Supplement 3. GRADE summary

Outcome	No. of studies	No. of patients	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Effect (95% CI)	Certainty	Importance
Curve progression and surgery	11	685 (experimental), 668 (control)	Retrospective cohort	Serious	Serious	None	None	High heterogeneity (I^2=92%)	–1.76 (–4.97 to 1.46)	★★★★ (low)	★★★★★ (High)
Curve progression >25 years	6	583	Retrospective cohort	Serious	Very serious	None	None	Extremely high heterogeneity (I^2=100%)	35.68 (22.85 to 48.50)	★★★★ (very low)	★★★★★ (High)
Patients requiring surgery	2	233	Retrospective cohort	Serious	Very serious	None	None	High heterogeneity (I^2=100%)	14.20% (0.87 to 27.53)	★★★★ (low)	★★★★★ (High)
Thoracic kyphosis	3	107 (initial), 112 (final)	Retrospective cohort	Serious	Serious	None	None	Moderate heterogeneity (I^2=65%)	–2.18 (–5.26 to 0.91)	★★★★ (moderate)	★★★★★ (Moderate)
Lumbar lordosis	3	108 (initial), 112 (final)	Retrospective cohort	Serious	Serious	None	None	Substantial heterogeneity (I^2=80%)	3.11 (–3.18 to 9.39)	★★★★ (low)	★★★★★ (Moderate)
Global back pain	2	252	Retrospective cohort	Serious	Very serious	None	None	Extremely high heterogeneity (I^2=100%)	63.35 (38.36 to 88.34)	★★★★ (very low)	★★★★★ (High)

GRADE, Grading of Recommendations Assessment Development and Evaluation; CI, confidence iinterval.