Supplementary Table 2. GRADE quality of evidence assessment for each outcome.

Outcome	Limitations	Indirectness	Imprecision	Inconsistency	Publication bias	MCID	Total number of participants	Conclusion	Quality of evidence
Analgesia									
Pain score at rest at 8–12 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	1523	Control clinically and statistically inferior to diclofenac and indomethacin Indomethacin clinically and statistically superior to celecoxib No other statistical differences between interventions, but, with MCID of 10, clinical differences possible	Low quality (⊕⊕)
Pain score on movement at 8–12 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	506	Control statistically inferior but clinically equivalent to diclofenac No statistical differences between interventions, but, with MCID of 10, clinical differences possible	Low quality (⊕⊕)
Pain score at rest at 24 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	1790	Control clinically and statistically inferior to diclofenac Control statistically inferior but clinically equivalent to tenoxicam No other statistical differences between interventions, but, with MCID of 10, clinical differences possible	Low quality (⊕⊕)
Pain score on movement at 24 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	582	Control clinically and statistically inferior to diclofenac No other statistical differences between interventions, but, with MCID of 10, clinical differences possible	Low quality (⊕⊕)
Pain score at rest at 48 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	571	No statistical differences between interventions, but, with MCID of 10, clinical differences possible	Low quality (⊕⊕)
Pain score on movement at 48 h (0–100)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10	235	Control clinically and statistically inferior to indomethacin No clinical or statistical difference between control and celecoxib + parecoxib	Low quality (⊕⊕)

Need for rescue analgesia (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	Serious publication bias[e]	20%	1586	Indomethacin clinically and statistically superior to celecoxib + parecoxib, diclofenac and ketorolac No other statistical differences between interventions, but, with MCID of 10, clinical differences possible Control statistically and clinically inferior to diclofenac, ketoprofen and tenoxicam Ketoprofen statistically and clinically superior to celecoxib + parecoxib No other statistical differences between interventions, but, with MCID of 20%, clinical differences possible Control clinically and statistically inferior to diclofenac, ketorolac and naproxen	Very low quality (⊕)
Time to first analgesic request (min)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	Serious inconsistency[d]	Serious publication bias[e]	60 min	1076	Diclofenac, ibuprofen, indomethacin and ketorolac clinically and statistically superior to celecoxib No other statistical differences between interventions, but, with MCID of 60 min, clinical differences possible Control clinically and statistically inferior to diclofenac	Very low quality (⊕)
Cumulative intravenous morphine equivalent consumption at 8–12 h (mg)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10 mg	364	No other statistical differences between interventions, but, with MCID of 10 mg, clinical differences possible Control clinically and statistically inferior to diclofenac, indomethacin and ketorolac	Low quality (⊕⊕)
Cumulative intravenous morphine equivalent consumption at 24 h (mg)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	Serious publication bias[e]	10 mg	1228	No other statistical differences between interventions, but, with MCID of 10 mg, clinical differences possible Pairwise comparison only Control clinically and statistically inferior to diclofenac (MD -46.29	Very low quality (⊕)
Cumulative intravenous morphine equivalent	Serious limitations[a]	No serious indirectness	No serious imprecision	Moderate inconsistency[c]	No serious publication bias	10 mg	320		Moderate quality (⊕⊕⊕)

consumption at 48 h (mg)								95% CI -60.71,-31.86; I^2 = 73%; p < 0.0001)	
Cumulative in-hospital intravenous morphine-equivalent consumption (mg)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	10 mg	404	Control clinically and statistically inferior to diclofenac, ketorolac and parecoxib No other statistical differences between interventions, but, with MCID of 10 mg, clinical differences possible	Low quality (⊕⊕)
Side effects									
Rate of postoperative nausea and/or vomiting at 24 h (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	20%	938	No statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Low quality (⊕⊕)
Rate of postoperative nausea and/or vomiting at 48 h (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	Serious publication bias[e]	20%	74	No statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Very low quality (⊕)
Rate of in-hospital nausea and/or vomiting (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	20%	1387	No statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Low quality (⊕⊕)
Rate of pruritis at 24 h (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	20%	293	No statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Low quality (⊕⊕)
Rate of pruritis at 48 h (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	Serious publication bias[e]	20%	74	No statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Very low quality (⊕)
Rate of in-hospital pruritis (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	20%	1043	Ketoprofen statistically and clinically superior to celecoxib + parecoxib No other statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Low quality (⊕⊕)
Rate of sedation at 24 h (%)	Serious limitations[a]	No serious indirectness	Serious imprecision[b]	No serious inconsistency	No serious publication bias	20%	630	Control statistically and clinically inferior to diclofenac No other statistical differences between interventions, but, with MCID of 20%, clinical differences possible	Low quality (⊕⊕)

Rate of sedation at 48 h (%)	-	-	-	-	-	20%	44	In one trial, no statistical difference between control and ketorolac	-
Rate of in-hospital sedation (%)	Serious limitations[a]	No serious indirectness	No serious imprecision	No serious inconsistency	No serious publication bias	20%	559	Pairwise comparison only Control clinically and statistically inferior to diclofenac (RR 0.43 95% CI 0.26, 0.73; I^2 = 13; p = 0.002)	Moderate quality (⊕⊕⊕)
Functional outcomes									
Length of hospital stay (h)	Serious limitations[a]	No serious indirectness	No serious imprecision	No serious inconsistency	No serious publication bias	6 h	317	Pairwise comparison Control statistically inferior but clinically equivalent to diclofenac (MD -0.48 95% CI -0.88,-0.08; I^2 = 0%; p = 0.02) In one trial, no statistical difference between ketorolac and parecoxib	Moderate quality (⊕⊕⊕)

[a]In many comparisons between interventions, some concerns or major concerns were present for the overall risk of bias. Final decision to rate down quality of evidence for serious limitations.

[b]Evidence of serious imprecision as some or most comparisons between interventions have a confidence and/or prediction interval which extend into clinically important or unimportant effects. Final decision to rate down quality of evidence for serious imprecision.

[c]Evidence of moderate imprecision as even though the I^2 was above 50%, the point estimates did not vary widely between studies. Final decision to not rate down quality of evidence for moderate inconsistency.

[d]Evidence of serious inconsistency as most comparisons between different interventions have variability of direct and indirect effects in relation to a clinically important size of effect. Final decision to rate down quality of evidence for serious inconsistency.

[e]Final decision to rate down quality of evidence for serious publication bias.