" OR AB "subjective ageing" OR TI "age stereotyp*" OR AB "age stereotyp*" OR TI "aging stereotyp*" OR AB "aging stereotyp*" OR TI "ageing stereotyp*" OR AB "ageing steretyp*" OR TI "views on aging" OR AB "views on aging" OR TI "views on ageing" OR AB "views on ageing" OR TI "self perceptio* of aging" OR AB "self perceptio* of aging" OR TI "self perceptio* of ageing" OR AB "self perceptio* of ageing" OR TI "self-perceptio* of aging" OR AB "self-perceptio* of aging" OR TI "self-perceptio* of ageing" OR AB "self-perceptio* of ageing" OR TI "perceptio* of aging" OR AB "perceptio* of aging" OR TI "perceptio* of ageing" OR AB "perceptio* of ageing"
Database Specific Restrictions	Scopus – Subject headings – Medicine, Nursing, Psychology, Social Sciences, Health Professions, Undefined. MEDLINE – MeSH terms and keyword searches were explored during preliminary searches, however, due to the quality of citations generated and the number of specific terms not covered by MeSH, a combined keyword and MeSH term search was run.

Secondary and Grey Literature Search Strategies	Google Scholar and Ethos were searched for grey literature; full text articles retrieved were hand searched via reference checking and forward and backwards citation screening/snowballing, and members of the Society for Personality and Social Psychology group were contacted to identify any additional studies.
Study Selection Process	All database and secondary searches were conducted by RK, who, following the removal of duplicates via Endnote X8 (Clarivate Analytics, US), uploaded the retrieved articles to Covidence and screened the titles and abstracts, coding "yes", "no", "maybe". A screening tool developed by RK, based on the inclusion criteria, was piloted on 5% of the articles; a second reviewer (JH) independently reviewed the screened titles for discrepancies to ensure there was no discordance between reviewers, or with regards to inclusion/exclusion criteria application. On completion of the screening process, a further 5% of studies and all articles coded "maybe" were reviewed by JH. The two reviewers independently reviewed all articles retrieved in full text against the pre-defined inclusion/exclusion criteria. Disagreements regarding were resolved by discussion with a third reviewer (AC) (initially $k = 0.75$, following discussion $k = 1$).

Supplementary Data 2

Information to Support Risk of Bias Assessments Non-Randomized Articles

	Outcome Domain	Potential Confounders	Bias due to confounding	Bias in selection of participants	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of the outcome	Bias in selection of the reported result	OVERALL JUDGEMENT
Belgrave (2011)	Psychosocial Well-Being	Contact with Grandchildren Baseline health status	Non-equivalent CG - potential differences in health status, cognition and contact with grandchildren Recruitment method No evidence of control of any potential confounders but did do baseline comparisons	Experimental and control groups were from different residential facilities Only 1 x 10-week intervention period Cannot rule out bias completely as unclear when participants allocated to groups - could be low but methods not really comparable to RCT	No serious concerns but not exactly comparable to an RCT	No indication of any deviations beyond what would be expected in usual practice	At least 95% available details of 1 x participant that deceased noted No serious concerns but not exactly comparable to an RCT	Researcher assessed and delivered sessions, self-report outcome measure	Only pre-post measures taken or one outcome measure	Serious risk of bias from confounding, no adjustment and lack of blinding
Brothers & Diehl (2017)	Physical Activity Age Stereotypes	Gender Age Health Status Baseline exercise status	No comparator group but does not appear to be any	No concerns	No concerns	No concerns	Data for 11 participants not included	Measure could have been influenced as measure is self-report	No information on multiple outcome measurements but no indication of multiple analyses	Based on multiple scores of moderate risk
Fujiwara et al. (2009)	Physical Function Subjective Health	Education Grandchildren Other volunteering	Potential from contact with grandchildren, other volunteering activities only education controlled for in analysis	No concerns	No concerns	Participants changed from control to intervention group but not likely to have affected the outcome as excluded from analysis	Data from multiple withdrawals not included but even though there are exclusions they have been identified	Outcome measures could not have been influenced by knowledge of intervention received but no information on whether outcome assessors aware of participant allocation	Lack of information	Based on confounder control plus multiple scores of moderate risk
Sakurai et al. (2016)	Physical Function Psychological Well-Being Subjective Health	Education Grandchildren Other volunteering	Potential from contact with grandchildren, other volunteering activities Only education controlled for in analysis	No concerns	No concerns	Participants changed from control to intervention group but not likely to have affected the outcome as excluded from analysis	Data from multiple withdrawals not included but even though there are exclusions they have been identified	Multiple participant led outcomes Participant led self-rated	Lack of information	Based on confounder control plus multiple scores of moderate risk

Information to Support Risk of Bias Assessments Randomized Articles

	Outcome Domain	Risk arising from randomization process	Risk due to deviations from the intended interventions	Risk due to missing outcome data	Risk in measurement of the outcome	Risk in selection of the reported result	OVERALL JUDGEMENT
Beyer et al. (2019)	Physical Function	Participants were randomly assigned to the IG or the CG using the software R (R Development Core Team, 2015) via the function 'sample' of the R 'base' package with predefined sizes for both groups. However, an additional non-randomized group was added at a later date	Participants aware of assigned interventions. The self-perceptions of aging intervention were delivered by a separate psychologist who did not have contact with the control group. Appropriate analysis used to estimate effect of assignment to the intervention	No apparent missing data	No information on whether outcome assessors aware of the intervention received	No information on analysis plan but results for all time points reported	Based on outcome of risk arising from the randomization process
	Psychological Well-Being			Missing data is described for the questionnaire but unclear if it refers to this measure	Participant led outcome and participants blinded		
	Age Stereotypes			No apparent missing data	Participant led outcome and participants blinded		
Emile et al. (2014)	Quality of Life	Participants randomly divided into two groups, but any further methods unclear	Participants aware of intervention, not possible to hide allocation from a non-intervention control group. Concerns over failure to analyze participants in their allocated groups - 34 randomly divided into two groups but total results only reported for n = 52	Potentially 18% data missing, with no explanation	Participant led outcome and participant aware of intervention	Insufficient information, appears all outcome measurements reported, but unclear about volume of analyses	Based on two high risk domains
Emile et al. (2014)	Physical Activity				Appears that the intervention and the outcome assessment are delivered by the same person. Judged to be some concerns not high risk as unclear as influence of intervention knowledge could be different for the objective and subjective measures		
Emile et al. (2014)	Psychological Well-Being				Participant led outcome and participant aware of intervention		
Klusmann et al. (2012)	Age Stereotypes	259 women met the eligibility criteria, were included in the baseline assessment, and then were randomized. Of these, 247 women were allocated to one of three study groups, that is, a physical exercise course, an active control (i.e., a computer course), or a passive control group - no information on method of randomization. Not enough information on baseline data	Participants probably unaware of intervention allocation as debriefed at end of study but intervention deliverers probably aware. Intention to treat analysis used. No information to judge deviation from intended intervention	> 95% of sample reported	Participants are outcome assessors and it is likely they knew their allocation but not aware of actual reason for study as debriefed at end	No indication analysis planned in advance but no concerns over lack of measurement reporting or multiple data analysis	Based on average risk of some concerns

Levy et al. (2014)	Age Stereotypes		Participants unaware of the hypotheses and to which group they had been assigned	No apparent missing data	Participants are outcome assessors and it is likely they knew their allocation but not aware of actual reason for study as debriefed at end	No indication analysis planned in advance but no concerns over lack of measurement reporting or multiple data analysis	Based on average risk of some concerns
Levy et al. (2014)	Physical Function	Statement on lack of differences between baseline measures in intervention groups, but no details on method of randomization	Three experimenters, two unaware of the hypotheses (tested 90% of participants) and the nature of condition assignment; the pattern of significant results did not differ between the three experimenters. No information to conclude if appropriate analysis. No concerns that any participant changed groups or were lost to follow up	No apparent missing data	Comparable timepoints (see Table 1) and assessors Three experimenters, two unaware of the hypotheses (tested 90% of participants) and the nature of condition assignment; the pattern of significant results did not differ between the three experimenters. Assessment consists of physical participant led measures	No information on analysis plan on measure of physical activity but whilst data for all groups are not reported, a 2x2 design has been utilized and data presented relative to all noted hypotheses	Based on average risk of some concerns
Warner et al. (2016)	Physical Activity	The software R (http://cran.r-project.org) via the function 'sample' of the R package 'base' was used to randomize participants into three groups using pre-defined group sizes for the intervention group Successful randomization to these four groups supported by analyses of variance and chi-square tests. No differences occurred between the groups for any demographic variables, and accelerometer-assessed PA, Functional Comorbidity Index or disadvice to be active from a physician at baseline	Minimal information to make judgements. Due to the nature of the intervention those delivering would know who was in the group Reasons for dropouts and Attrition analyses undertaken but no information as to whether ITT or mITT carried out	Unclear, not all reasons for drop out stated (Figure 1 p. 1149, Warner, 2016) Figure 1 reasons for attrition varied – see consort diagram, multiple dropouts due to health-related outcomes	Participant reported and objective measure through accelerometer. No indication that assessment of the outcome influenced by knowledge of the intervention	No information on analysis plan but results for all time points reported, multiple different covariates etc. but non-significant results reported	Judged as some concerns not high-risk following reviewer discussion
Wolff et al. (2014)	Physical Activity	The software R (http://cran.r-project.org) via the function 'sample' of the R package 'base' was used to randomize participants into three groups using pre-defined group sizes for the intervention group	Minimal information to make judgements. Due to the nature of the intervention those delivering would know who was in the group Reasons for dropouts and Attrition analyses undertaken but no information as to whether ITT or mITT carried out	Unclear, not all reasons for drop out stated See supplementary data; reasons for attrition varied – see consort diagram, multiple dropouts due to health-related outcomes	Participant led outcome and participant aware of intervention	No information on analysis plan but results for all time points reported, multiple different covariates etc. but non-significant results reported. Data for passive control group not reported but this is not what the question is asking	Judged as some concerns not high-risk following reviewer discussion
Wolff et al. (2014)	Age Stereotypes						