

Appendix 4: GRADE assessment criteria

As per the recommendations for prognostic research (1), observational longitudinal study designs were ranked as 'high-quality' (rating +4). Outcomes were rated down where there was 'serious concern' (-1 rating) or 'very serious concern' (-2 rating) for measures of risk of bias (completeness of follow-up, study attrition, representativeness of cases), inconsistency (overall trend in trajectory over time, variability in point estimates that would lead to alternative management approaches), imprecision (minority [$<50\%$] of confidence intervals spanned ≥ 20 [out of 100] [-1 point], majority [$\geq 50\%$] of confidence intervals spanned ≥ 20 [out of 100] [-2 points], where variability of ≥ 20 was deemed to be clinically meaningful (2)), indirectness (generalisability of represented samples), and publication bias. Outcomes were rated up if events over time followed a well-defined pattern and estimates fell close to the trend line that was calculated from the pooled data (+1 points if 'followed a pattern', +2 if 'followed a well-defined pattern' (1)).

References:

1. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ : British Medical Journal*. 2015;350:h870.
2. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. Minimal Clinically Important Change for Pain Intensity and Disability in Patients With Nonspecific Low Back Pain. *Spine*. 2007;32(25):2915-20.