Search strategy (Ovid MEDLINE(R))

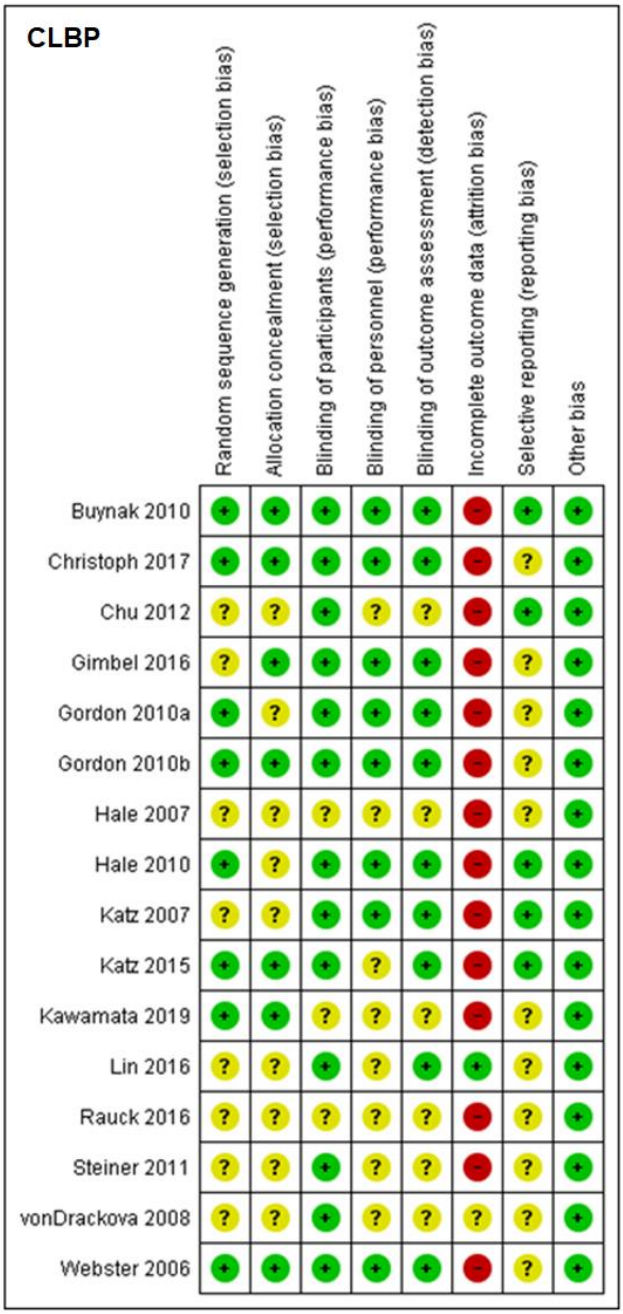
Database(s): Ovid MEDLINE(R) 1946 to June Week 5 2019, Ovid MEDLINE(R) Daily Update July 03, 2019, Ovid MEDLINE(R) Epub Ahead of Print July 03, 2019, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations 1946 to July 03, 2019

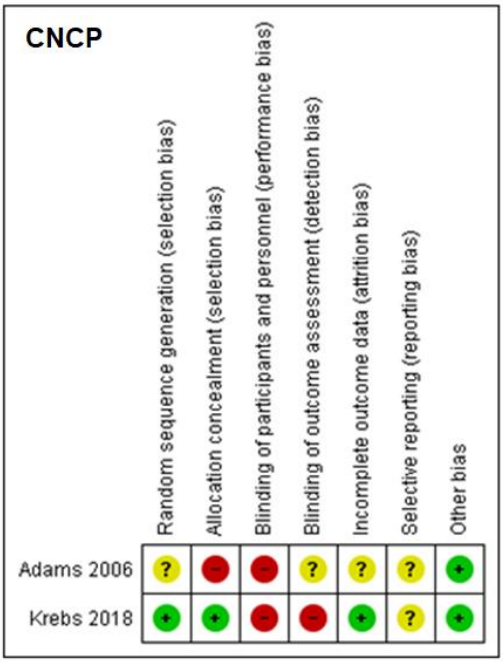
1 exp Analgesics, Opioid/
2 (opioid* or opiate*).mp.
3 Opiate Substitution Treatment/
4 (buprenorphin* or fentan* or hydromorphon* or morphin* or oxycodon*).mp.
5 (butorphanol* or codein* or dihydrocodein* or hydroxycodein* or isocodein* or oxycodein* or dihydrohydroxycodein* or hydrocodon* or hydrocodeinonebitartrat* or meperidin* or methadon* or normethadon* or methadyl acetate or opium or pentazocin* or phenazocin* or tapentadol or tramadol or levomethadon* or methylnaltrexon* or naltrexon* or naloxon* or piritramid* or morphin or morphine or morphina or morphium or beta-casomorphin* or dihydromorphin* or ethylmorphin* or methylmorphin* or morfin* or morphia or morphinium or morphinene or n-methylmorphin* or oxymorphon* or hydromorphon* or heroin* or phentan* or sufentan*).mp.
6 (alfentan* or alphaprodin* or carfentan* or deltorphin* or dextromethorphan* or dezocin* or encephalin* or ethylketocyclazocin* or etorphin* or ketobemidon* or levorphanol or lofentan* or meptazinol or nalbuphin* or phenoperidin* or pirinitramid* or promedol* or propoxyphen* or remifentan* or tilidin* or tapentadol or adolonta or anpec or ardinex or asimadolin* or alvimopam or amadol or biodalgic or biokanol or codinovo or contramal or demerol or dicodid or dihydrone or dilaudid or dinarkon or dolsin or dolosal or dolin or dolantin* or dolargan or dolcontral or duramorph or duromorph or duragesic or durogesic or eucodal or fedotzine or fentanest or fentora or fortral or hycodan or hycon or isonipecain * or jutadol or laudacon or l dromoran or levodroman or levorphan* or levo-dromoran or levodromoran or lexir or lidol* or lydol* or ms contin* or nobligan or numorphan or oramorph or oxiconum or oxycone or oxycontin or palladon* or pancodine or pethidin* or prontofort or robidone or skenan or sublimaze or sufenta or takadol or talwin or theocodin* or tramadol hameln or tramadolor or tramadura or tramagetic or tramagit or tramake or tramal* or tramex or tramundin or trasedal or theradol or tiral or topalgic or tradol or tradolpuren or tradonal or tralgiol or tramadorsch or tramadin or tramadoc or ultram or zamudol or zumalgic or zydol or zytram).mp.
7 or/1-6
8 ((unspecific or unspecified or "not-specified" or "not further specified") adj3 pain*).mp.
9 ((noncancer* or non-cancer* or recurrent or non-malign* or non-tumo* or refractory) adj3 pain*).mp.
10 exp Back pain/
11 (back pain* or backpain* or backache* or back-ache*).mp.
12 or/8-11
13 chronic*.mp.
14 exp Chronic Disease/
15 13 or 14
16 12 and 15
17 Pain, intractable/
18 (intractable adj3 pain*).mp.
19 17 or 18
20 16 or 19
21 7 and 20
22 animals/ not humans/
23 21 not 22
24 case reports/
25 23 not 24
26 remove duplicates from 25

Supplementary figures

Risk of bias assessments in RCTs

Figure S1. Risk of bias summaries with judgements about each bias domain for the included CLBP and CNCP RCTs





Forest plots CLBP

Figure S2. Sleep quality - overall: Mean changes from baseline; treatment duration min. 8 to max. 14 weeks; assessed with self-reported CPSI and PSQ

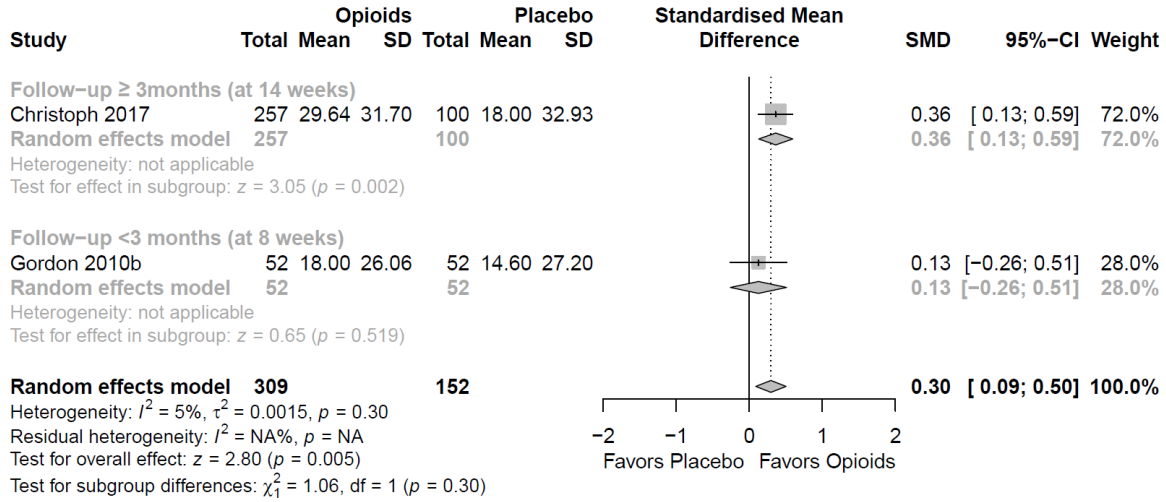


Figure S3. Sleep quality - pain interference with/impact on sleep: Mean changes from baseline; treatment duration min. 5 to max. 8 weeks; assessed with self-reported PSQ and BPI sleep interference subscale

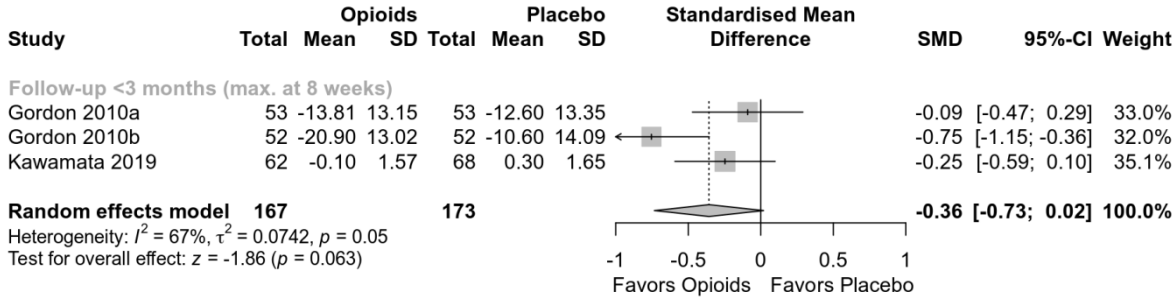


Figure S4. Trial discontinuations (overall): Treatment duration min. 4 to max. 15 weeks

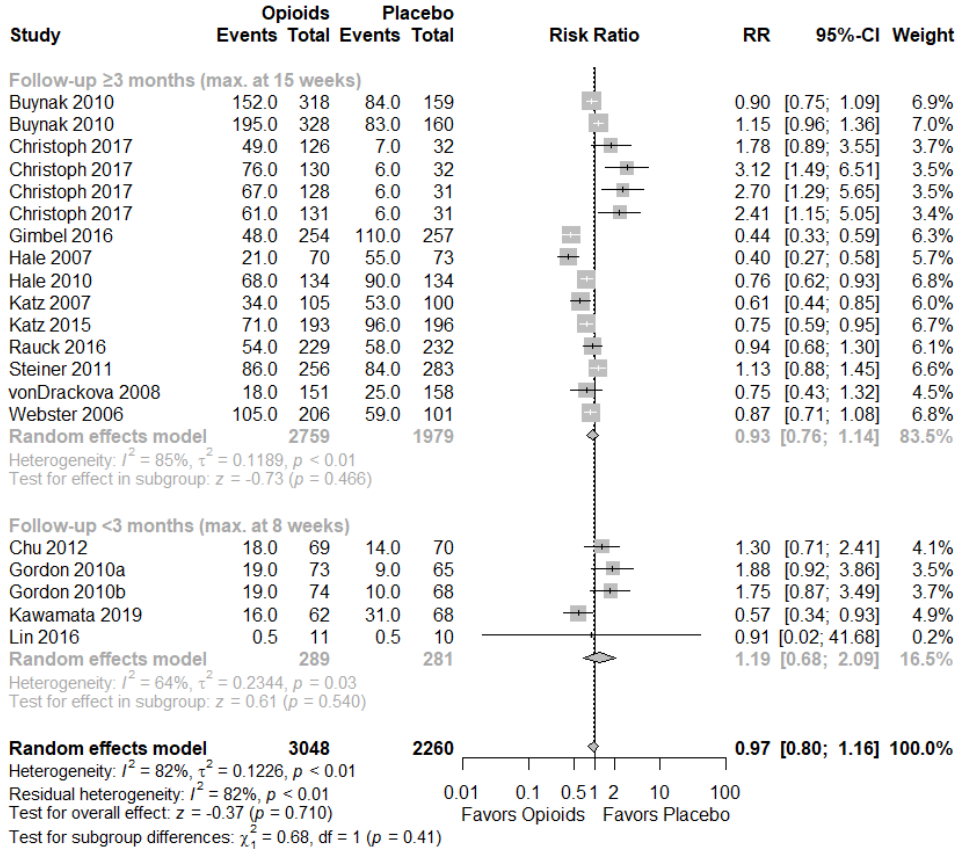


Figure S5. Trial discontinuations due to AEs: Treatment duration min. 4 to max. 15 weeks

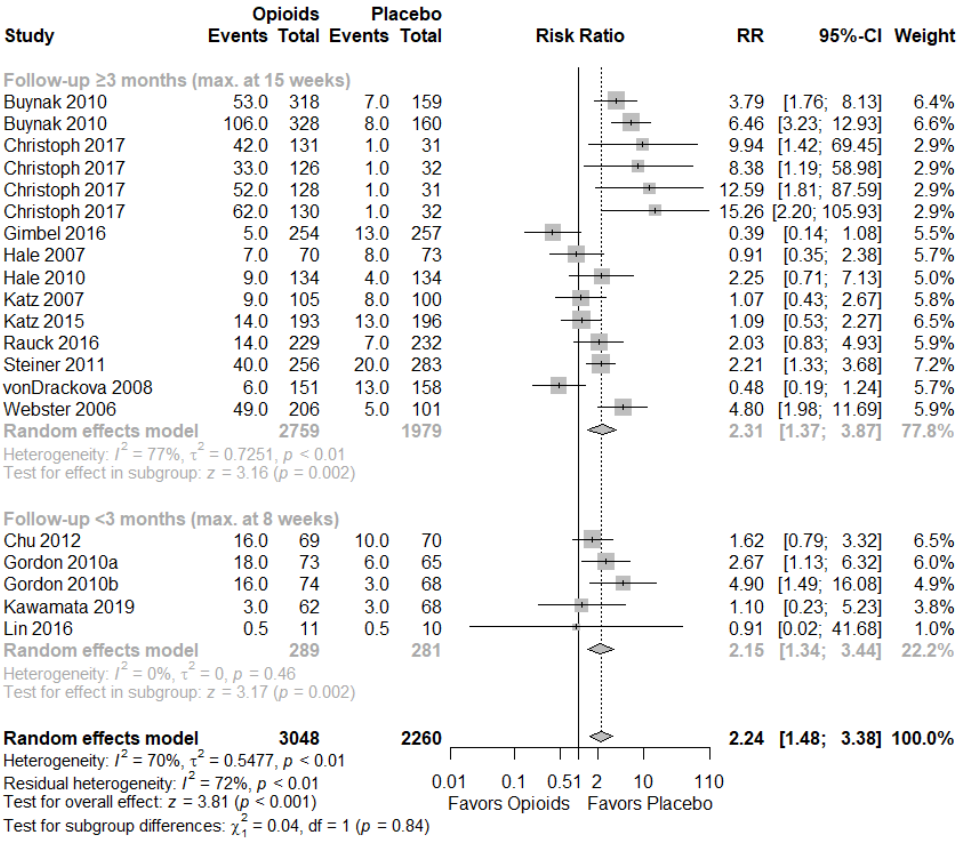


Figure S6. Trial discontinuations due to efficacy lack: Treatment duration min. 4 to max. 15 weeks

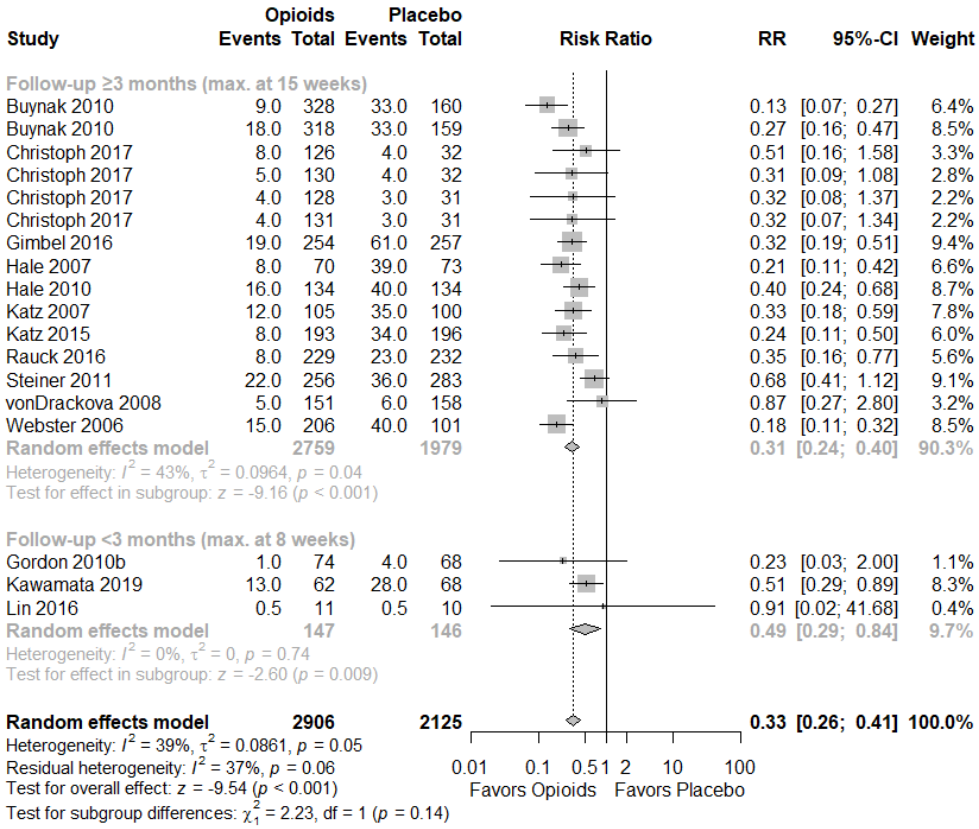


Figure S7. Opioid withdrawal symptoms: Treatment duration min. 4 to max. 15 weeks

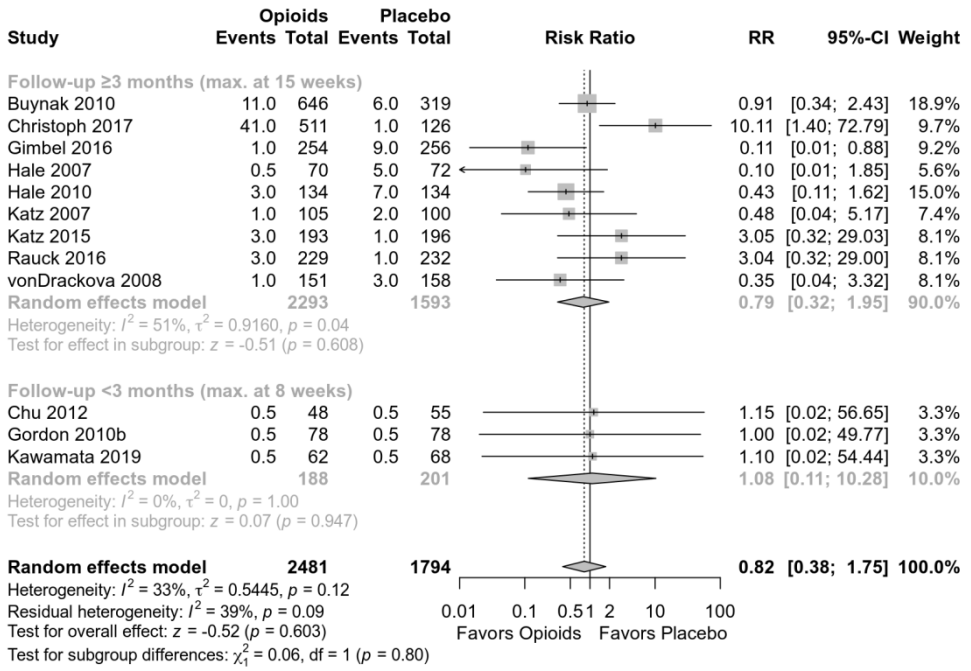


Figure S8. Adverse events (any): Treatment duration min. 5 to max. 15 weeks

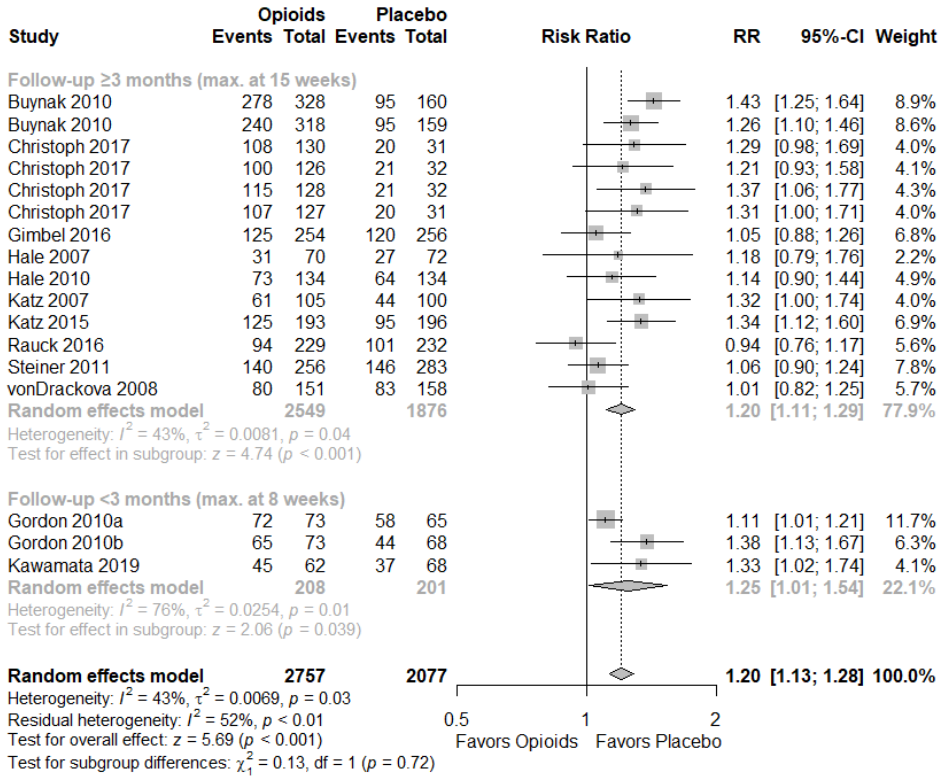


Figure S9. Adverse events (serious): Treatment duration min. 4 to max. 15 weeks

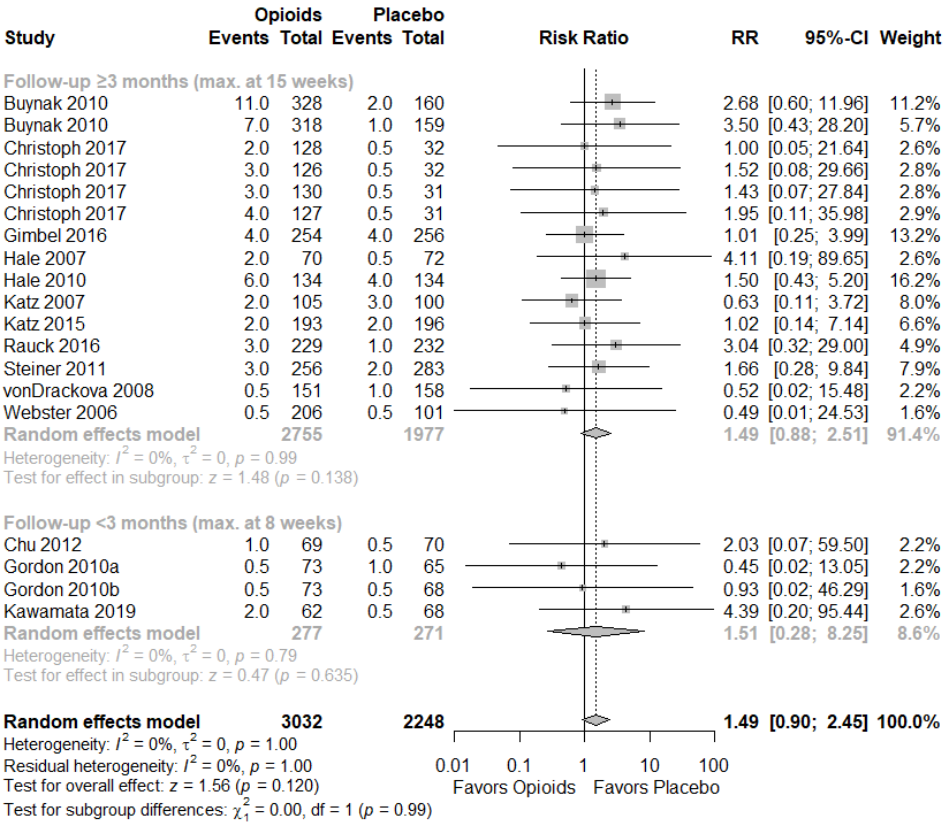


Figure S10. Nausea: Treatment duration min. 4 to max. 15 weeks

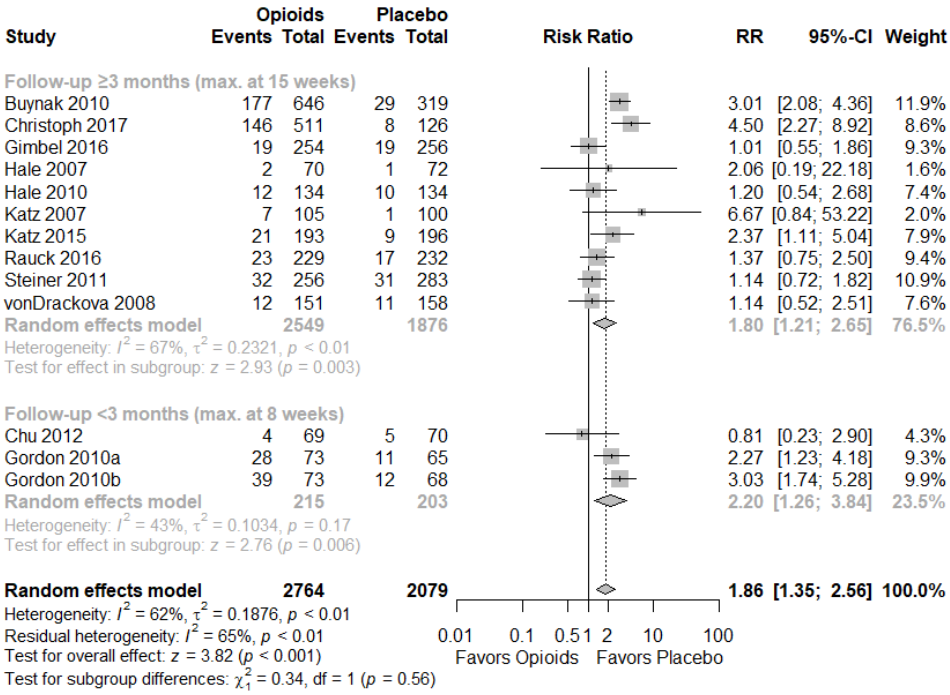


Figure S11. Vomiting: Treatment duration min. 4 to max. 15 weeks

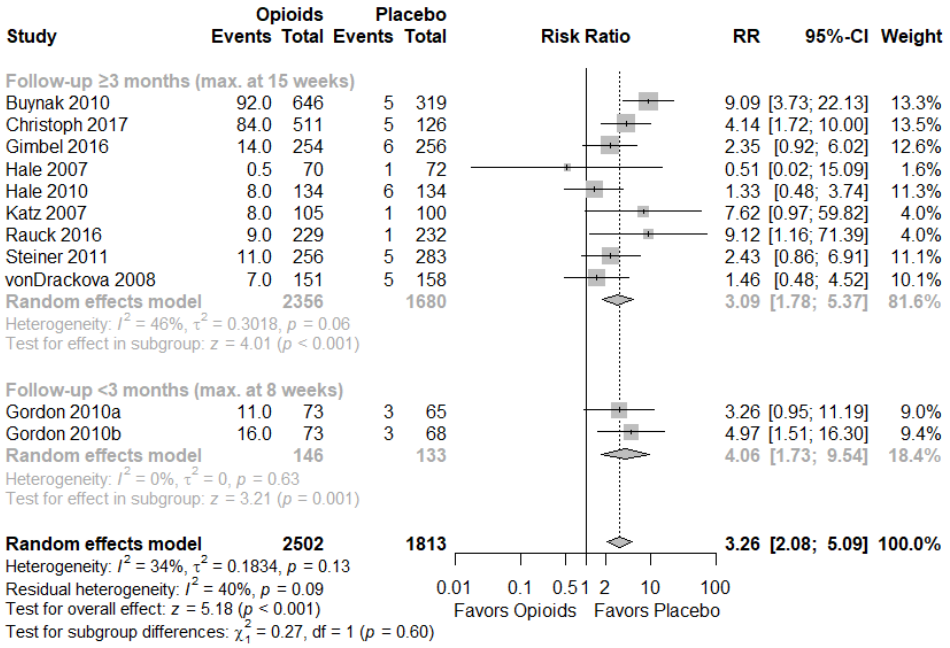


Figure S12. Constipation: Treatment duration min. 4 to max. 15 weeks

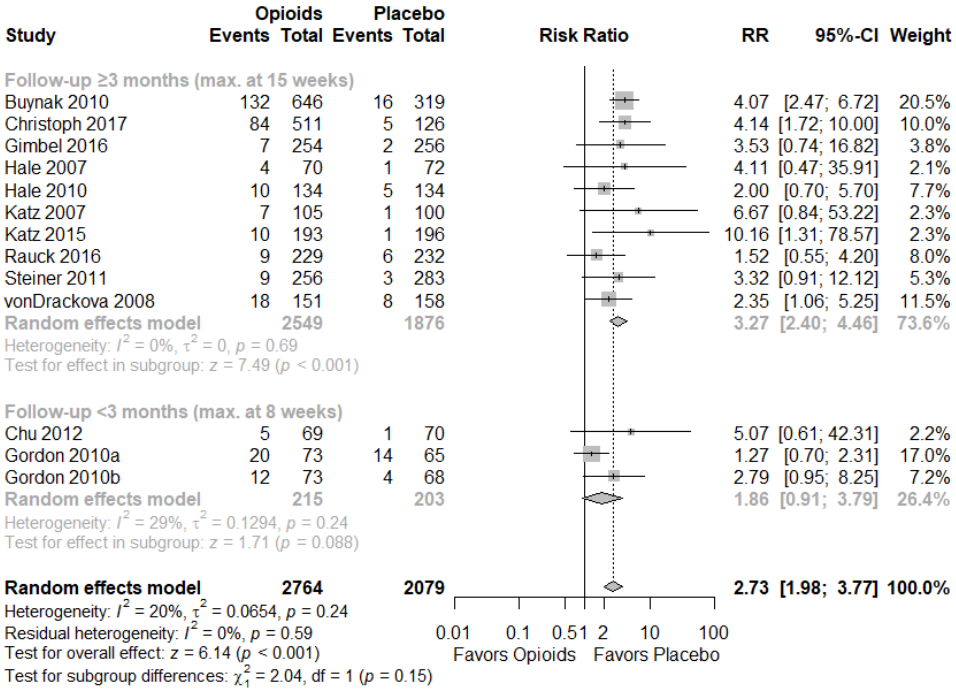


Figure S13. Dizziness: Treatment duration min. 8 to max. 15 weeks

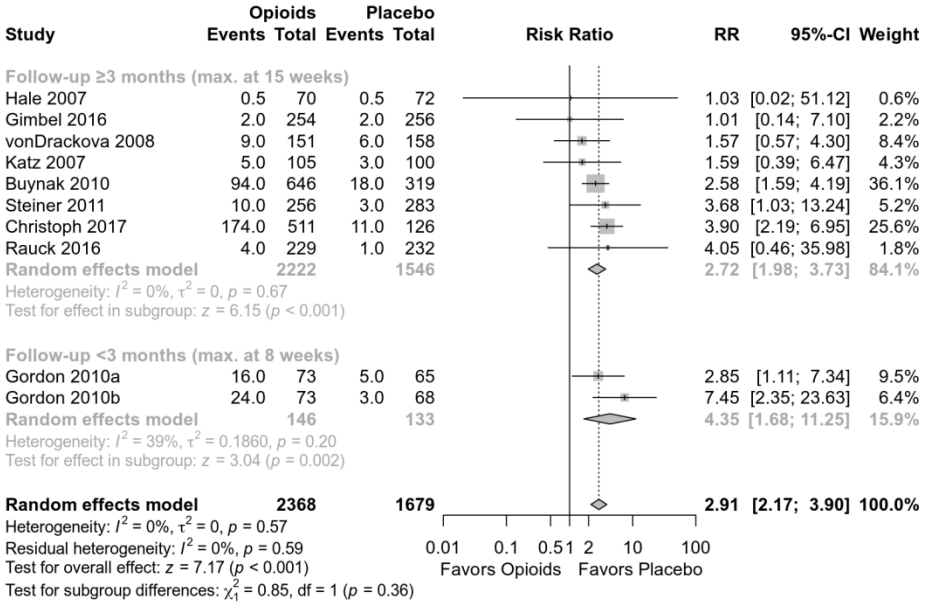


Figure S14. Somnolence: Treatment duration min. 8 to max. 15 weeks

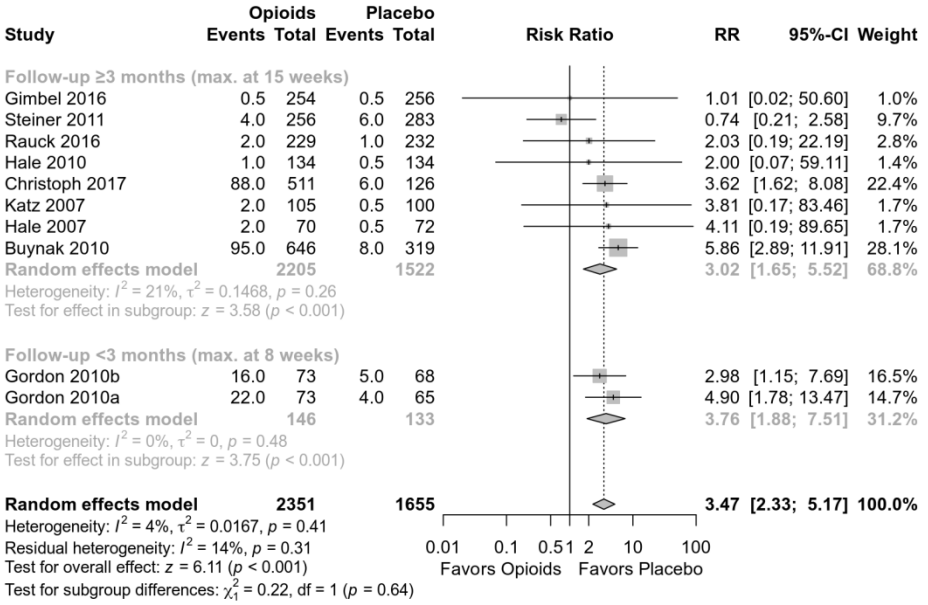


Figure S15. Headache: Treatment duration min. 8 to max. 15 weeks

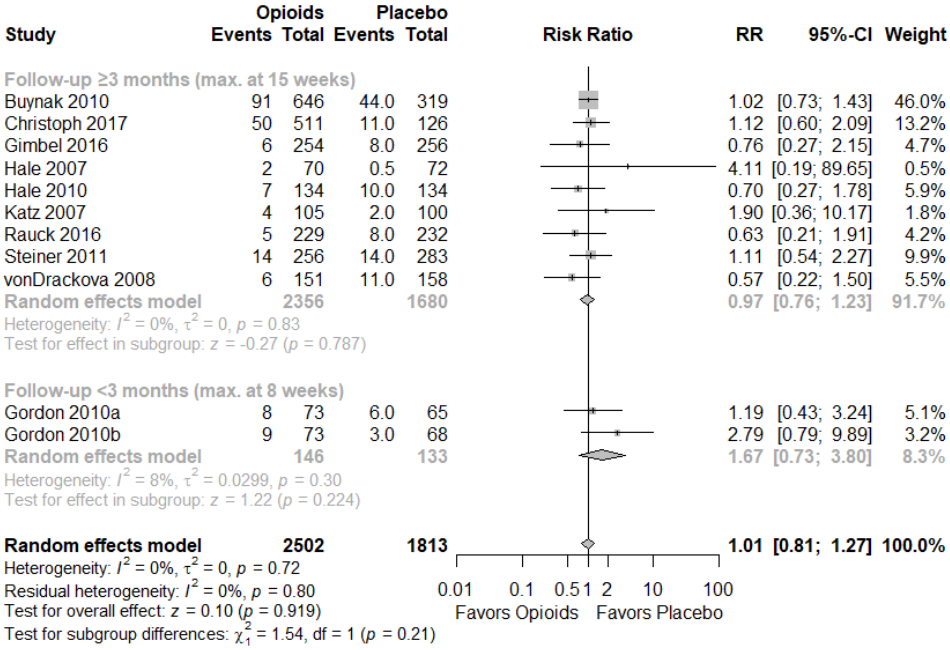


Figure S16. Depression and Anxiety: Mean changes from baseline; treatment duration min. 5 to max. 12 weeks; assessed with self-reported SF-36 MH, SF-12v2 MCS and SF-36v2 MCS

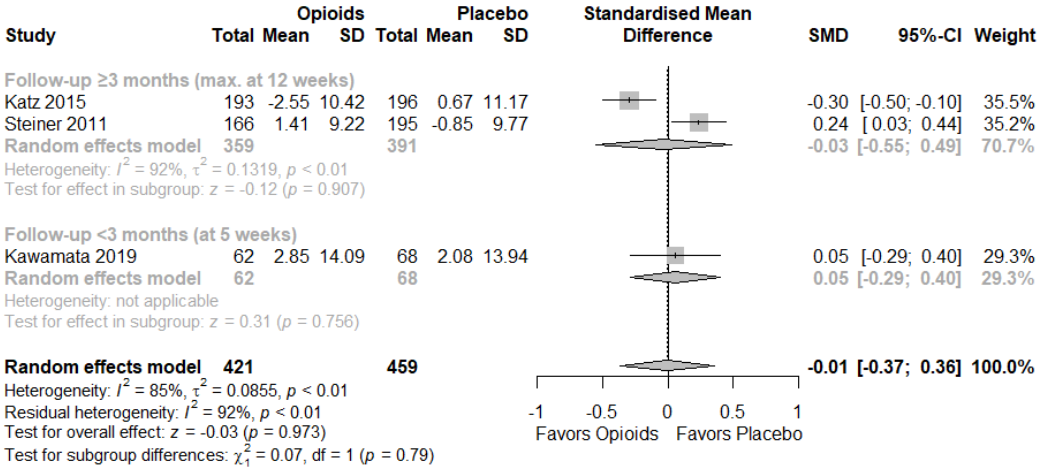


Figure S17. PGIC *much or very much improved*: treatment duration at max. 15 weeks

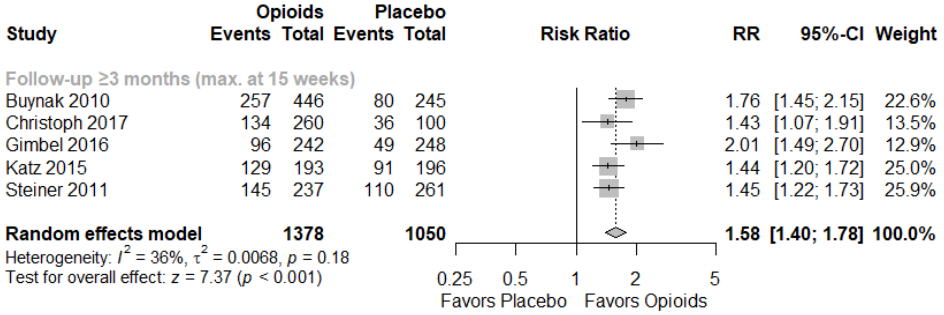


Figure S18. PGR study medication *good/very good/excellent*: treatment duration at 12 weeks

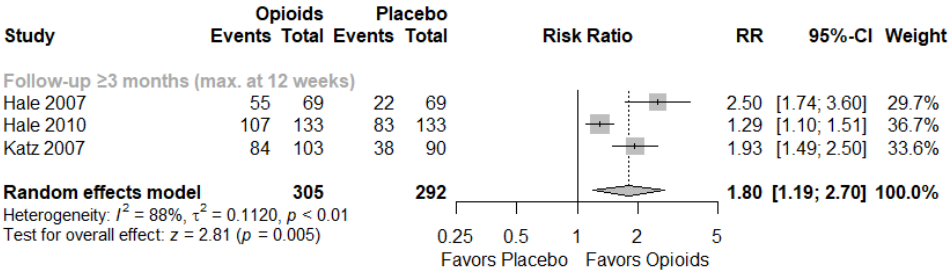
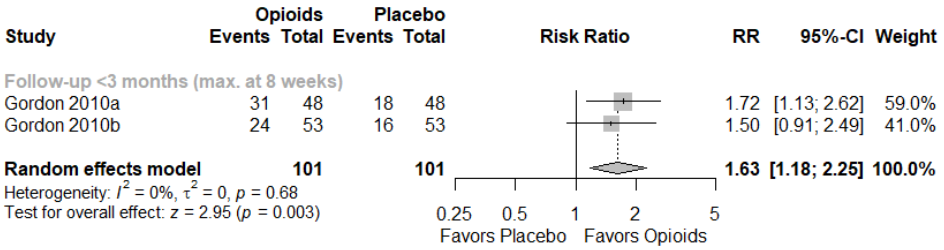


Figure S19. Patient assessed treatment effectiveness *moderately or highly effective*: treatment duration
at 8 weeks



Forest plots CNCP

Figure S20. Global change in pain ≥ moderately better: treatment duration of 3, 6, 9 and 12 months

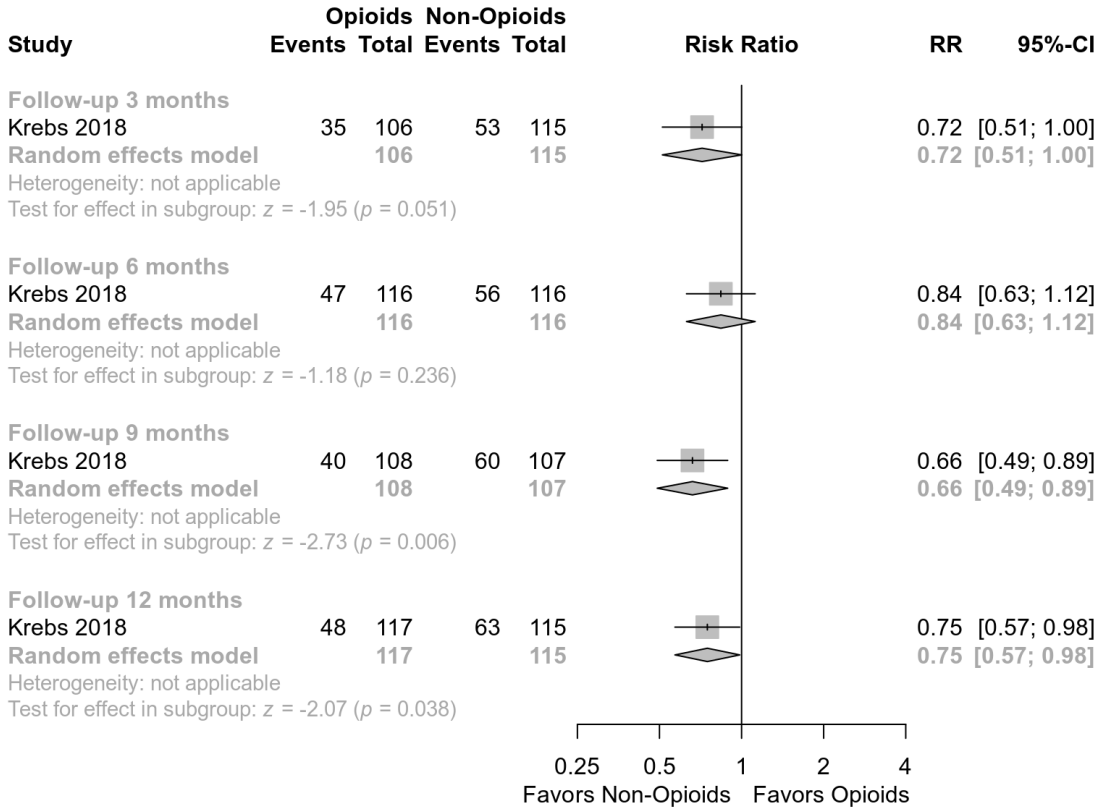


Figure S21. Pain severity and disability: treatment duration ≥6 months; events refer to the number of patients with high disability and moderately or severely limiting pain

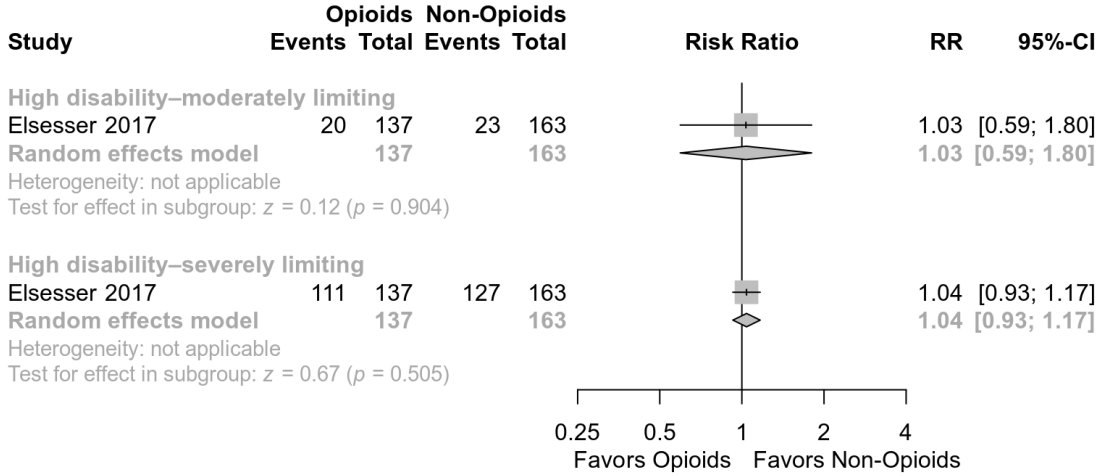


Figure S22. Drug abuse: Events refer to the number of patients with ≥1 positive score(s) or case(s) on the Abuse Index or a clinician-assessed ABC-score of ≥3 during 12-month follow-up

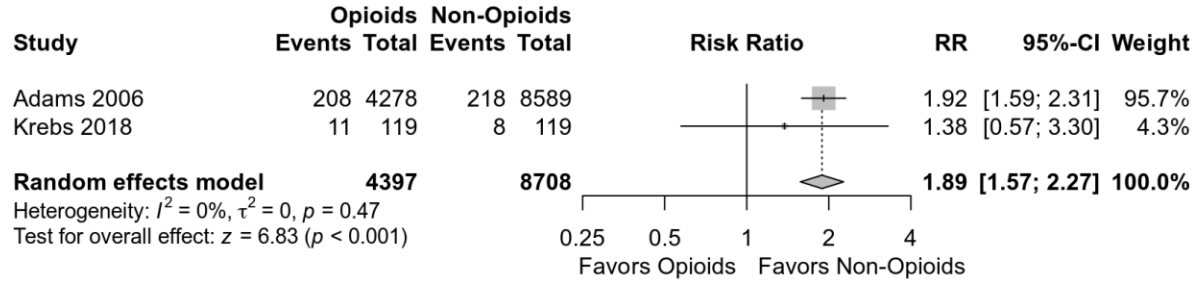


Figure S23. Falls: Events refer to the number of patients with falls in the 12 months after enrolment

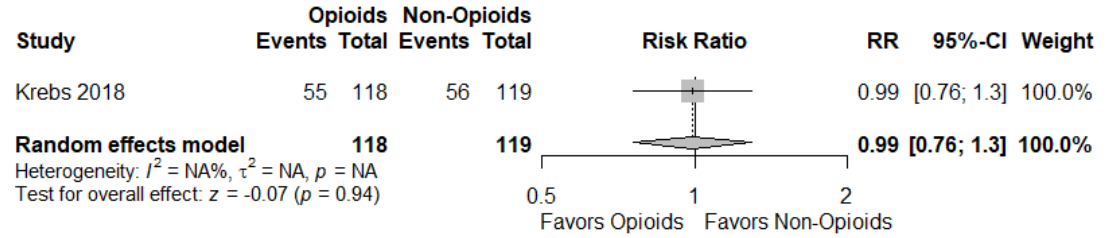


Figure S24. Opioid Abuse or Dependence: Events refer to the number of patients with an opioid abuse or dependence diagnosis

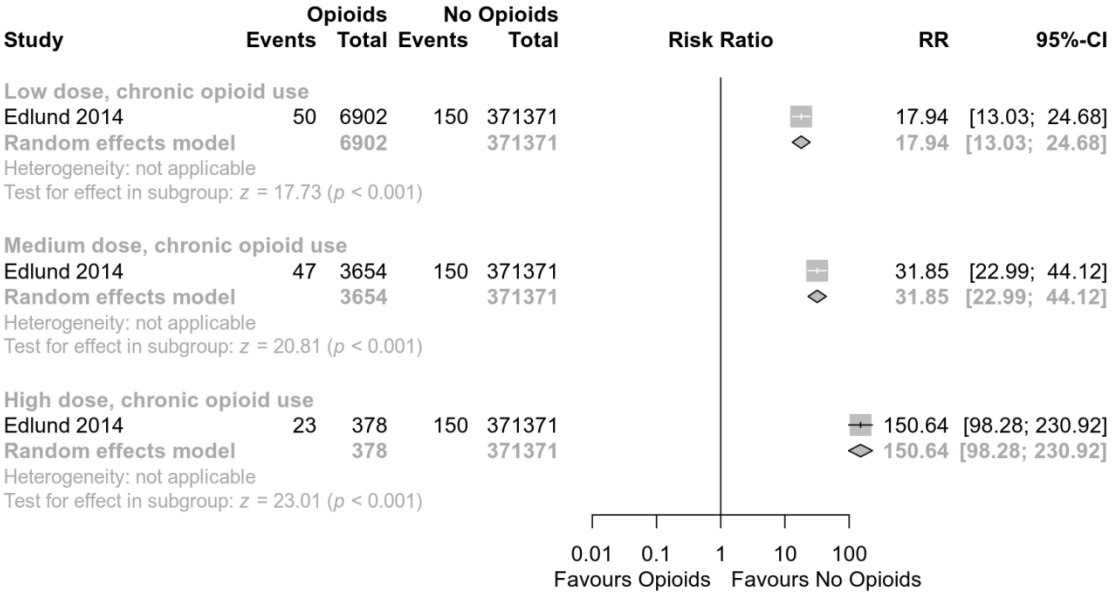


Figure S25. Any adverse events: Events refer to the number of any adverse events that occurred during the study follow-up (treatment duration ≥6 months)

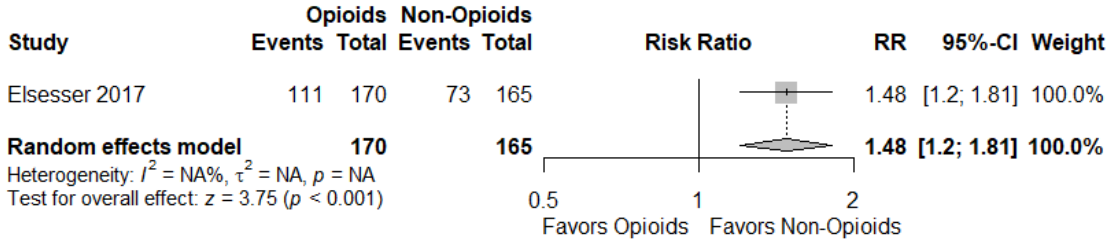
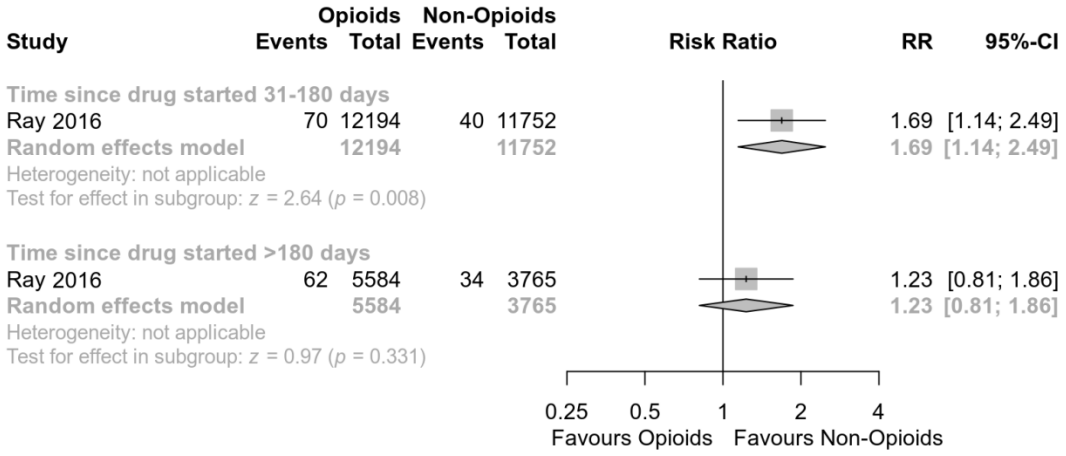


Figure S26. Deaths: Events refer to the number of deaths that occurred during the study follow-up.


Study
Opioids
Events
Total
Non-Opioids
Events
Total
Risk Ratio
RR
95%-CI
Time since drug started 31-180 days
Ray 2016 70 12194 40 11752 1.69 [1.14; 2.49]
Random effects model 12194 11752 1.69 [1.14; 2.49]
Heterogeneity: not applicable
Test for effect in subgroup: z = 2.64 (p = 0.008)
Time since drug started >180 days
Ray 2016 62 5584 34 3765 1.23 [0.81; 1.86]
Random effects model 5584 3765 1.23 [0.81; 1.86]
Heterogeneity: not applicable
Test for effect in subgroup: z = 0.97 (p = 0.331)
0.25
0.5
1
2
4
Favours Opioids
Favours Non-Opioids

Supplementary tables

Risk of bias assessments in NRSI

Table S1. Risk of Bias in non-randomized studies (NRSI)

Study	Bias caused by confounding	Bias in the selection of participants	Bias in the classification of the intervention[#]	Bias due to deviations from the intended interventions[$]	Attrition bias due to missing data	Detection bias in the measurement of outcomes	Reporting bias	Overall judgement
Edlund 2014	Serious approaches (adjusted ORs) to control for predefined prognostic factors were described, but only a few of known confounders* were addressed	Serious 568640 participants retrospectively included, but the selection process was not described	Moderate some aspects of the assignments of intervention status were determined retrospectively	No information	Low data reported for all participants initially included	No information	No information	SERIOUS
Ray 2016	Serious/Moderate approaches (e.g. matching, adjusted HRs/RDs) to control for predefined prognostic factors were described	Serious 45824 of 155191 participants retrospectively included	Moderate some aspects of the assignments of intervention status were determined retrospectively	No information	Low/Moderate data reported for all participants initially included	Moderate	No information	SERIOUS/ MODERATE
Elsesser 2017	Serious approaches (adjusted scores, subgroup analyses) to control for predefined prognostic factors were described, but only a small selection (e.g. age, pain duration, opioid potency) of known confounders were addressed	Serious 333 participants retrospectively included, using a non-consecutively patient enrollment	Moderate some aspects of the assignments of intervention status were determined retrospectively	No information	Low data reported for all participants initially included	Serious interviews conducted by unblinded investigators and pain questionnaires completed by unblinded patients	No information	SERIOUS

* Baseline confounders (i.e., factors that [may] predict whether an individual receives one or the other intervention of interest) identified in a systematic review/study on predicting factors for opioid misuse and abuse in chronic pain patients: age, sex, race, SES/income, pain severity, opioid type (WHO), personal history substance abuse, family history substance abuse, personal history of psychiatric diagnosis, childhood abuse, history of legal problems, DUI/drug conviction, disability level, past motor vehicle accident, current cigarette smoking, positive toxicology screen, lost/stolen prescriptions, unsanctioned dose escalation , unscheduled clinic/ER visits, multiple clinic phone calls, supplemental sources to obtain opioids, and prescription forgery.[12, 13]

[#] Bias in the classification of the intervention: due to the nature of the comparison groups (opioid vs. no opioid/non-opioid treatment) misclassification is unlikely.

[$] Bias due to deviations from the intended interventions: retrospective study design: there is no/insufficient information on the actual intake of additional medications (e.g., pain relievers) or on the use of co-interventions and whether these co-interventions were balanced across the groups.

OR: Odds Ratio; **HR:** hazard ratio; **RD**: Risk Difference

Subgroup analyses

Study design

Table S2: Subgroup analysis for efficacy endpoints comparing EERW vs. Parallel vs. Cross-over trials

	RR (95% CI)	p for interaction
30% pain reduction		
All comparisons (n=9)	1.40 (1.26, 1.56)	
EERW (n=7)	1.44 (1.23, 1.69)	0.49
Parallel (n=2)	1.33 (1.14, 1.55)	
50% pain reduction		
All comparisons (n=8)	1.49 (1.30, 1.70)	
EERW (n=6)	1.57 (1.27, 1.93)	0.38
Parallel (n=2)	1.37 (1.11, 1.69)	
	SMD (95% CI)	**p for interaction**
Pain intensity		
All comparisons (n=15)	-0.40 (-0.46, -0.34)	
EERW (n=8)	-0.44 (-0.53, -0.34)	
Parallel (n=5)	-0.34 (-0.44, -0.25)	0.32
Cross-over (n=2)	-0.30 (-0.57, -0.03)	
Disability		
All comparisons (n=9)	-0.21 (-0.30, -0.12)	
EERW (n=6)	-0.21 (-0.32, -0.11)	
Parallel (n=2)	-0.27 (-0.47, -0.07)	0.35
Cross-over (n=1)	0.04 (-0.34, 0.42)	
Sleep quality (pain interference/impact)		
All comparisons (n=3)	-0.36 (-0.73, 0.02)	
EERW (n=1)	-0.25 (-0.59, 0.10)	0.65
Cross-over (n=2)	-0.42 (-1.07, 0.23)	

Table S3: Subgroup analysis for safety endpoints comparing EERW vs. Parallel vs. Cross-over trials

	RR (95% CI)	p for interaction
Opioid withdrawal symptoms		
All comparisons (n=12)	0.82 (0.38, 1.75)	
EERW (n=7)	0.56 (0.21, 1.52)	
Parallel (n=4)	1.41 (0.30, 6.68)	0.62
Cross-over (n=1)	1.00 (0.02, 49.77)	
Adverse events (any)		
All comparisons (n=13)	1.20 (1.13, 1.28)	
EERW (n=8)	1.15 (1.04, 1.26)	
Parallel (n=3)	1.27 (1.17, 1.39)	0.27
Cross-over (n=2)	1.22 (0.94, 1.59)	
Serious adverse events		
All comparisons (n=15)	1.49 (0.90, 2.45)	
EERW (n=8)	1.38 (0.73, 2.61)	
Parallel (n=5)	1.87 (0.79, 4.40)	0.67
Cross-over (n=2)	0.61 (0.05, 7.86)	
Nausea		
All comparisons (n=13)	1.86 (1.35, 2.56)	
EERW (n=7)	1.32 (1.01, 1.74)	
Parallel (n=4)	2.15 (1.11, 4.19)	**0.02**
Cross-over (n=2)	2.66 (1.76, 4.01)	
Vomiting		
All comparisons (n=11)	3.26 (2.08, 5.09)	
EERW (n=6)	2.33 (1.37, 3.96)	

Parallel (n=3)	3.99 (1.46, 10.92)	0.44
Cross-over (n=2)	4.06 (1.73, 9.54)	
Constipation		
All comparisons (n=13)	2.73 (1.98, 3.77)	
EERW (n=7)	2.71 (1.60, 4.61)	0.16
Parallel (n=4)	3.65 (2.50, 5.31)	
Cross-over (n=2)	1.65 (0.79, 3.44)	
Dizziness		
All comparisons (n=10)	2.91 (2.17, 3.90)	
EERW (n=5)	2.23 (1.03, 4.85)	0.56
Parallel (n=3)	2.79 (1.83, 4.26)	
Cross-over (n=2)	4.35 (1.68, 11.25)	
Somnolence		
All comparisons (n=10)	3.47 (2.33, 5.17)	
EERW (n=6)	1.27 (0.50, 3.19)	0.05
Parallel (n=2)	4.75 (2.79, 8.08)	
Cross-over (n=2)	3.76 (1.88, 7.51)	
Headache		
All comparisons (n=11)	1.01 (0.81, 1.27)	
EERW (n=6)	0.91 (0.59, 1.41)	0.44
Parallel (n=3)	0.99 (0.75, 1.31)	
Cross-over (n=2)	1.67 (0.73, 3.80)	

Table S4: Subgroup analysis for trial discontinuations comparing EERW vs. Parallel vs. Cross-over trials

	RR (95% CI)	**p for interaction**
Discontinuations (overall)		
All comparisons (n=16)	0.97 (0.80, 1.16)	
EERW (n=8)	0.67 (0.53, 0.86)	**<0.0001**
Parallel (n=6)	1.27 (0.99, 1.63)	
Cross-over (n=2)	1.81 (1.10, 2.98)	
Discontinuations due to AEs		
All comparisons (n=16)	2.24 (1.48, 3.38)	
EERW (n=8)	1.28 (0.85,1.94)	**0.0088**
Parallel (n=6)	3.82 (1.87, 7.80)	
Cross-over (n=2)	3.29 (1.64, 6.61)	
Discontinuations due to efficacy lack		
All comparisons (n=14)	0.33 (0.26, 0.41)	
EERW (n=8)	0.37 (0.28, 0.48)	0.21
Parallel (n=5)	0.90 (0.34, 2.39)	
Cross-over (n=1)	0.23 (0.03, 2.00)	

Table S5: Subgroup analysis for patient ratings comparing EERW vs. Parallel vs. Cross-over trials

	RR (95% CI)	**p for interaction**
PGIC (much or very much improved)		
All comparisons (n=5)	1.58 (1.40, 1.78)	
EERW (n=3)	1.56 (1.31, 1.86)	0.73
Parallel (n=2)	1.63 (1.34, 1.99)	

Study/treatment duration

Table S6: Subgroup analysis for efficacy endpoints comparing ≥3 months vs. <3 months trials

	RR (95% CI)	p for interaction
30% pain reduction		
All comparisons (n=9)	1.40 (1.26, 1.56)	
≥3 months (n=8)	1.41 (1.25, 1.58)	0.76
<3 months (n=1)	1.35 (1.06, 1.72)	
	SMD (95% CI)	**p for interaction**
Pain intensity		
All comparisons (n=15)	-0.40 (-0.46, -0.34)	
≥3 months (n=10)	-0.41 (-0.48, -0.34)	0.33
<3 months (n=5)	-0.34 (-0.50, -0.13)	
Disability		
All comparisons (n=9)	-0.21 (-0.30, -0.12)	
≥3 months (n=6)	-0.21 (-0.31, 0.11)	0.91
<3 months (n=3)	-0.22 (-0.48, 0.04)	
Sleep quality (overall)		
All comparisons (n=2)	0.30 (0.09, 0.5)	
≥3 months (n=1)	0.36 (0.13, 0.59)	0.30
<3 months (n=1)	0.13 (-0.26, 0.51)	
Depression/Anxiety		
All comparisons (n=3)	-0.01 (-0.37, 0.36)	
≥3 months (n=2)	-0.03 (-0.55, 0.49)	0.79
<3 months (n=1)	0.05 (-0.29, 0.40)	

Table S7: Subgroup analysis for safety endpoints comparing ≥3 months vs. <3 months trials

	RR (95% CI)	p for interaction
Opioid withdrawal		
All comparisons (n=12)	0.82 (0.38, 1.75)	
≥3 months (n=9)	0.79 (0.32, 1.95)	0.80
<3 months (n=3)	1.08 (0.11, 10.28)	
Adverse events (any)		
All comparisons (n=13)	1.20 (1.13, 1.28)	
≥3 months (n=10)	1.20 (1.11, 1.29)	0.72
<3 months (n=3)	1.25 (1.01, 1.54)	
Serious adverse events		
All comparisons (n=15)	1.49 (0.90, 2.45)	
≥3 months (n=11)	1.49 (0.88, 2.51)	0.99
<3 months (n=4)	1.51 (0.28, 8.25)	
Nausea		
All comparisons (n=13)	1.86 (1.35, 2.56)	
≥3 months (n=10)	1.80 (1.21, 2.65)	0.56
<3 months (n=3)	2.20 (1.26, 3.84)	
Vomiting		
All comparisons (n=11)	3.26 (2.08, 5.09)	
≥3 months (n=9)	3.09 (1.78, 5.37)	0.60
<3 months (n=2)	4.06 (1.73, 9.54)	
Constipation		
All comparisons (n=13)	2.73 (1.98, 3.77)	
≥3 months (n=10)	3.27 (2.40, 4.46)	0.15
<3 months (n=3)	1.86 (0.91, 3.79)	
Dizziness		
All comparisons (n=10)	2.91 (2.17, 3.90)	
≥3 months (n=8)	2.72 (1.98, 3.73)	0.36
<3 months (n=2)	4.35 (1.68, 11.25)	
Somnolence		
All comparisons (n=10)	3.47 (2.33, 5.17)	
≥3 months (n=8)	3.02 (1.65, 5.52)	0.64

<3 months (n=2)	3.76 (1.88, 7.51)	
Headache		
All comparisons (n=11)	1.01 (0.81, 1.27)	
≥3 months (n=9)	0.97 (0.76, 1.23)	0.21
<3 months (n=2)	1.67 (0.73, 3.80)	

Table S8: Subgroup analysis for trial discontinuations comparing ≥3 months vs. <3 months trials

	RR (95% CI)	**p for interaction**
Discontinuations (overall)		
All comparisons (n=16)	0.97 (0.80, 1.16)	
≥3 months (n=11)	0.93 (0.76, 1.14)	0.41
<3 months (n=5)	1.19 (0.68, 2.09)	
Discontinuations due to AEs		
All comparisons (n=16)	2.24 (1.48, 3.38)	
≥3 months (n=11)	2.31 (1.37, 3.87)	0.84
<3 months (n=5)	2.15 (1.34, 3.44)	
Discontinuations due to efficacy lack		
All comparisons (n=14)	0.33 (0.26, 0.41)	
≥3 months (n=11)	0.54 (0.33, 0.86)	0.81
<3 months (n=3)	0.49 (0.29, 0.84)	

Opioid experience status at trial start

Table S9: Subgroup analysis for efficacy endpoints comparing Opioid-naïve vs. Opioid-experienced vs. Opioid-naïve and-experienced patients

	RR (95% CI)	**p for interaction**
30% pain reduction		
All comparisons (n=9)	1.40 (1.26, 1.56)	
Opioid-naïve (n=3)	1.25 (1.13, 1.39)	
Opioid-experienced (n=2)	1.99 (1.66, 2.39)	**<0.0001**
Opioid-naïve and-experienced (n=4)	1.37 (1.22, 1.54)	
50% pain reduction		
All comparisons (n=8)	1.49 (1.30, 1.70)	
Opioid-naïve (n=3)	1.31 (1.13, 1.51)	
Opioid-experienced (n=2)	2.27 (1.74, 2.97)	**0.0017**
Opioid-naïve and-experienced (n=3)	1.43 (1.20, 1.70)	
	SMD (95% CI)	**p for interaction**
Pain intensity		
All comparisons (n=15)	-0.40 (-0.46, -0.34)	
Opioid-naïve (n=4)	-0.42 (-0.54, -0.30)	
Opioid-experienced (n=4)	-0.48 (-0.68, -0.27)	0.38
Opioid-naïve and-experienced (n=7)	-0.35 (-0.43, -0.27)	
Disability		
All comparisons (n=)	-0.21 (-0.30, -0.12)	
Opioid-naïve (n=)	-0.26 (-0.55, 0.04)	
Opioid-experienced (n=)	-0.23 (-0.37, -0.08)	0.77
Opioid-naïve and-experienced (n=2)	-0.17 (-0.30, -0.03)	
Sleep quality (overall)		
All comparisons (n=2)	0.30 (0.09, 0.5)	
Opioid-experienced (n=1)	0.13 (-0.26, 0.51)	0.30
Opioid-naïve and-experienced (n=1)	0.36 (0.13, 0.59)	
Sleep quality (pain interference/impact)		
All comparisons (n=3)	-0.36 (-0.73, 0.02)	
Opioid-experienced (n=1)	-0.75 (-1.15, -0.36)	**0.02**

Opioid-naïve and-experienced (n=2)	-0.18 (-0.43, 0.08)	
Depression/Anxiety		
All comparisons (n=3)	-0.01 (-0.37, 0.36)	
Opioid-naïve (n=1)	0.24 (0.03, 0.44)	0.06
Opioid-naïve and-experienced (n=2)	-0.15 (-0.49, 0.19)	

Table S10: Subgroup analysis for safety endpoints comparing Opioid-naïve vs. Opioid-experienced vs. Opioid-naïve and-experienced patients

	RR (95% CI)	**p for interaction**
Opioid withdrawal		
All comparisons (n=12)	0.82 (0.38, 1.75)	
Opioid-naïve (n=2)	1.25 (0.20, 7.71)	
Opioid-experienced (n=5)	0.29 (0.11, 0.72)	**0.029**
Opioid-naïve and-experienced (n=5)	2.03 (0.62, 6.67)	
Adverse events (any)		
All comparisons (n=13)	1.20 (1.13, 1.28)	
Opioid-naïve (n=3)	1.08 (0.91, 1.27)	
Opioid-experienced (n=5)	1.14 (1.01, 1.29)	0.12
Opioid-naïve and-experienced (n=5)	1-28 (1.17, 1.39)	
Serious adverse events		
All comparisons (n=15)	1.49 (0.90, 2.45)	
Opioid-naïve (n=3)	1.32 (0.44, 3.96)	
Opioid-experienced (n=5)	1.28 (0.56, 2.95)	0.83
Opioid-naïve and-experienced (n=7)	1.78 (0.83, 3.80)	
Nausea		
All comparisons (n=13)	1.86 (1.35, 2.56)	
Opioid-naïve (n=3)	1.34 (0.83, 2.16)	
Opioid-experienced (n=5)	1.50 (0.89, 2.53)	**0.05**
Opioid-naïve and-experienced (n=5)	2.67 (1.85, 3.86)	
Vomiting		
All comparisons (n=11)	3.26 (2.08, 5.09)	
Opioid-naïve (n=3)	3.69 (1.58, 8.63)	0.06
Opioid-experienced (n=5)	2.05 (1.21, 3.45)	
Opioid-naïve and-experienced (n=3)	5.32 (2.88, 9.84)	
Constipation		
All comparisons (n=13)	2.73 (1.98, 3.77)	
Opioid-naïve (n=3)	2.38 (1.13, 5.03)	
Opioid-experienced (n=5)	2.53 (1.53, 4.18)	0.81
Opioid-naïve and-experienced (n=5)	3.25 (1.58, 6.70)	
Dizziness		
All comparisons (n=10)	2.91 (2.17, 3.90)	
Opioid-naïve (n=3)	2.71 (1.14, 6.46)	
Opioid-experienced (n=3)	2.96 (0.81, 10.88)	0.99
Opioid-naïve and-experienced (n=4)	2.93 (2.09, 4.12)	
Somnolence		
All comparisons (n=10)	3.47 (2.33, 5.17)	
Opioid-naïve (n=3)	1.08 (0.38, 3.07)	
Opioid-experienced (n=4)	2.83 (1.20, 6.65)	**0.03**
Opioid-naïve and-experienced (n=3)	4.78 (2.98, 7.65)	
Headache		
All comparisons (n=11)	1.01 (0.81, 1.27)	
Opioid-naïve (n=3)	1.02 (0.58, 1.79)	
Opioid-experienced (n=5)	0.92 (0.50, 1.69)	0.93
Opioid-naïve and-experienced (n=3)	1.05 (0.79, 1.40)	

Table S11: Subgroup analysis for trial discontinuations comparing Opioid-naïve vs. Opioid-experienced vs. Opioid-naïve and-experienced patients

	RR (95% CI)	**p for interaction**
Discontinuations (overall)		
All comparisons (n=16)	0.97 (0.80, 1.16)	
Opioid-naïve (n=4)	0.88 (0.63, 1.23)	**0.03**
Opioid-experienced (n=5)	0.66 (0.44, 0.98)	
Opioid-naïve and-experienced (n=7)	1.19 (0.94, 1.51)	
Discontinuations due to AEs		
All comparisons (n=16)	2.24 (1.48, 3.38)	
Opioid-naïve (n=4)	1.88 (1.26, 2.79)	0.05
Opioid-experienced (n=5)	1.09 (0.45, 2.64)	
Opioid-naïve and-experienced (n=7)	3.55 (2.07, 6.10)	
Discontinuations due to efficacy lack		
All comparisons (n=14)	0.33 (0.26, 0.41)	
Opioid-naïve (n=4)	0.46 (0.30, 0.70)	0.21
Opioid-experienced (n=5)	0.34 (0.24, 0.48)	
Opioid-naïve and-experienced (n=5)	0.72 (0.31, 1.62)	

Table S12: Subgroup analysis for patient ratings comparing Opioid-naïve vs. Opioid-experienced vs. Opioid-naïve and-experienced patients

	RR (95% CI)	**p for interaction**
PGIC (much or very much improved)		
All comparisons (n=5)	1.58 (1.40, 1.78)	
Opioid-naïve (n=1)	1.45 (1.22, 1.73)	0.18
Opioid-experienced (n=1)	2.01 (1.49, 2.70)	
Opioid-naïve and-experienced (n=3)	1.55 (1.35, 1.79)	
PGA of study medication (good/very good/excellent)		
All comparisons (n=3)	1.80 (1.19, 2.70)	
Opioid-naïve (n=1)	1.93 (1.49, 2.50)	0.80
Opioid-experienced (n=2)	1.76 (0.89, 3.48)	
Patient assessed treatment effectiveness		
All comparisons (n=2)	1.63 (1.18, 2.25)	
Opioid-experienced (n=1)	1.50 (0.91, 2.49)	0.68
Opioid-naïve and-experienced (n=2)	1.72 (1.13, 2.62)	

GRADE Evidence Profiles for CLBP outcomes (RCTs)

Certainty assessment							№ of patients		Effect		Certainty
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Opioids	Placebo	Relative (95% CI)	Absolute (95% CI)	
Pain intensity (follow up: range 4 weeks to 15 weeks; assessed with: self-reported NRS [0–10]; lower is better; the MID = 2-points)											
15	randomised trials	serious [a]	not serious	serious [b]	not serious	none	2703	1916	-	MD **0.9 lower** (1.03 lower to 0.76 lower)	⊕⊕◯◯ LOW
30% Pain reduction at the end of treatment (follow up: range 5 weeks to 15 weeks)											
9	randomised trials	serious [c]	not serious	serious [b]	not serious	none	1081/2080 (52.0%)	607/1606 (37.8%)	**RR 1.40** (1.26 to 1.56)	**151 more per 1.000** (from 98 more to 212 more)	⊕⊕◯◯ LOW
50% Pain reduction at the end of treatment (follow up: range 12 weeks to 15 weeks)											
8	randomised trials	serious [c]	not serious	serious [b]	not serious	none	738/2018 (36.6%)	394/1538 (25.6%)	**RR 1.49** (1.30 to 1.70)	**126 more per 1.000** (from 77 more to 179 more)	⊕⊕◯◯ LOW
Disability (follow up: range 4 weeks to 14 weeks; assessed with: self-reported RMDQ [0-24]; lower is better; the MID = 5-points)											
9	randomised trials	serious [a]	not serious	serious [b]	not serious	none	1354	1235	-	MD **1.09 lower** (1.56 lower to 0.63 lower)	⊕⊕◯◯ LOW
Sleep quality (follow up: range 8 weeks to 14 weeks; assessed with: self-reported VAS [0–100]; higher is better; the MID = 10 mm)											
2	randomised trials	serious [a]	not serious	not serious	not serious	none	309	152	-	MD **8.8 higher** (2.64 higher to 14.67 higher)	⊕⊕⊕◯ MODERATE
Sleep quality: pain interference/impact on sleep (follow up: range 5 weeks to 8 weeks; assessed with: self-reported NRS [0-10]; lower is better; the MID = 1-point)											
3	randomised trials	serious [a]	serious [f]	not serious	serious [e]	none	167	173	-	MD **0.58 lower** (1.18 lower to 0.03 higher)	⊕◯◯◯ VERY LOW
Trial discontinuations (Overall) (follow up: range 4 weeks to 15 weeks)											
16	randomised trials	serious [a]	serious [d]	serious [b]	serious [g]	none	1177/3048 (38.6%)	886/2260 (39.2%)	**RR 0.97** (0.80 to 1.16)	**12 fewer per 1.000** (from 78 fewer to 63 more)	⊕◯◯◯ VERY LOW
Trial discontinuations (adverse events) (follow up: range 4 weeks to 15 weeks)											
16	randomised trials	serious [a]	serious [f]	serious [b]	not serious	none	554/3048 (18.2%)	132/2260 (5.8%)	**RR 2.26** (1.49 to 3.43)	**74 more per 1.000** (from 29 more to 142 more)	⊕◯◯◯ VERY LOW
Trial discontinuations (efficacy lack) (follow up: range 4 weeks to 15 weeks)											
14	randomised trials	serious [a]	not serious	serious [b]	not serious	none	175/2906 (6.0%)	426/2125 (20.0%)	**RR 0.33** (0.26 to 0.41)	**134 fewer per 1.000** (from 148 fewer to 118 fewer)	⊕⊕◯◯ LOW
Opioid withdrawal symptoms (follow up: range 12 weeks to 15 weeks)											
12	randomised trials	serious [a]	not serious	serious [b]	serious [h]	none	64/2481 (2.6%)	35/1794 (2.0%)	**RR 0.82** (0.38 to 1.75)	**4 fewer per 1.000** (from 12 fewer to 15 more)	⊕◯◯◯ VERY LOW
Opioid dependency (follow up: 5 weeks)											
1	randomised trials	serious [a]	not serious	serious [i]	not serious	none	Kawamata et al. reported that *"no patients were judged to have developed drug dependency by the Data and Safety Monitoring Board"* in either the opioid (n = 62) or placebo group (n = 68).				⊕⊕◯◯ LOW
Opioid misuse or abuse (follow up: range 5 weeks to 12 weeks)											

3	randomised trials	serious [a]	not serious	serious [i]	not serious	none	No cases of opioid abuse were reported in the opioid group (total n = 572) or placebo group (total n = 607).				⊕⊕◯◯ LOW
Adverse events (any) (follow up: range 5 weeks to 15 weeks)											
13	randomised trials	serious [a]	not serious	serious [b]	not serious	none	1859/2757 (67.4%)	1091/2077 (52.5%)	**RR 1.20** (1.13 to 1.28)	**105 more per 1.000** (from 68 more to 147 more)	⊕⊕◯◯ LOW
Adverse events (serious) (follow up: range 4 weeks to 15 weeks)											
15	randomised trials	serious [a]	not serious	serious [b]	serious [g]	none	55/3032 (1.8%)	23/2248 (1.0%)	**RR 1.44** (0.88 to 2.37)	**5 more per 1.000** (from 1 fewer to 14 more)	⊕◯◯◯ VERY LOW
Deaths (follow up: range 5 weeks to 15 weeks)											
10	randomised trials	serious [a]	not serious	serious [b]	not serious [j]	none	None of the 10 trials addressing mortality reported any treatment-related deaths in either intervention arm.				⊕⊕◯◯ LOW
Nausea (follow up: range 4 weeks to 15 weeks)											
13	randomised trials	serious [a]	serious [f]	serious [b]	not serious	none	522/2764 (18.9%)	164/2079 (7.9%)	**RR 1.86** (1.35 to 2.56)	**68 more per 1.000** (from 28 more to 123 more)	⊕◯◯◯ VERY LOW
Vomitting (follow up: range 4 weeks to 15 weeks)											
11	randomised trials	serious [a]	not serious	serious [b]	not serious	none	260/2502 (10.4%)	41/1813 (2.3%)	**RR 3.22** (2.04 to 5.09)	**50 more per 1.000** (from 24 more to 92 more)	⊕⊕◯◯ LOW
Constipation (follow up: range 4 weeks to 15 weeks)											
13	randomised trials	serious [a]	not serious	serious [b]	not serious	none	327/2764 (11.8%)	67/2079 (3.2%)	**RR 2.73** (1.98 to 3.77)	**56 more per 1.000** (from 32 more to 89 more)	⊕⊕◯◯ LOW
Dizziness (follow up: range 8 weeks to 15 weeks)											
10	randomised trials	serious [a]	not serious	serious [b]	not serious	none	338/2368 (14.3%)	52/1679 (3.1%)	**RR 2.91** (2.17 to 3.90)	**59 more per 1.000** (from 36 more to 90 more)	⊕⊕◯◯ LOW
Somnolence (follow up: range 8 weeks to 15 weeks)											
10	randomised trials	serious [a]	not serious	serious [b]	not serious	none	232/2351 (9.9%)	30/1655 (1.8%)	**RR 3.47** (2.33 to 5.17)	**45 more per 1.000** (from 24 more to 76 more)	⊕⊕◯◯ LOW
Headache (follow up: range 8 weeks to 15 weeks)											
11	randomised trials	serious [a]	not serious	serious [b]	serious [g]	none	202/2502 (8.1%)	117/1813 (6.5%)	**RR 1.01** (0.81 to 1.27)	**1 more per 1.000** (from 12 fewer to 17 more)	⊕◯◯◯ VERY LOW
Depression and Anxiety (follow up: range 5 weeks to 12 weeks; assessed with: self-reported SF-36v2 MCS [0-100]; higher is better; surrogate outcome [no MID])											
3	randomised trials	serious [a]	serious [d]	serious [k]	serious [e]	none	421	459	-	MD **0.1 lower** (3.52 lower to 3.43 higher)	⊕◯◯◯ VERY LOW
Suicidal ideation or behavior (follow up: range 12 weeks to 14 weeks)											
2	randomised trials	serious [a]	not serious	serious [b]	serious [l]	none	Christoph et al. reported that no events occurred in either the placebo (n = 126) or opioid group (n = 511). Steiner et al. reported that only 1 event of suicidal ideation occurred in the placebo group (n = 283) compared to none in the opioids group (n = 256). Hale et al. reported one in the intervention group (n = 134) and none in the placebo group (n = 134).				⊕◯◯◯ VERY LOW
PGIC: much improved or very much improved (follow up: 15 weeks)											
5	randomised trials	serious [a]	not serious	serious [b]	not serious	none	761/1378 (55.2%)	366/1050 (34.9%)	**RR 1.58** (1.40 to 1.78)	**202 more per 1.000** (from 139 more to 272 more)	⊕⊕◯◯ LOW
PGR study medication: good, very good, or excellent (follow up: 12 weeks)											
3	randomised trials	serious [c]	serious [d]	serious [b]	not serious	none	246/305 (80.7%)	143/292 (49.0%)	**RR 1.80** (1.19 to 2.70)	**392 more per 1.000** (from 93 more to 833 more)	⊕◯◯◯ VERY LOW

Patient assessed treatment effectiveness: moderately or highly effective (follow up: 8 weeks)											
2	randomised trials	serious [a]	not serious	not serious	serious [m]	none	55/101 (54.5%)	34/101 (33.7%)	**RR 1.63** (1.18 to 2.25)	**212 more per 1.000** (from 61 more to 421 more)	⊕⊕◯◯ LOW

CI: Confidence interval; **MD:** Mean difference; **RR:** Risk ratio

Explanations
a. Risk of bias downgraded by one level: attrition bias (missing outcome data) and selective reporting cannot be excluded.
b. Indirectness downgraded by one level: the study population in at least half of the included trials consisted of opioid responders only (EERW design)
c. Risk of bias downgraded by one level: attrition bias (missing outcome data)
d. Inconsistency downgraded one level: I^2 >75% (considerable heterogeneity)
e. Imprecision downgraded by one level: 95%-CI included zero, i.e. 95%-CI consistent with the possibility of improving and the possibility of worsening sleep quality/symptoms.
f. Inconsistency downgraded by one level: I^2 >50% (substantial heterogeneity)
g. Imprecision downgraded by one level: 95%-CI included zero, i.e. 95%-CI consistent with the possibility of less discontinuations/cases and the possibility of more discontinuations/cases
h. Imprecision downgraded by one level: 95%-CI included 1, i.e. CI consistent with the possibility of harm (more opioid withdrawal) and the possibility of benefit (less opioid withdrawal)
i. Indirectness downgraded by one level: the study population consisted only of opioid responders as the trial/trials had an EERW design.
j. Difficult to assess imprecision as no events occurred in either intervention arm in all of the included studies. However, the difference in effect estimate is so small that it is sufficiently precise (less than 1 per 1000 fewer).
k. Indirectness downgraded by one level: out of the 3 trials, the study population in 2 trials with an EERW design only consisted of opioid responders & surrogate outcome for depression and anxiety.
l. Imprecision downgraded by one level: low number of events (i.e. only 1 event in the placebo group).
m. Imprecision downgraded by one level: low number of participants

GRADE Evidence Profiles for CNCP outcomes (RCTs and NRSIs)

Certainty assessment						№ of patients		Effect	Certainty

№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Opioids	Non-Opioids	Relative (95% CI)	Absolute (95% CI)	
Pain intensity (follow-up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	serious [b]	none	117	117	-	MD **0.5 higher** (0.05 higher to 0.95 higher)	⊕⊕◯◯ LOW
Disability/Pain-related function (follow-up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	very serious [c]	none	117	117	-	MD **0.2 higher** (0.41 lower to 0.81 higher)	⊕◯◯◯ VERY LOW
30% reduction in BPI pain severity score (follow up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	serious [b]	none	48/117 (41.0%)	63/117 (53.8%)	**RR 0.76** (0.58 to 1.00)	**129 fewer per 1.000** (from 226 fewer to 0 fewer)	⊕⊕◯◯ LOW
30% reduction in BPI interference score (follow up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	very serious [c]	none	69/117 (59.0%)	71/117 (60.7%)	**RR 0.97** (0.79 to 1.20)	**18 fewer per 1.000** (from 127 fewer to 121 more)	⊕◯◯◯ VERY LOW
Patient-reported global change in pain ≥ moderately better (follow up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	serious [b]	none	48/117 (41.0%)	63/115 (54.8%)	**RR 0.75** (0.57 to 0.98)	**137 fewer per 1.000** (from 236 fewer to 11 fewer)	⊕⊕◯◯ LOW
Drug abuse (follow up: 12 months)											
2	randomised trials	serious [d]	not serious	serious [e]	not serious	none	219/4397 (5.0%)	226/8708 (2.6%)	**RR 1.89** (1.57 to 2.27)	**23 more per 1.000** (from 15 more to 33 more)	⊕⊕◯◯ LOW
Falls (follow up: 12 months)											
1	randomised trials	serious [a]	not serious	not serious	very serious [c]	none	55/119 (46.2%)	56/119 (47.1%)	**RR 0.99** (0.76 to 1.30)	**5 fewer per 1.000** (from 113 fewer to 141 more)	⊕◯◯◯ VERY LOW
Pain Severity and Disability (therapy duration ≥6 months)											
1	observational studies	very serious [f]	not serious	not serious	very serious [c]	none	111/137 (81.0%)	127/163 (77.9%)	**RR 1.04** (0.93 to 1.17)	**31 more per 1.000** (from 55 fewer to 132 more)	⊕◯◯◯ VERY LOW
Opioid Abuse or Dependence (follow up: 18 months)											
1	observational studies	very serious [f]	not serious	not serious	serious [g]	none	47/3654 (1.3%)	150/371371 (0.0%)	**RR 31.85** (22.99 to 44.12)	**12 more per 1.000** (from 9 more to 17 more)	⊕◯◯◯ VERY LOW
Any adverse events (therapy duration ≥6 months)											
1	observational studies	very serious [f]	not serious	not serious	not serious [h]	none	111/170 (65.3%)	73/165 (44.2%)	**RR 1.48** (1.20 to 1.81)	**212 more per 1.000** (from 88 more to 358 more)	⊕⊕◯◯ LOW
Deaths (time since drug started >180 days)											
1	observational studies	serious [i]	not serious	not serious	very serious [j]	none	62/5584 (1.1%)	34/3765 (0.9%)	**RR 1.23** (0.81 to 1.86)	**2 more per 1.000** (from 2 fewer to 8 more)	⊕◯◯◯ VERY LOW

CI: Confidence interval; **RR:** Risk ratio

Explanations

a. Risk of bias downgraded by one level: performance bias and detection bias cannot be excluded
b. Imprecision downgraded by one level: small sample size
c. Imprecision downgraded by two levels: small sample size and 95%-CI consistent with the possibility of harm and the possibility of benefit.
d. Risk of bias downgraded by one level: selection bias and performance bias cannot be excluded

e. Indirectness downgraded by one level: one study concerned a natural history study, in which physicians could prescribe whatever medication was therapeutically appropriate based on response to the initial medication; thus, some subjects may have been taking opioids and non-opioids at different times during the study.
f. Risk of bias downgraded by two levels: major concerns for confounding and selection bias; detection bias (i.e. lack of blinding) also cannot be excluded.
g. Imprecision downgraded by one level: low number of events
h. Opitmal information size criterium met (87 per group; $\alpha = 0.05$ and power = 80%)
i. Risk of bias downgraded by one level: major concern for selection bias; confounding and detection bias also cannot be excluded.
j. Imprecision downgraded by one level: low number of events and 95%-CI consistent with the possibility of harm and the possibility of benefit.

Supplementary Methods S2: Assessing the Certainty of Evidence (GRADE)

The GRADE approach considers the direct and size of effect estimates as well as factors that may affect the certainty in the estimates[1]. The certainty of evidence is graded for each outcome separately, i.e. a comparison of an intervention vs control may have different levels of evidence certainty based on the outcome assessed. Using this approach, one of the following levels of certainty of evidence is assigned for each outcome across studies.

High: We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very Low: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

The following domains were assessed for issues that may affect and lead to downgrading of the certainty of evidence:

Risk of bias: When all included trials were judged as "low" risk of bias (RoB) for the examined outcome, the evidence was not downgraded. The evidence was downgraded by one level when at least half of the trials included for an outcome had ≤ 3 RoB domains judged as "high or unclear". We downgraded the evidence by 2 points when more than half of the included trials for an outcome had more than three domains judged as "high or unclear" RoB.

Inconsistency: Inconsistency concerns an unexplained heterogeneity of results. When multiple studies show consistent effects, the certainty is highest for an outcome. Inconsistent effects across studies may be explained by differences in study populations (e.g. greater relative effects of drugs in sicker populations), interventions (e.g. larger effects due to higher drug doses) and outcomes (e.g. effects differing due to follow-up duration). Inconsistency was assessed by examining how much point estimates differed and to what extent the confidence intervals overlapped across studies. In addition, the I^2 statistic was used to quantify the proportion of variation in point estimates due to differences across studies. When heterogeneity was large (e.g. I^2 >75%) the certainty of evidence was downgraded by one point. The certainty of evidence was downgraded by two points in case of large heterogeneity and inconsistency arising from differences in population, interventions or outcomes.

Indirectness: The certainty of evidence may decrease when patients, interventions or outcomes differ from those of interest or when interventions are not tested in direct head-to-head comparisons. When the outcome studied is a surrogate for a different outcome, indirectness can also occur. Indirectness was assessed by examining if the research question addressed in this systematic review deviated from the available evidence concerning the study population, intervention, comparison or outcome. The certainty of evidence was downgraded by one point if there was indirectness ≤2 areas and by two points in case of indirectness in >2 areas.

Imprecision: Findings are imprecise when studies include relatively few patients or few events were observed, resulting in wide confidence intervals around the effect estimate. We determined whether sufficient information was available for making precise effect estimates by assessing the total number of participants and events. In addition, we examined whether the confidence interval around the effect estimate included consistent or contradictory conclusions, i.e. no effect *and* benefit or harm. We downgraded the certainty of evidence with one point when a) there were a total of <400 events (dichotomous outcomes) or 400 participants (continuous outcomes) across both intervention and control group, or b) when the 95% CI around the

pooled effect estimate included both no effect *and* benefit or harm. The evidence was downgraded by two levels when there was imprecision due to both (a) and (b).

Other considerations: Other aspects that were examined were the probability of publication bias and factors that may upgrade the evidence from non-randomized studies. We assessed whether all conducted studies addressing the research question were identified (i.e. the thoroughness of the literature search) and whether findings from inconclusive or negative studies that were not widely published appeared to be missing. As suggested by GRADE, the certainty of evidence was rated down by a maximum of one level when there was serious suspicion of publication bias. If the evidence from non-randomized studies was not downgraded for any of the domains (e.g. no risk of bias, no inconsistency, etc.), we assessed whether it could be additionally upgraded due to 1) a large magnitude of effect, 2) a dose-response effect, or 3) a plausible residual confounding effect (i.e. when all plausible residual, unaccounted confounding from non-randomized studies work to reduce the demonstrated effect or increase the effect, in case no effect was observed). None of the included non-randomized studies could be upgraded in our study.

References:

[1] Schünemann H, Brożek J, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. The GRADE Working Group, 2013. Available from: guidelinedevelopment.org/handbook.

Supplementary Figures (SF): Funnel plots of the CLBP trials

Figure SF1: Funnel plot of strong opioids compared to placebo for pain intensity

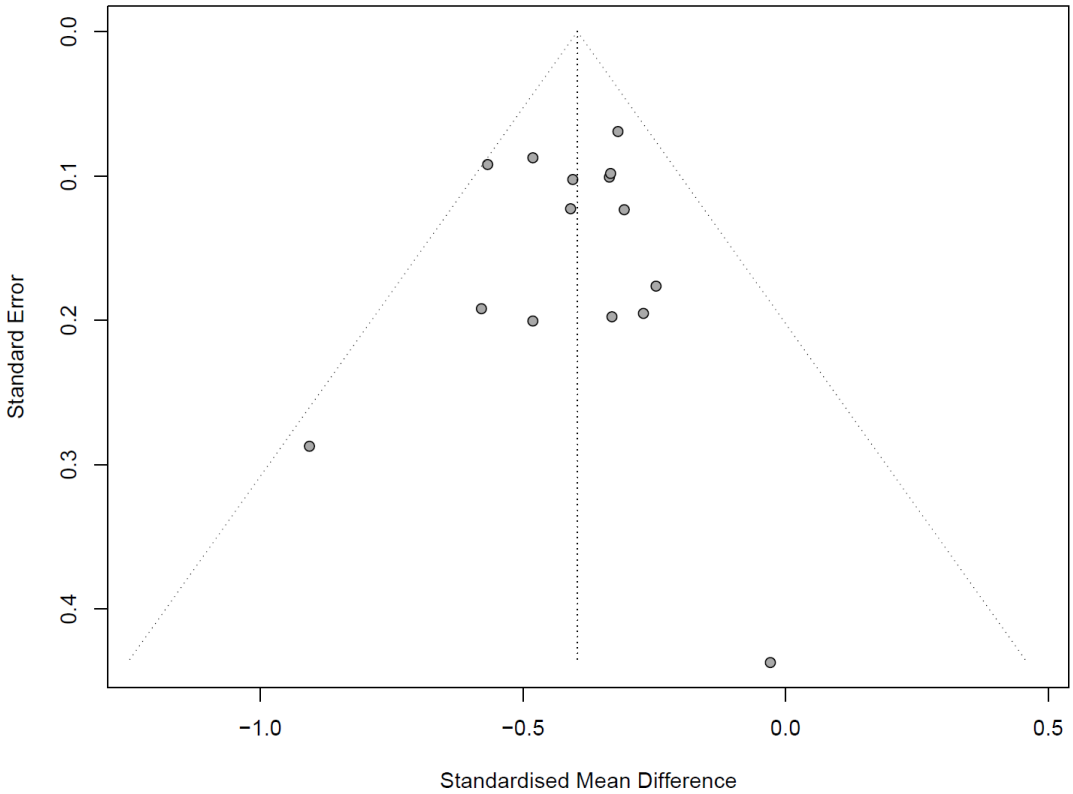


Figure SF2: Funnel plot of strong opioids compared to placebo for overall trial discontinuations

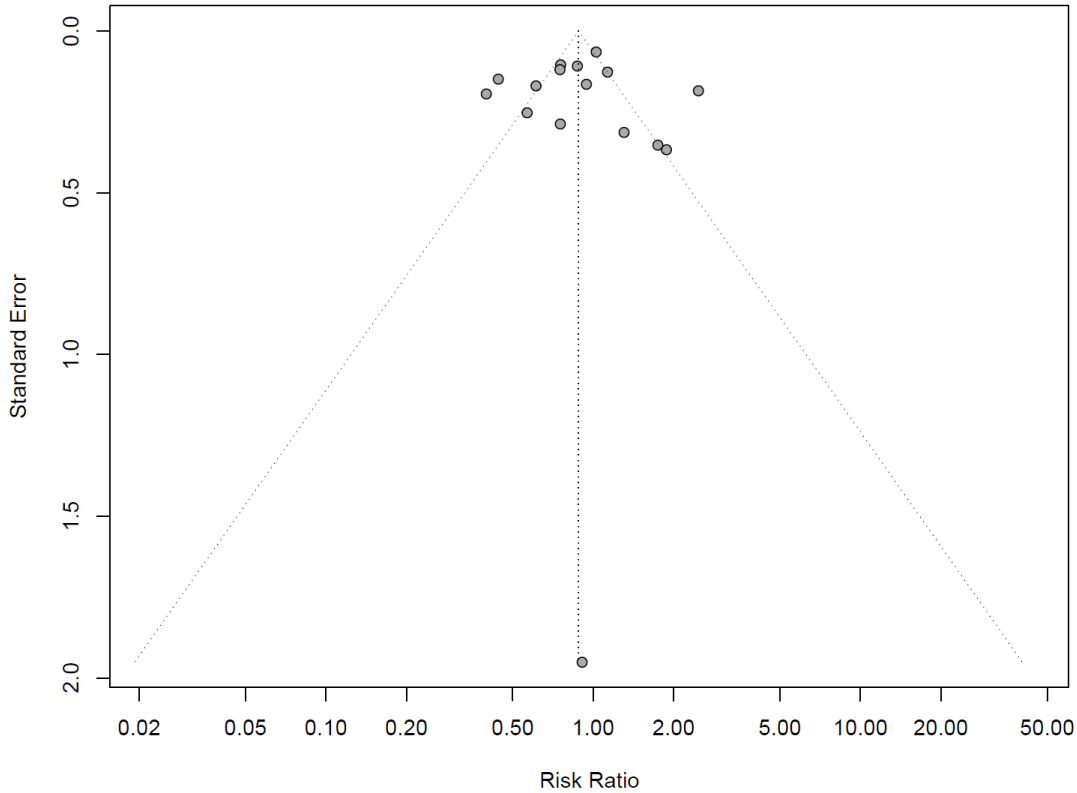


Figure SF3: Funnel plot of strong opioids compared to placebo for trial discontinuations due to AEs

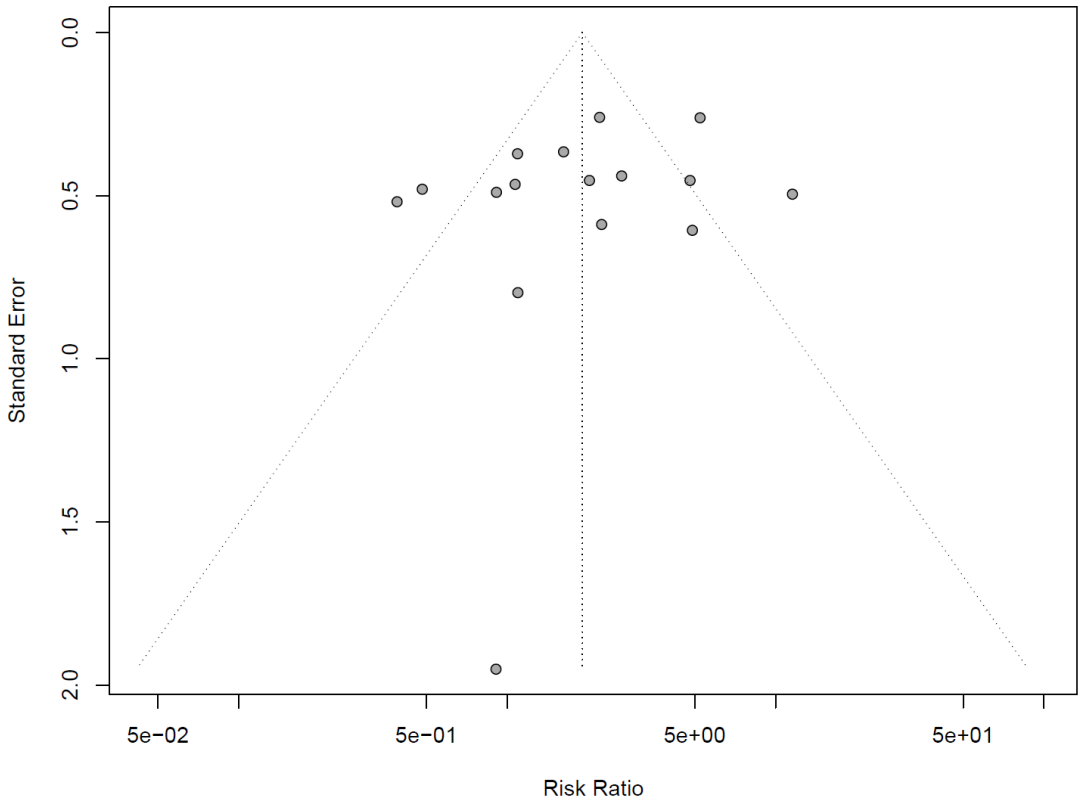


Figure SF4: Funnel plot of strong opioids compared to placebo for trial discontinuations due to efficacy lack

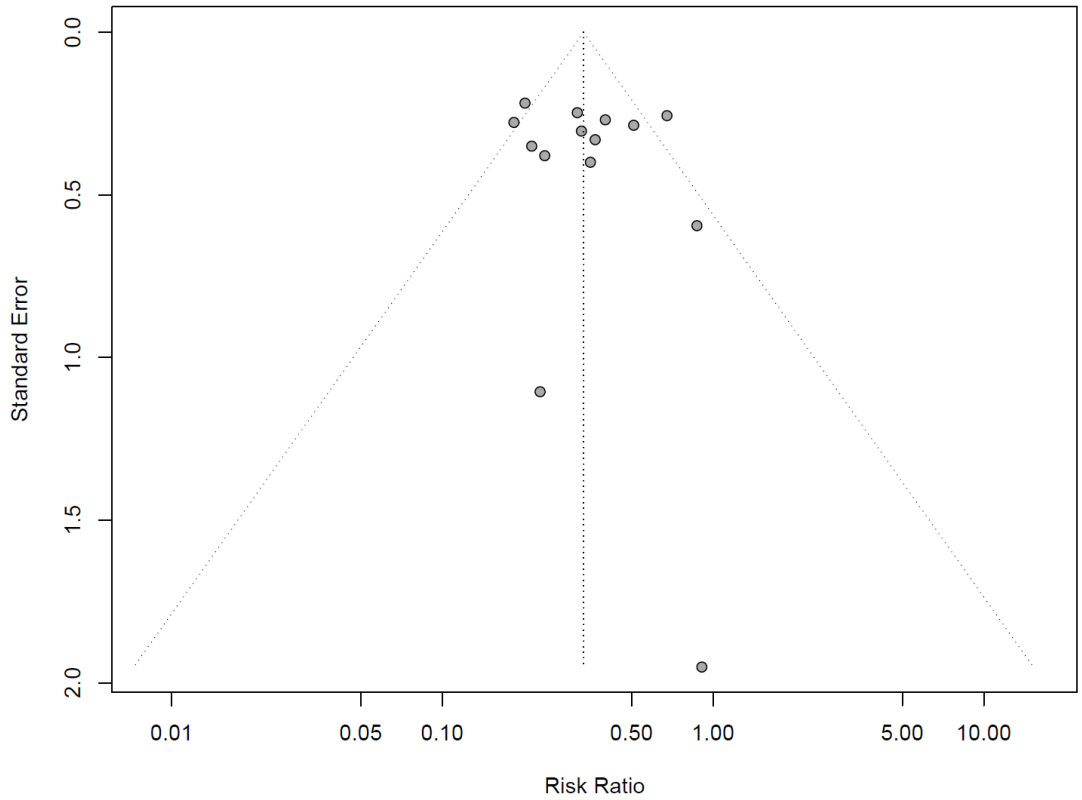


Figure SF5: Funnel plot of strong opioids compared to placebo for opioid withdrawal

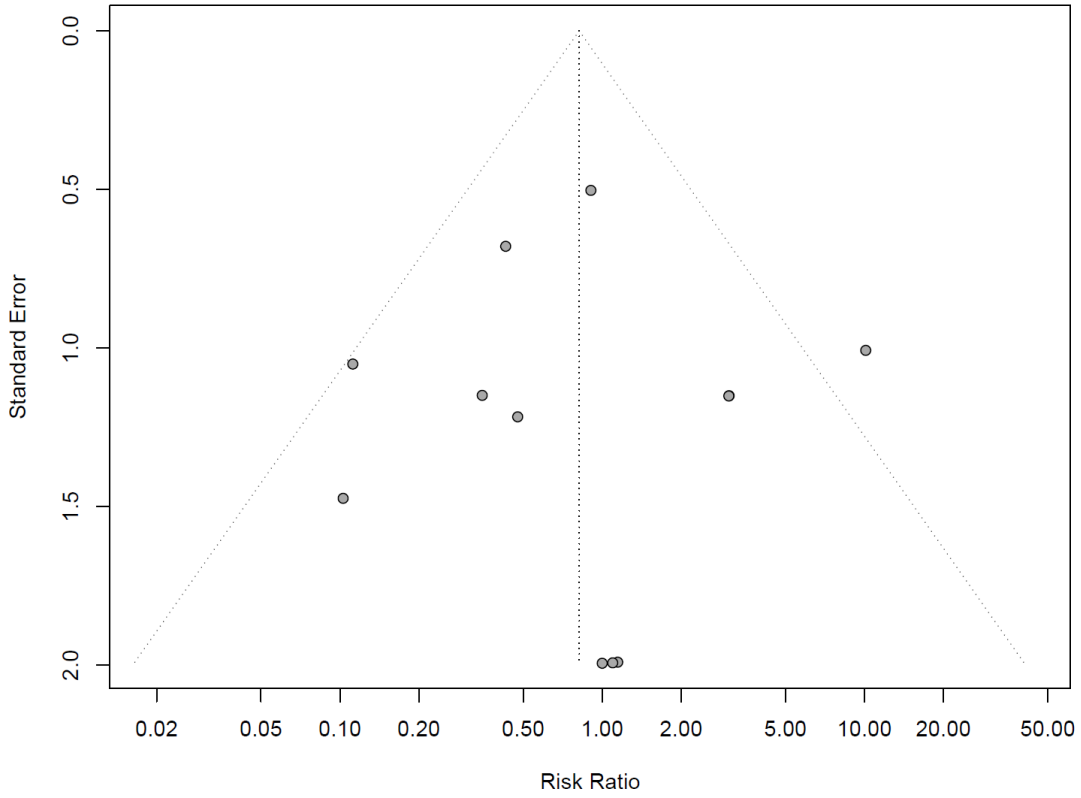


Figure SF6: Funnel plot of strong opioids compared to placebo for any adverse events

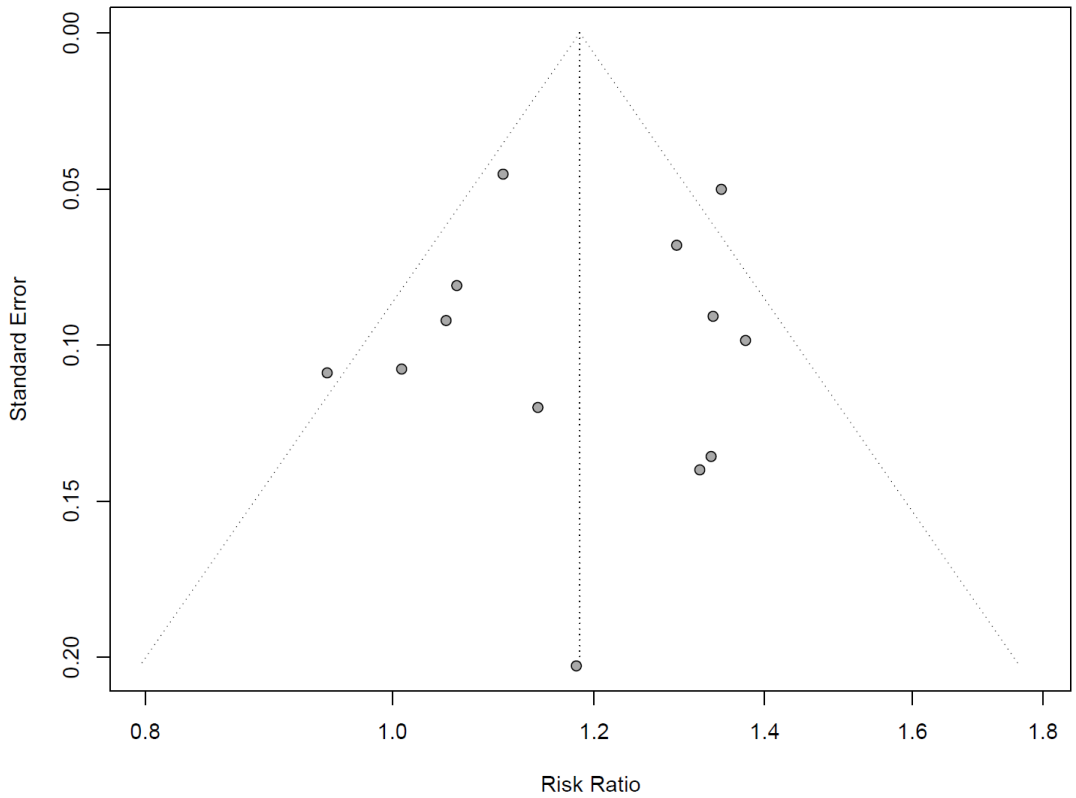


Figure SF7: Funnel plot of strong opioids compared to placebo for serious adverse events

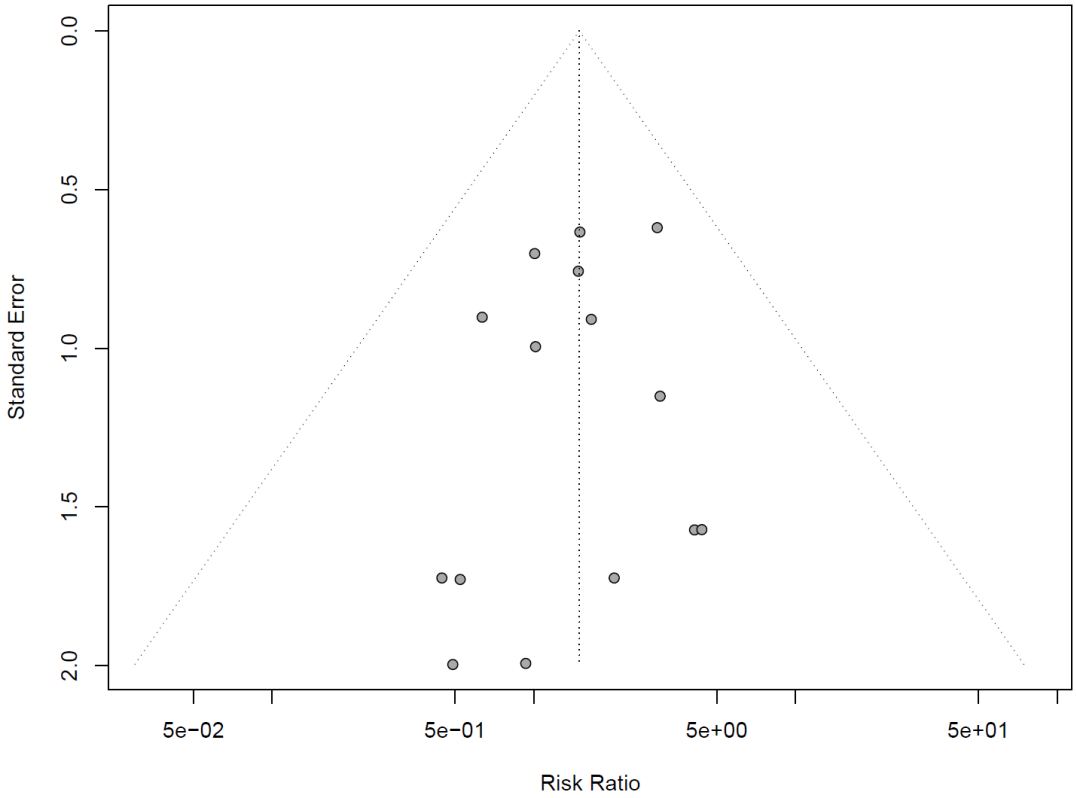


Figure SF8: Funnel plot of strong opioids compared to placebo for nausea

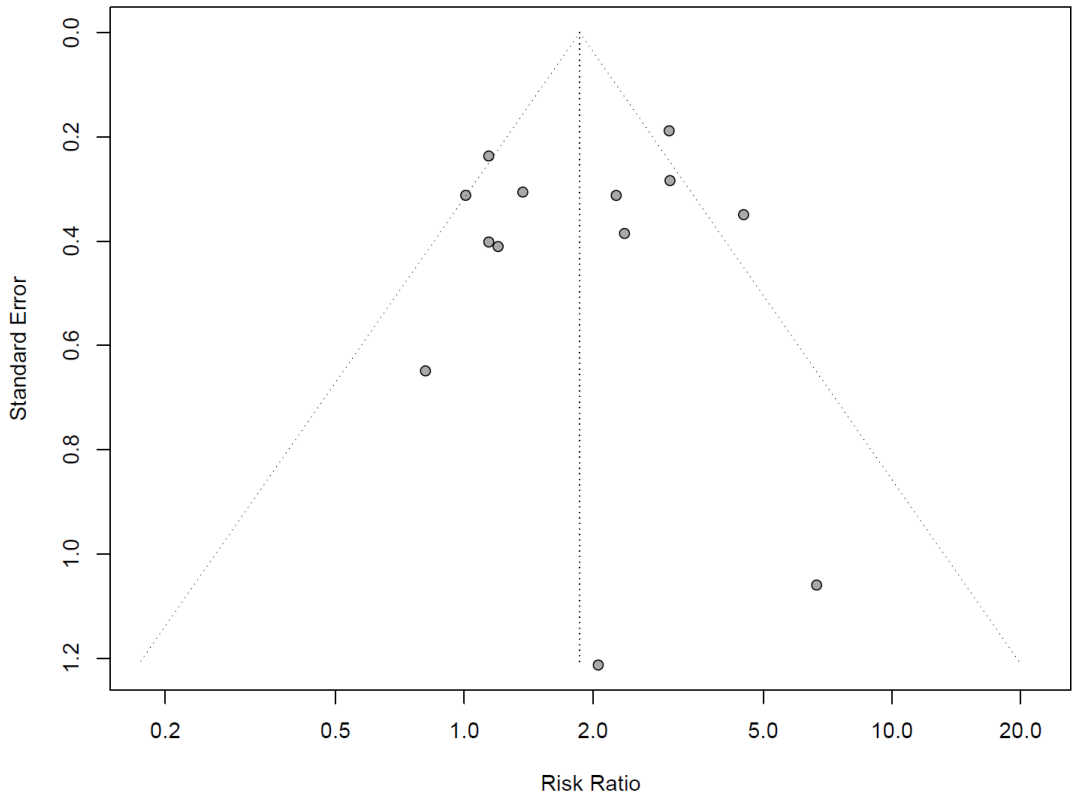


Figure SF9: Funnel plot of strong opioids compared to placebo for constipation

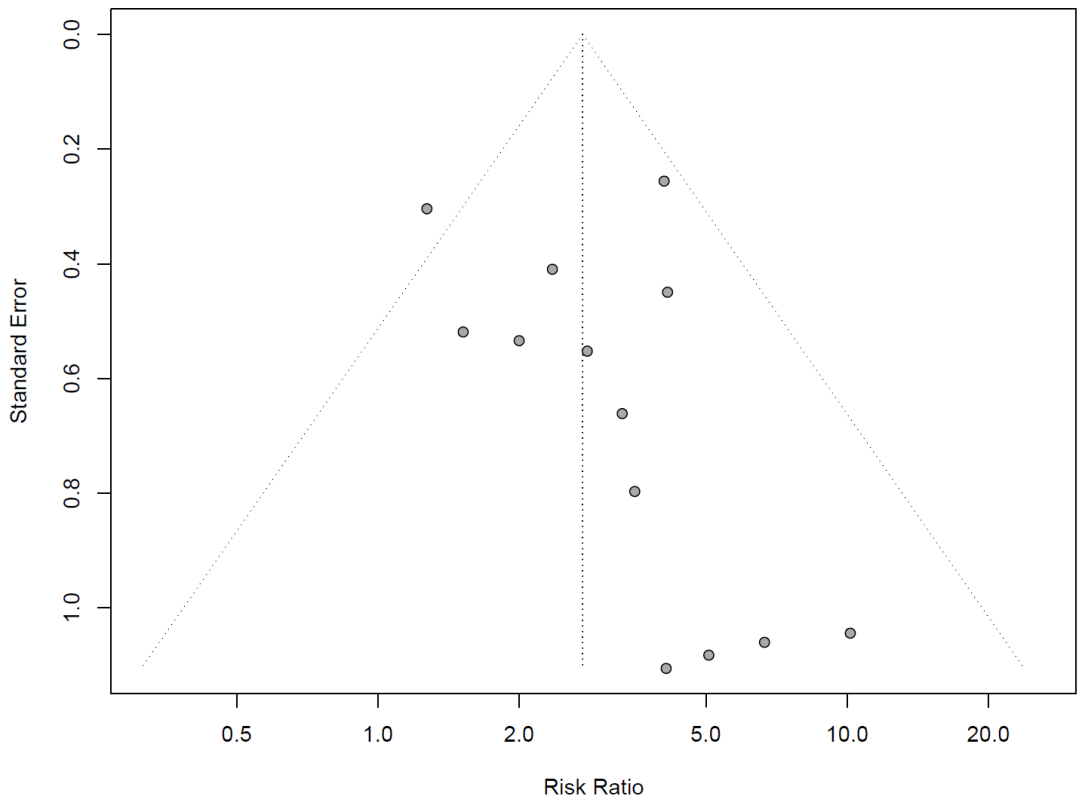


Figure SF10: Funnel plot of strong opioids compared to placebo for vomiting

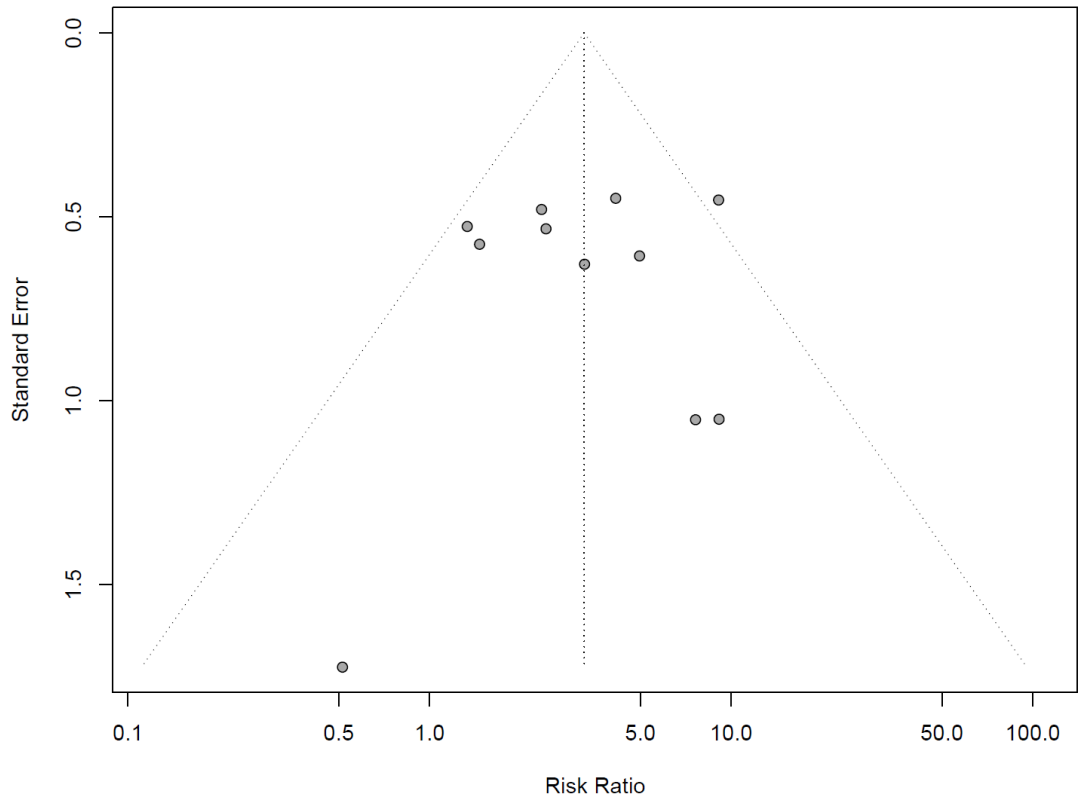


Figure SF11: Funnel plot of strong opioids compared to placebo for dizziness

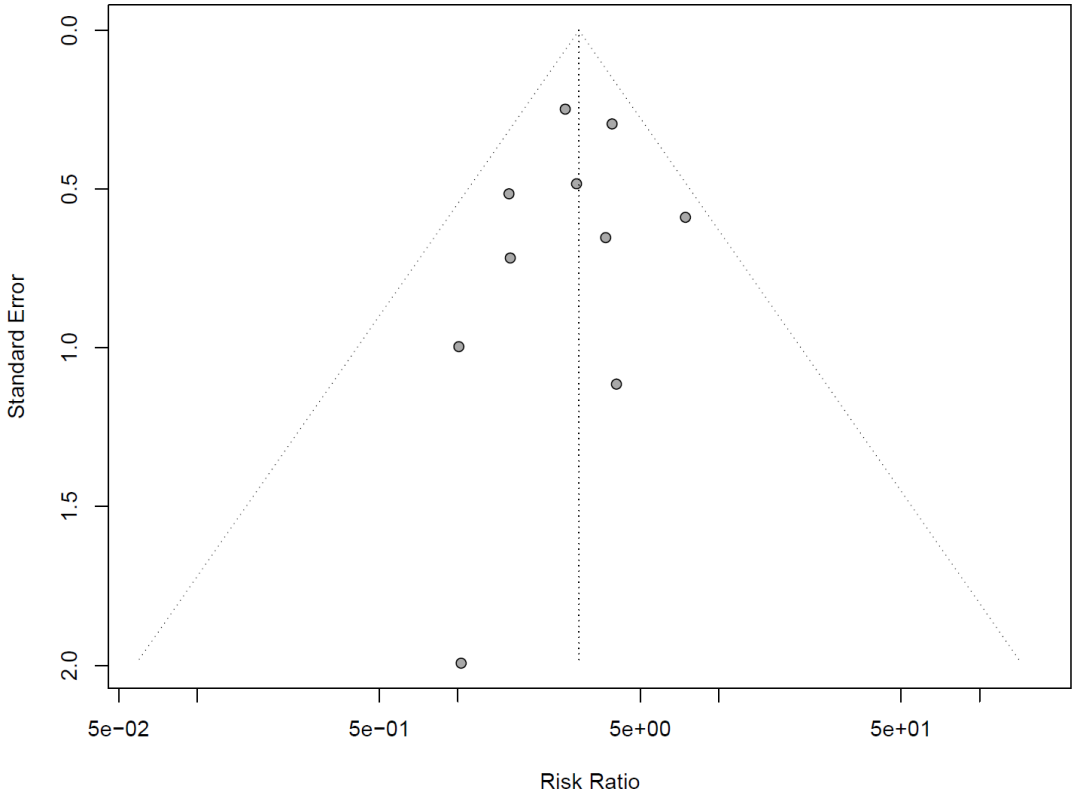


Figure SF12: Funnel plot of strong opioids compared to placebo for somnolence

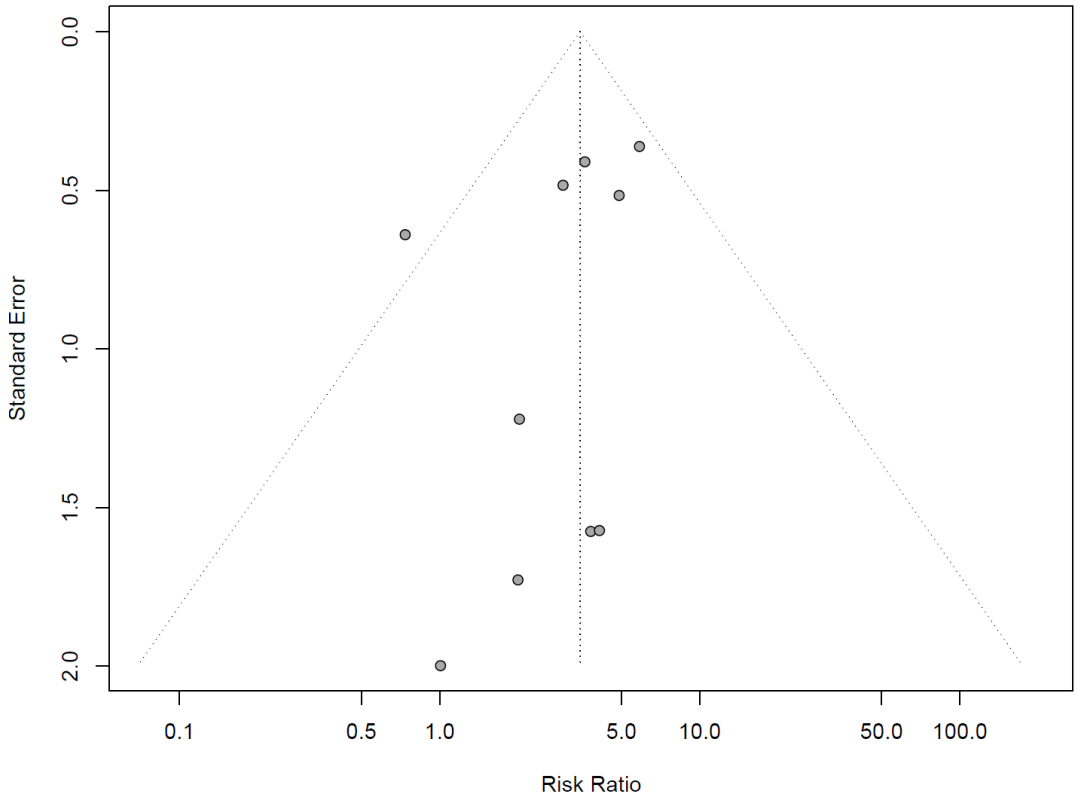
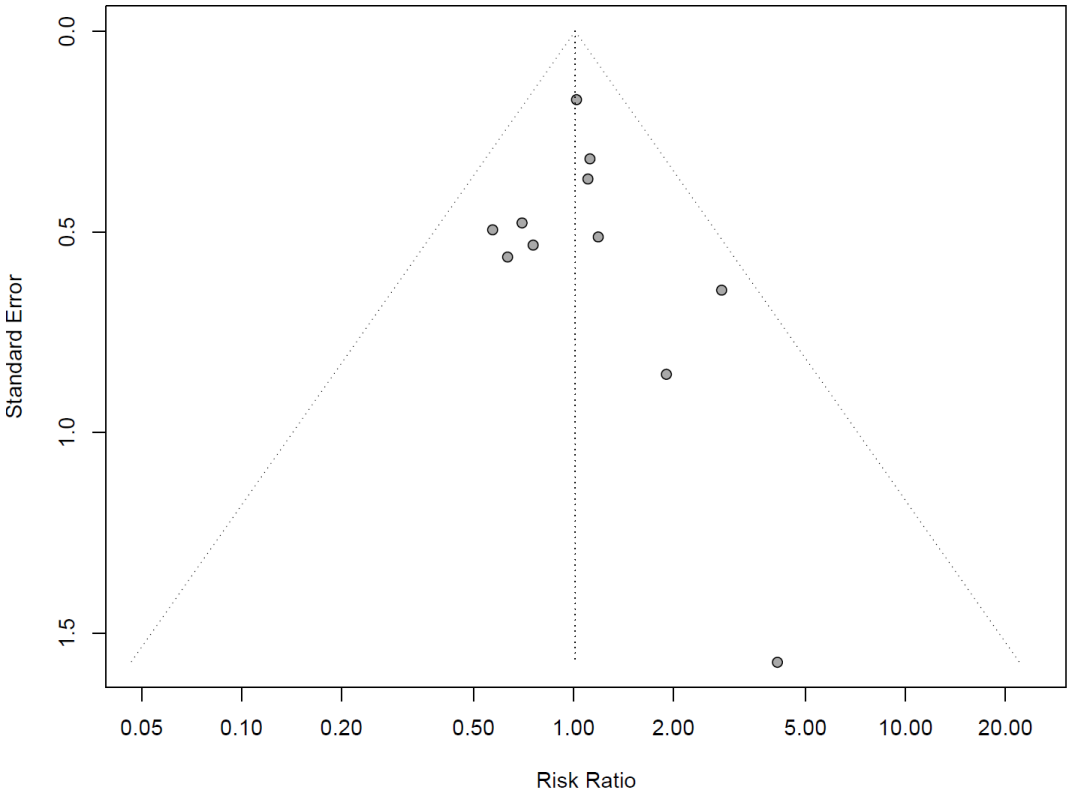


Figure SF13: Funnel plot of strong opioids compared to placebo for headache


Standard Error
0.0
0.5
1.0
1.5
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
20.00
Risk Ratio