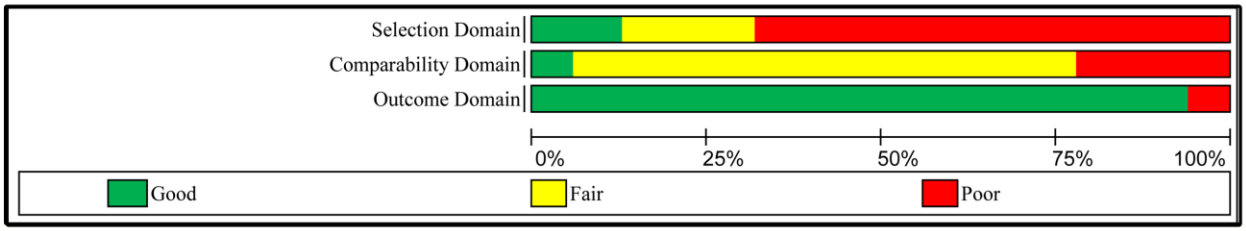
Supplemental Digital Content 4 – Quality Assessment

Study	Selection Domain			Comparability Domain		Outcome Domain		Overall
	Sample representativeness	Sample size	Health status	Physical activity status	Confounders	Main outcome	Statistical test	# of stars
Allman and Rice (2001)			*			*	*	3
Bieuzen et al. (2010)			*	*			*	3
Callahan et al. (2009)			**	*		**	*	6
Callahan and Kent-Braun (2011)	*		**	*		**	*	6
Chan et al. (2000)			**	*	*	*	*	6
Chung et al. (2007)			**	*		*	*	5
Dalton et al. (2010)			*			**	*	4
Dalton et al. (2012)			*			**	*	4
Dalton et al. (2015)			*		*	**	*	5
Ditor and Hicks (2000)			*		*	*	*	4
Griffith et al. (2010)			*	*		**	*	5
Hakkinen (1995b)			*			**	*	4
Hunter et al. (2004)			*	*		*	*	4
Hunter et al. (2005)			*	*		*	*	4
Hunter et al. (2008)			*	*		*	*	4
Justice et al. (2014)	*		**	*		*	*	6
Kent-Braun et al. (2002)	*		**	*			*	5
Lanza et al. (2004)			**	*		**	*	6
Lanza et al. (2005)			*	*	*	*	*	5
McNeil and Rice (2007)			*			**	*	4
Power et al. (2012)			*		*	**	*	5
Senefeld et al. (2017)	*		**	*		**	*	7
Solianik et al. (2017)			*		*	**	*	5
Stackhouse et al. (2001)			*			**	*	4
Thompson et al. (2015)			**		*	**	*	6
Walker et al. (2013)			**		*	*	*	5
Wallace et al. (2016)			*			**	*	4
Yoon et al. (2008)			*	*		*	*	4
Yoon et al. (2012)			*	*		*	*	4
Yoon et al. (2013)			*	*		*	*	4
Yoon et al. (2015)			*	*		*	*	4
Frequency/Mean	**4/31**	**0/31**	**41/62**	**18/31**	**8/31**	**44/62**	**31/31**	**4.68**

Frequency of ratings		
≤ 3 stars	3	10%
4 stars	13	42%
5 stars	8	26%
6 stars	5	16%
≥ 7 stars	2	7%



The mean of points (stars) of the included studies as 4.68 and the median was 4. Selection Domain: 13% “good”, 19% “fair”, and 68% “poor”. Comparability Domain: 6% “good”, 72% “fair”, and 22% “poor”. Outcome Domain: 94% “good”, and 6% “poor”.