Appendix 6. GRADE assessment for the single leg hop test applied to all outcomes*

	Study design	Risk of bias	Inconsistency (I^2)	Indirectness	Imprecision	Publication bias	Odds Ratio	GRADE
Return to Sport								
Ardern et al, 2015 Ebert et al, 2019 Moksnes et al, 2009 Nawasreh et al, 2017 Toole et al, 2017 Faleide et al, 2021 Kitaguchi et al, 2020 Webster et al, 2019 Welling et al, 2020	**(-2)** Prospective	**(-1) Serious:** Most did not control for confounders, had inadequate sample size and poor follow-up	**(-1) Not serious:** I^2 = 70.6% (P= 0.001) Overlapping confidence intervals from 8/10 studies. Odds ratio between 0.94 to 16.44 without large discrepancies in weighting	**(0) Not serious** Similar populations, timepoints of testing and outcomes	**(0)** **Total n= 638** **CI (1.30, 3.54)**	N/A	**(0)** 1.97 (1.24, 3.13)	Very low
Patient-reported symptoms and function								
Cristiani et al, 2020 Culvenor et al, 2016 Ebert et al, 2019 Ericsson et al, 2013 Filbay et al, 2021 Logerstedt et al, 2012 Mansson et al, 2013 McGrath et al, 2017 Oiestad et al, 2012 Stropnik et al, 2020 Welling et al, 2020	**(-2)** Prospective	**(-1) Serious:** Most did not control for confounders, had inadequate sample size and poor follow-up	**(-1) Serious:** I^2 = 69.2% (P= 0.001) Overlapping confidence intervals of most studies with two outlying studies. Odds ratio between 1.28 to 18.95. Discrepancies in weighting.	**(-1) Serious** Differing populations (age), differing timepoints of outcome, and outcomes used (IKDC and KOOS)	**(0)** **Total n=1737** **CI (1.62, 3.88)**	**(-1) Serious** P = 0.002 Eggers test for small study effects (as there were 10 studies)	**(+1)** 2.51 (1.62, 3.88)	Very low
Success with ACL deficiency								
Button et al, 2006 Eitzen et al, 2010 Ekas et al, 2019 Ericsson et al, 2013 Fitzgerald et al, 2000 Grindem et al, 2018 Mosknes et al, 2009	**(-2)** Prospective	**(-1) Serious:** Most did not control for confounders, had inadequate sample size and poor follow-up	**(0) Not serious:** I^2 = 54.1% (P= 0.042) Overlapping confidence intervals from 5/7 studies. Odds ratio between 0.59 to 6.65 without large discrepancies in weighting	**(-1) Serious** Differing outcome definitions	**(0)** **Total n= 228** **CI (0.75, 2.32)**	N/A	**(0)** 1.32 (0.75, 2.32)	Very low
Subsequent knee-injury after ACLR								
Cristiani et al, 2021 Falstrom et al, 2021 Grindem et al, 2016 King et al, 2021 Kyritsis et al, 2016 Webster et al, 2019 Wellstandt et al, 2017	**(-2)** Prospective	**(-1) Serious:** Most did not control for confounders, had inadequate sample size and poor follow-up	**(0) Not serious:** I^2 = 47.2% (P= 0.078)	**(-1) Serious** Different outcome definition for knee-injury or re-injury	**(0)** **Total n= 6970** **CI (0.58, 1.11)**	N/A	**(0)** 0.81 (0.58, 1.11)	Very low

Knee Osteoarthritis								
Filbay et al, 2021 Janssen et al, 2013 Patterson et al, 2018 Pinczewski et al, 2007 Wellstandt et al, 2018	**(-2)** Prospective	**(-1) Serious:** Most did not control for confounders, had inadequate sample size and poor follow-up	**(-1) Serious:** I^2 = 75.8% (P= 0.006)	**(-1)** ? maybe high due to timing of outcome (and therefore population) and outcome definition	**(-1)** **Total n= 222** **CI (0.70, 3.98)* wide**	N/A	**(0)** 1.67 (0.70, 3.98)	Very low

Grade of evidence was assigned using the GRADE system, which has 4 categories HIGH, MODERATE, LOW or VERY LOW. Evidence is initially assigned as HIGH from randomised trials. The grade of evidence was then reduced if there was serious (-1) or very serious (-2) limitations to study quality or uncertainties about directness of association; important inconsistency (-1), imprecise or sparse data (-1) or a high probability of reporting bias (-1). Grade of evidence was increased if strong evidence of association was seen (e.g., RR >2 or <0.5) from ≥2 observational studies with no plausible confounders (+1) or very strong direct evidence (RR >5 or <0.2) with no major threats to validity (+2); if there was evidence of a dose-response gradient (+1) or if all plausible confounders would have reduced the effect/association seen (+1). The interpretation of GRADE evidence assessments is that for HIGH certainty evidence further research is very unlikely to change our confidence in the estimate of effect; for MODERATE certainty evidence further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate; for LOW certainty evidence further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate; and for VERY LOW certainty evidence any estimate of effect is very uncertain.

*Only level of evidence of the single-forward hop test was assessed as this was used across all outcomes and studies. The level of evidence was found to be very low across all outcomes using this test and so we decided there was not need to complete this assessment for other tests as they would likely yield the same result and have less data from which to draw conclusions of evidence certainty.

KOOS, Knee osteoarthritis outcome score, IKDC, international knee documentation committee score