Appendix D. Downs and Black tool used for the risk of bias assessment of included cohort studies with guidance.

Category	Item	Guidance
Reporting	Are the characteristics of the study's patients clearly described?	Are the patients' age, gender, and malocclusion described?
	Are the interventions of interest clearly defined?	Are the appliances and the treatment approach (ex/non-ex) described to a minimum?
	Are the distributions of principal confounders in each group clearly described?	Are confounders (DI, Tx time) described in each group?
	Does the study provide estimates of the random variability in the data for the main outcomes?	Is the variability of the estimate given (SD, SE or 95% CI)?
External validity	Were the subjects asked to participate in the study representative of the entire population?	Are patients representative of the average orthodontic patient in terms of age (10-20 years) and sex (%male 30%-60%)
	Can we be confident that finishing quality not used as a patient selection criterion?	Patients must not be selected according to the treatment results.
	Were the staff, places, and facilities where the patients were treated, representative of what the majority of patients receive?	Patients should be treated by a skilled clinician (either orthodontist, last year postgraduate, or experienced clinician)
	Can we be confident that patients were treated?	In this case, treatment is fairly obvious to ascertain
Internal validity -bias	Was the study prospectively planned and conducted?	Prospective design
	Was an attempt made to blind those measuring the main outcome?	Was blinding implemented during model scoring with OGS?
	If any of the results of the study were based on "data dredging", was this made clear?	Are there more than 5 subgroups/comparisons?
	Were the main outcome measures used accurate (valid and reliable)?	For us, it is straightforward, as OGS is very valid.
Power	Did the study have sufficient power to accurately detect an existing effect?	Sample size for the study (cohort) or for each compared group (comparative cohort studies with at least two groups). Cut-offs used to give 0, 1, 2, 3, 4 or 5 points: 0-20, 20-40, 40-60, 60-80, 80-100, >100 patients.

DI, discrepancy index; Tx, treatment; SD, standard deviation; SE, standard error; CI, confidence interval; OGS, objective grading system.