Appendix 2. Risk of Bias Tool

Modified Newcastle-Ottawa Scale from Wells, G. et al, The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp* 2013;

Selection	
1. Definition of ACL injured population Low risk of bias = a High risk of bias = b	a) Clearly described if the inclusion/exclusion criteria of an ACL injured person stated **both** of the following criteria: i) Diagnosed ACL injury with clinical/imaging or surgical confirmation (e.g. Lachman's or pivot shift test ± MRI/arthroscopic confirmation), ii) Reports of surgical or non-surgical management b) Not described OR used minimal criteria for inclusion/exclusion.
2. Source population Low risk of bias = a High risk if = b, c	a) A consecutive sample or random selection from a source population that is well described and representative of the condition under study (e.g. surgeon's clinic, outpatient clinic). b) A consecutive sample or random selection from a population that is not highly representative of the condition under study. c) Cannot be defined or enumerated (i.e. volunteering or self-recruitment).
3. Typical of the average ACL injured population (representativeness of cohort) Low risk of bias = a High risk if = b	a) Truly representative of the average ACL injured person in the community if **all** of the following criteria are present: i) Including men and women, ii) Typical age range at time of ACL injury/surgery (mean age = 16-35), iii) If surgery, then 'typical' surgical procedure (arthroscopic and not synthetic graft*) **If non-surgical management then N/A for this point* b) Above criteria are not present then not truly representative of the average ACL injured population.
4. Sample size Low risk of bias = a High risk if = b, c	a) Power analysis completed and sample size adequate to detect meaningful difference. b) Power analysis completed but sample size not adequate to detect meaningful difference. c) No power analysis completed.

Exposure	
5. Methods for assessment of functional performance (i.e. ascertainment of exposure) Low risk of bias = a High risk if = b, c	a) Well described methods for functional tests - including an appropriately trained or appropriate profession as assessor **AND** describes or cites reliability. b) Well described methods for functional tests including an appropriately trained or appropriate profession as assessor) **OR** describes or cites reliability. c) Not described.
6. Demonstration that outcome of interest was not present at ascertainment of exposure (i.e. outcome that is compared to exposure) Low risk = a High risk if = b	a) **True** if baseline score of outcome of interest for both exposed/non-exposed (poor/good functional performance) has been accounted for (for example as a covariate or change in score or not present at ascertainment of exposure). b) No demonstration that the baseline score of outcome of interest has been accounted for.

Comparability	
7. Comparability of cohorts on the basis of the design or analysis Low risk = a High risk if = b	a) Comparability exists if study cohort (exposed/non-exposed) was a priori matched for **at least one covariate**, or confounding controlled for in statistical analysis. Covariate examples: i) Age, ii) BMI, iii) Sex b) Study not controlled in design or analysis and no confounders acknowledged. (Statements of no differences between groups or that differences were not statistically significant are not sufficient for establishing comparability)

Outcome	
8. Validity and reliability of outcome(s) of interest Low risk = a High risk if = b	a) Outcome measure(s) of interest are clearly described, and references other article(s) which found outcome measure to be valid & reliable **OR** demonstrates the outcome measure(s) of interest are valid and reliable. (note all outcome(s) of interest must be valid and reliable for (a)) b) If outcome measure(s) of interest were not explained in reproducible detail, or validity and reliability not proven/reported.
9. Assessment of outcome(s) of interest Low risk = a High risk if = b	a) Assessor has suitable qualification to interpret findings (e.g. musculoskeletal radiologist) **AND** blind to participant baseline exposure/non-exposure. **N/A: Blinding not needed for self-reported outcomes* b) Poor or no description.
10. Adequacy of follow-up of cohorts Low risk = a High risk if = b	a) Adequacy of follow-up if **either** of the following are satisfied: i) <15% lost to follow up + description of those lost, ii) <5% lost to follow up with no description b) >15% lost to follow up or not explicitly stated with number of participants lost to follow-up **OR** characteristics of those lost to follow-up were not described.

*risk of bias assessed from published paper, not considering extra data if provided by the authors