

Supplementary file 4. Risk of bias summary.

	Confounding	Selection of participants	Classification of physical activity	Departures from the intended interventions	Missing data	Measurement of mortality	Selective reporting	Overall bias
Ammitzbøll 2016	−	+	?	−	−	+	+	−
Bao 2015	+	+	?	+	+	+	+	+
Bertram 2011	+	+	?	+	+	+	+	+
Bradshaw 2014	+	+	−	+	−	+	+	−
Chen 2011	+	+	?	+	+	+	+	+
de Glas 2014	+	+	?	+	+	?	+	+
Holick 2008	+	−	?	−	+	+	+	−
Holmes 2005	−	+	?	−	+	+	+	−
Irwin 2008	−	+	?	−	+	+	+	−
Irwin 2011	+	+	?	+	+	+	+	+
Maliniak 2018	+	+	?	−	+	+	+	−
Sternfeld 2009	−	−	?	−	−	+	+	−
Glenn 2015	−	+	?	−	+	+	+	−
Gregg 2003	−	+	−	−	+	+	+	−
Hu 2004	−	+	−	−	+	+	+	−
Sluik 2012	−	+	?	−	+	+	+	−
Sone 2013	−	+	?	−	+	+	+	−
Tanasescu 2003	+	+	?	+	?	+	+	+
Cheng 2018	−	+	?	−	+	+	+	−
Garcia-Aymerich 2006	−	+	?	−	+	+	+	−
Gerber 2011	+	+	?	+	+	+	+	+
Janssen 2006	+	+	?	+	+	+	+	+
Lahtinen 2018	+	+	?	+	+	+	+	+
Moholdt 2017	+	+	?	+	?	+	+	+
Mons 2014	−	+	−	+	?	+	+	−
Stewart 2017	−	+	?	−	+	?	+	−
Tian 2017	−	−	−	−	+	+	+	−
Wannamethee 2000	+	+	?	+	+	+	+	+

Low: comparable to a well-performed RCT;
 Moderate: sound evidence but not comparable to a well-performed RCT;
 Serious: the study has some important problems;
 Critical: the study is too critical to provide any useful evidence and should not be included in any synthesis;
 No information: not enough information to reach a judgment.