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 fentanest or fentora or fortral or duramorph 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 I dromoran or levodroman or levorphan* or or amaroph 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 tramadol or tiral or topalgic or tradol or tradol or tradol or tadol or tadolor tramadol or tamadol or tramadorsch or tramador or ultram or zamudol or zumalgic 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

Study	Total	Op Mean	oioids SD	Total	Pl Mean	acebo SD	Stand D	lardised M Difference	lean	SMD	95%-CI	Weight
Follow–up ≥ 3months Christoph 2017 Random effects model Heterogeneity: not applica Test for effect in subgroup:	(at 14 v 257 257 ble z = 3.0	veeks) 29.64 5 (p = 0	31.70	100 100	18.00	32.93		.		0.36 0.36	[0.13; 0.59] [0.13; 0.59]	72.0% 72.0%
Follow-up <3 months Gordon 2010b Random effects model Heterogeneity: not applica Test for effect in subgroup:	(at 8 we 52 52 ble z = 0.6	eeks) 18.00 5 (p = 0	26.06 0.519)	52 52	14.60	27.20				0.13 0.13	[-0.26; 0.51] [-0.26; 0.51]	28.0% 28.0%
Random effects model Heterogeneity: $I^2 = 5\%$, τ^2 Residual heterogeneity: I^2 Test for overall effect: $z = 2$ Test for subgroup difference	309 = 0.001 = NA%, 2.80 (<i>p</i> = ees: χ ₁ ² =	5, p = 0 , p = NA = 0.005; : 1.06, c	0.30 A) If = 1 (<i>p</i>	152	0)	-	2 −1 Favors Plac	0 ebo Favo	1 2 ors Opioids	0.30	[0.09; 0.50]	100.0%

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

		0	pioids		Pla	acebo	Standardised Mea	n		
Study	Total	Mean	SD	Total	Mean	SD	Difference	SMD	95%-Cl	Weight
Follow-up <3 months (max. a	t 8 wee	eks)							
Gordon 2010a	53	-13.81	13.15	53	-12.60	13.35		-0.09	[-0.47; 0.29]	33.0%
Gordon 2010b	52	-20.90	13.02	52	-10.60	14.09	< +	-0.75	[-1.15; -0.36]	32.0%
Kawamata 2019	62	-0.10	1.57	68	0.30	1.65		-0.25	[-0.59; 0.10]	35.1%
Random effects model Heterogeneity: $I^2 = 67\%$, τ^2	167 2 = 0.074	42, p = 0	0.05	173				-0.36	[-0.73; 0.02]	100.0%
Test for overall effect: $z = -1.86 (p = 0.063)$							-1 -0.5 0 0.5 Favors Opioids Favors F	1 Placebo		

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

Study Events Total Risk Ratio RR 95%-CI We Follow-up ≥3 months (max. at 15 weeks) 1 1 1 1 1	6.9% 7.0% 3.7%
Follow-up ≥3 months (max. at 15 weeks)	6.9% 7.0% 3.7%
Dumpel: 2040 452.0. 240 04.0. 460 0.00 [0.75; 4.00] 6	6.9% 7.0% 3.7%
Buynak 2010 152.0 318 84.0 159 📮 0.90 [0.75, 1.09] 6	7.0% 3.7%
Buynak 2010 195.0 328 83.0 160 1.15 [0.96; 1.36] 7	3.7%
Christoph 2017 49.0 126 7.0 32 1.78 [0.89; 3.55] 3	
Christoph 2017 76.0 130 6.0 32 3.12 [1.49; 6.51] 3	3.5%
Christoph 2017 67.0 128 6.0 31 2.70 [1.29; 5.65] 3	3.5%
Christoph 2017 61.0 131 6.0 31 2.41 [1.15; 5.05] 3	3.4%
Gimbel 2016 48.0 254 110.0 257 0.44 [0.33; 0.59] 6	5.3%
Hale 2007 21.0 70 55.0 73 🛨 0.40 [0.27; 0.58] 5	5.7%
Hale 2010 68.0 134 90.0 134 0.76 [0.62; 0.93] 6	6. 8%
Katz 2007 34.0 105 53.0 100 10.061 [0.44; 0.85] 6	5.0%
Katz 2015 71.0 193 96.0 196 0.75 [0.59; 0.95] 6	5. 7%
Rauck 2016 54.0 229 58.0 232 0.94 [0.68; 1.30] 6	5.1%
Steiner 2011 86.0 256 84.0 283 1.13 [0.88; 1.45] 6	5.6%
vonDrackova 2008 18.0 151 25.0 158 0.75 [0.43; 1.32]	4.5%
Webster 2006 105.0 206 59.0 101 0.87 [0.71; 1.08] 6	6.8%
Random effects model 2759 1979 4 0.93 [0.76; 1.14] 83	3.5%
Heterogeneity: $I^2 = 85\%$, $\tau^2 = 0.1189$, $p < 0.01$	
Test for effect in subgroup: $z = -0.73$ ($p = 0.466$)	
Follow-up <3 months (max. at 8 weeks)	
Chu 2012 18.0 69 14.0 70 1.30 [0.71; 2.41]	4.1%
Gordon 2010a 19.0 73 9.0 65 1.88 [0.92; 3.86] 3	3.5%
Gordon 2010b 19.0 74 10.0 68 1.75 [0.87; 3.49]	3.7%
Kawamata 2019 16.0 62 31.0 68 - 0.57 [0.34; 0.93]	4.9%
Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] (0.2%
Random effects model 289 281 - 1.19 [0.68; 2.09] 16	5.5%
Heterogeneity: $J^2 = 64\%$, $\tau^2 = 0.2344$, $p = 0.03$	
Test for effect in subgroup: $z = 0.61$ ($p = 0.540$)	
Random effects model 3048 2260 0.97 [0.80; 1.16] 100	0.0%
Heterogeneity: $I^2 = 82\%, \tau^2 = 0.1226, p < 0.01$	
Residual heterogeneity: $I^2 = 82\%$, $p < 0.01$ 0.01 0.1 0.51 2 10 100 Test for overall effect: $z = -0.37$ ($p = 0.710$) Eavors Onioids Eavors Placebo	
Test for subgroup differences: $\gamma_{r}^{2} = 0.68$, df = 1 (p = 0.41)	

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

Study Events Total Risk Ratio RR 95%-Cl Weight Follow-up ≥ 3 months (max. at 15 weeks) Buynak 2010 53.0 318 7.0 159 Buynak 2010 106.0 328 8.0 160 6.46 [3.23] 12.93] 6.64% Christoph 2017 32.0 126 1.0 32 8.38 [1.19, 58.98] 2.9% Christoph 2017 52.0 128 1.0 31 1.526 [2.20] 105.93] 2.9% Gimbel 2016 5.0 254 13.0 257 0.39 [0.14] 1.08] 5.5% Hale 2007 7.0 70 8.0 100 107 [0.43] 2.38] 5.7% Katz 2015 14.0 193 13.0 196 1.07 [0.43] 2.67 5.8% VonDrackova 2008 6.0 151 13.0 196 1.09 [0.53] 2.27] 6.5% Gordon 2010a 18.0 7.3 6.0 65 1.01<		Opioids	s Pla	icebo						
Follow-up ≥3 months (max. at 15 weeks) Buynak 2010 53.0 318 7.0 159 Buynak 2010 106.0 328 8.0 160 Christoph 2017 13.0 126 10 32 Christoph 2017 52.0 128 10 31 Christoph 2017 52.0 128 10 31 Christoph 2017 62.0 128 1259 1.81; 87.59 2.9% Gimbel 2016 5.0 22.5 17; 7.13; 5.0% 5.26 2.20; 105.93 2.259 1.9% Katz 2007 9.0 105 8.0 100 1.07; 0.43; 2.27] 6.5% Katz 2016 14.0 129 7.0 2.28 1.03; 3.28 7.2% Veebser 2006 49.0 2.06 10.0 70 7.8 4	Study	Events Tota	Events	Total	Risk Ratio	RR	95%-CI	Weight		
Buynak 2010 53.0 318 7.0 159 Buynak 2010 106.0 328 8.0 160 Christoph 2017 42.0 131 1.0 31 Christoph 2017 52.0 128 1.0 31 Christoph 2017 52.0 128 1.0 31 Christoph 2017 62.0 130 1.0 32 Gimbel 2016 5.0 2.54 13.0 257 Hale 2010 9.0 134 4.0 134 2.25 10.71 7.13 5.0% Katz 2015 14.0 193 13.0 196 1.07 [0.43; 2.67] 5.8% Katz 2016 14.0 2.9 7.0 2.02 2.03 [0.83; 4.93] 5.9% Rauck 2016 14.0 2.9 <	Follow-up ≥3 months	(max. at 15 we	eks)							
Buynak 2010 106.0 328 8.0 160 Christoph 2017 42.0 131 1.0 31 Christoph 2017 33.0 126 1.0 32 Christoph 2017 52.0 128 1.0 31 Christoph 2017 62.0 130 1.0 32 Gimbel 2016 5.0 254 13.0 257 Hale 2007 7.0 70 8.0 73 Hale 2010 9.0 134 4.0 134 Christoph 2017 6.2 0.130 1.0 32 Gimbel 2016 5.0 254 13.0 257 Hale 2010 9.0 134 4.0 134 Katz 2007 9.0 105 8.0 100 follow-up 2016 14.0 229 7.0 232 Construction 2016 14.0 229 7.0 232 Construction 2016 151 13.0 158 Webster 2006 49.0 206 5.0 101 Test for effect in subgroup: $z = 3.16 (p = 0.002)$ Random effects model 289 Christoph 2017 $(p = 0.002)$ Random effects model 289 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01 Random effects model 289 Heterogeneity: l^2 = 70\%, r^2 = 0.5477, p < 0.01 Random effects model 289 Heterogeneity: l^2 = 70\%, r^2 = 0.5477, p < 0.01 Random effects model 289 Heterogeneity: l^2 = 70\%, r^2 = 0.5477, p < 0.01 Test for effect in subgroup: z = 3.17 (p = 0.002)Chartoph 2010 2000 2000 2000Christoph 2010 2000 2000 2000 2000 2000 2000 200$	Buynak 2010	53.0 318	3 7.0	159		3.79	[1.76; 8.13]	6.4%		
Christoph 2017 42.0 131 1.0 31 Christoph 2017 33.0 126 1.0 32 Christoph 2017 52.0 128 1.0 31 Christoph 2017 52.0 128 1.0 31 Christoph 2017 52.0 128 1.0 32 Christoph 2017 62.0 130 1.0 32 Christoph 2017 62.0 130 1.0 32 Christoph 2017 62.0 128 1.0 31 Christoph 2017 62.0 128 1.0 32 Christoph 2017 62.0 128 1.0 32 Christoph 2017 62.0 128 1.0 32 Christoph 2017 52.0 128 1.0 31 Christoph 2017 52.0 128 1.0 32 Christoph 2017 52.0 128 1.0 32 Christoph 2017 52.0 128 1.0 32 Christoph 2017 52.0 128 1.0 32 Hale 2007 7.0 70 8.0 73 Hale 2010 9.0 134 4.0 134 Chaz 2010 9.0 105 8.0 100 1.07 [0.43; 2.67] 5.8% Katz 2015 1.4.0 193 13.0 196 1.09 [0.53; 2.27] 6.5% Katz 2015 1.4.0 193 13.0 196 1.09 [0.53; 2.27] 6.5% Katz 2016 1.4.0 229 7.0 232 Christoph 2018 4.0 136 1.0 158 Christoph 2017 7.0 7.0 8.0 70 Steiner 2011 4.0.0 256 2.0.0 283 VonDrackova 2008 6.0 151 13.0 158 Chu 2012 16.0 69 10.0 70 Test for effect in subgroup: $z = 3.16 (p = 0.002)$ Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 0.5 11 0.5 10 Charcon effects model 289 281 Charcon effects model 280 281 Charcon effects model 280 281 Charcon effects model 280 281 Charcon effects model 280 281 Charc	Buynak 2010	106.0 328	8.0	160		6.46	[3.23; 12.93]	6.6%		
Christoph 2017 33.0 126 1.0 32 Christoph 2017 52.0 128 1.0 31 Christoph 2017 52.0 128 1.0 32 Gimbel 2016 5.0 254 13.0 257 Hale 2007 7.0 70 8.0 73 Hale 2010 9.0 134 4.0 134 Katz 2007 9.0 105 8.0 100 1.07 [0.43; 2.67] 5.8% Katz 2015 14.0 193 13.0 196 Katz 2015 14.0 229 7.0 232 Steiner 2011 40.0 256 20.0 283 VonDrackova 2008 6.0 151 13.0 158 Webster 2006 49.0 206 5.0 101 Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ Test for effect in subgroup: $z = 3.16$ ($p = 0.002$) Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Lin 2016 0.5 11 0.5 11 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2002) Random effects model 2002) Random effects model 2004 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2004 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2004 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2004 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2005 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2004 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 2005 Heterogeneity: l^2	Christoph 2017	42.0 13	1.0	31		9.94	[1.42; 69.45]	2.9%		
Christoph 2017 52.0 128 1.0 31 Christoph 2017 62.0 130 1.0 32 Gimbel 2016 5.0 254 13.0 257 Hale 2007 7.0 70 80 73 Hale 2010 9.0 134 4.0 134 Katz 2007 9.0 105 8.0 100 Katz 2015 14.0 193 13.0 196 Katz 2015 14.0 193 13.0 196 Katz 2016 14.0 229 7.0 232 Log 10.83; 2.67] 5.8% Katz 2016 14.0 229 7.0 232 Log 10.83; 3.68] 7.2% Webster 2006 49.0 206 5.0 101 Test for effect in subgroup: $z = 3.16 (p = 0.002)$ Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 0.5 11 0.5 10 Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 0.5 11 0.5 10 Chu 2012 16.0 69 10.0 70 Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 0.5 11 0.5 10 Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.5477$, $p < 0.01$ Test for effect in subgroup: $z = 3.17 (p = 0.002)$ Fandom effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Test for effect in subgroup: $z = 3.17 (p = 0.002)$ Fandom effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 289 281 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Random effects model 3048 H	Christoph 2017	33.0 126	5 1.0	32		8.38	[1.19; 58.98]	2.9%		
Christoph 2017 62.0 130 1.0 32 Gimbel 2016 50 254 13.0 257 Hale 2007 7.0 70 8.0 73 Hale 2010 9.0 134 4.0 134 Hale 2010 9.0 134 4.0 134 Katz 2015 14.0 193 13.0 196 Rauck 2016 14.0 229 7.0 232 vonDrackova 2008 6.0 151 13.0 158 Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 1.40 0.25 1.1 0.5 10 Chu 2012 16.0 69 10.0 70 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 1.51 1.05 10 Random effects model 289 Lin 2016 0.5 1.1 0.5 10 Random effects model 289 Lin 2016 0.5 1.1 0.5 10 Random effects model 289 Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = 0.5477$, $p < 0.01$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $r^2 = $	Christoph 2017	52.0 128	3 1.0	31		12.59	[1.81; 87.59]	2.9%		
Gimbel 2016 5.0 254 13.0 257 0.39 $[0.14]$; 1.08] 5.5% Hale 2007 7.0 7.0 7.0 8.0 73 0.91 $[0.35]$; 2.38] 5.7% Hale 2010 9.0 134 4.0 134 225 $[0.71]$; 7.13] 5.0% Katz 2007 9.0 105 8.0 100 107 $[0.43]$; 2.67] 5.8% Katz 2015 14.0 193 13.0 196 1.09 $[0.53]$; 2.27] 6.5% Rauck 2016 14.0 229 7.0 232 203 $[0.83]$; 4.93] 5.9% Steiner 2011 40.0 256 20.0 283 2.21 $[1.38]$; 11.69] 5.9% Random effects model 2759 1979 2.31 $[1.37]$; 3.87] 77.8% Heterogeneity: $l^2 = 77\%$, $r^2 = 0.7251$, $p < 0.01$ 70 76 76 76 77.8% Gordon 2010a 18.0 73 6.0 65 2.67 [1.33] 6.3% Kawamata 2019 3.0 62 3.0 68 99	Christoph 2017	62.0 130) 1.0	32		15.26	[2.20; 105.93]	2.9%		
Hale 2007 7.0	Gimbel 2016	5.0 254	13.0	257		0.39	[0.14; 1.08]	5.5%		
Hale 2010 9.0 134 4.0 134 Katz 2007 9.0 105 8.0 100 1.07 [0.43; 2.67] 5.8% Katz 2015 14.0 193 13.0 196 1.09 [0.53; 2.27] 6.5% Steiner 2011 40.0 256 20.0 283 2.21 [1.33; 3.68] 7.2% vonDrackova 2008 6.0 151 13.0 158 0.48 [0.19; 1.24] 5.7% Webster 2006 49.0 206 5.0 101 4.80 [1.98; 11.69] 5.9% Random effects model 2759 1979 2.31 [1.37; 3.87] 77.8% Heterogeneity: $l^2 = 77\%, r^2 = 0.7251, p < 0.01$	Hale 2007	7.0 70	0.8	73		0.91	[0.35; 2.38]	5.7%		
Katz 2007 9.0 105 8.0 100 1.07 $[0.43; 2.67]$ 5.8% Katz 2015 14.0 193 13.0 196 1.09 $[0.53; 2.27]$ 6.5% Rauck 2016 14.0 229 7.0 232 2.03 $[0.83; 4.93]$ 5.9% Steiner 2011 40.0 256 20.0 283 2.21 $[1.33; 3.68]$ 7.2% wonDrackova 2008 6.0 151 13.0 158 0.48 $[0.19; 1.24]$ 5.7% Webster 2006 49.0 206 5.0 101 4.80 $[1.98; 11.69]$ 5.9% Random effects model 2759 1979 2.31 $[1.37; 3.87]$ 77.8% Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 73 6.0 65 2.67 $[1.13; 6.32]$ 6.0% Gordon 2010a 18.0 73 6.0 68 1.10 $[0.22; 5.23]$ 3.8% Lin 2016 0.5 11 0.5 10 0.91 $[0.02; 41.68]$ 1.0% Random effects model 289 281 2260 <t< td=""><td>Hale 2010</td><td>9.0 134</td><td>4.0</td><td>134</td><td></td><td>2.25</td><td>[0.71; 7.13]</td><td>5.0%</td></t<>	Hale 2010	9.0 134	4.0	134		2.25	[0.71; 7.13]	5.0%		
Katz 2015 14.0 193 13.0 196 1.09 [0.53; 2.27] 6.5% Rauck 2016 14.0 229 7.0 232 203 [0.83; 4.93] 5.9% Steiner 2011 40.0 256 20.0 283 2.21 [1.33; 3.68] 7.2% wonDrackova 2008 6.0 151 13.0 158 0.48 [0.19; 1.24] 5.7% Webster 2006 49.0 206 5.0 101 4.80 [1.96] 1.69 5.9% Random effects model 2759 1979 2.31 [1.37; 3.87] 77.8% Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 73 6.0 65 2.67 [1.13; 6.32] 6.0% Gordon 2010a 18.0 73 6.0 68 4.90 [1.49; 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 4.90 [1.49; 16.08] 4.9% In 2016 0.5 11 0.5 10 91 [0.02; 41.68] 1.0% Random effects model 289 281 2260 224 </td <td>Katz 2007</td> <td>9.0 105</td> <td>i 8.0</td> <td>100</td> <td></td> <td>1.07</td> <td>[0.43; 2.67]</td> <td>5.8%</td>	Katz 2007	9.0 105	i 8.0	100		1.07	[0.43; 2.67]	5.8%		
Rauck 2016 14.0 229 7.0 232 2.03 $[0.83; 4.93]$ 5.9% Steiner 2011 40.0 256 20.0 283 2.21 $[1.33; 3.68]$ 7.2% vonDrackova 2008 6.0 151 13.0 158 0.48 $[0.19; 1.24]$ 5.7% Webster 2006 49.0 206 5.0 101 4.80 $[1.98; 11.69]$ 5.9% Random effects model 2759 1979 167 2.31 $[0.79; 3.32]$ 6.5% Gordon 2010a 18.0 73 6.0 65 2.67 $[1.13; 6.32]$ 6.0% Gordon 2010b 16.0 74 3.0 68 4.90 $[1.49; 16.08]$ 4.9% Kawamata 2019 3.0 62 3.0 68 4.90 $[1.49; 16.08]$ 4.9% In 2016 0.5 11 0.5 10 1.0 $[0.23; 52.3]$ 3.8% In 2016 0.5 11 0.5 10 2260 2.24 $[1.48; 3.38]$ 100.0% Random effects model 289 281 2260<	Katz 2015	14.0 193	3 13.0	196		1.09	[0.53; 2.27]	6.5%		
Steiner 2011 40.0 256 20.0 283 vonDrackova 2008 6.0 151 13.0 158 2.21 [1.33; 3.68] 7.2% Webster 2006 49.0 206 5.0 101 4.80 [1.98; 11.69] 5.9% Random effects model 2759 1979 2.31 [1.37; 3.87] 77.8% Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 16.0 69 10.0 70 1.62 [0.79; 3.32] 6.5% Gordon 2010a 18.0 73 6.0 65 2.67 [1.13; 6.32] 6.0% Gordon 2010b 16.0 74 3.0 68 4.90 [1.49; 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 1.10 [0.23; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 281 224 [1.48; 3.38] 100.0% Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ 0.01 0.1 0.51 2 10 1	Rauck 2016	14.0 229) 7.0	232		2.03	[0.83; 4.93]	5.9%		
vonDrackova 2008 6.0 151 13.0 158 0.48 [0.19; 1.24] 5.7% Webster 2006 49.0 206 5.0 101 4.80 [1.98; 11.69] 5.9% Random effects model 2759 1979 2.31 [1.37; 3.87] 77.8% Heterogeneity: $l^2 = 77%, r^2 = 0.7251, p < 0.01$ 16.0 69 10.0 70 1.62 [0.79; 3.32] 6.5% Gordon 2010a 18.0 73 6.0 65 2.67 [1.13; 6.32] 6.0% Gordon 2010b 16.0 74 3.0 68 4.90 [1.49; 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 1.10 [0.23; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 281 2260 2.15 [1.34; 3.38] 100.0% Heterogeneity: $l^2 = 70\%, r^2 = 0.5477, p < 0.01$	Steiner 2011	40.0 256	5 20.0	283		2.21	[1.33; 3.68]	7.2%		
Webster 2006 49.0 206 5.0 101 Random effects model 2759 1979 Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 1979 Follow-up <3 months (max. at 8 weeks)	vonDrackova 2008	6.0 151	13.0	158		0.48	[0.19; 1.24]	5.7%		
Random effects model 2759 1979 Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 1979 2.31 [1.37; 3.87] 77.8% Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ 16.0 69 10.0 70 1.62 [0.79; 3.32] 6.5% 1.62 [0.79; 3.32] 6.5% Chu 2012 16.0 69 10.0 70 1.62 [0.79; 3.32] 6.5% 1.62 [0.79; 3.32] 6.5% Gordon 2010a 18.0 73 6.0 65 2.67 [1.13; 6.32] 6.0% Kawamata 2019 3.0 62 3.0 68 1.10 [0.23; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48; 3.38] 100.0% Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.1 0.51 2 10 110 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) Favors Opioids Favors Placebo 2.24 [1.48; 3.38] 100.0%	Webster 2006	49.0 206	5.0	101		4.80	[1.98; 11.69]	5.9%		
Heterogeneity: $l^2 = 77\%$, $\tau^2 = 0.7251$, $p < 0.01$ Test for effect in subgroup: $z = 3.16$ ($p = 0.002$) Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010a 18.0 73 6.0 65 Gordon 2010b 16.0 74 3.0 68 Lin 2016 0.5 11 0.5 10 Heterogeneity: $l^2 = 0$, $\tau^2 = 0$, $p = 0.46$ Test for effect in subgroup: $z = 3.17$ ($p = 0.002$) Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Test for overall effect: $z = 3.81$ ($p < 0.001$) Chu 2012 16.0 69 10.0 70 Lin 2016 16.0 74 3.0 68 Lin 2016 16.0 74 3.0 68 Lin 2016 289 281 Lin 2016 289 281 Lin 2016 289 281 Lin 2016 289 281 Lin 2016 224 [1.48; 3.38] 100.0% Lin 2010 10.1 0.51 2 10 110 Favors Opioids Favors Placebo	Random effects mode	2759)	1979		2.31	[1.37; 3.87]	77.8%		
Test for effect in subgroup: $z = 3.16 (p = 0.002)$ Follow-up <3 months (max. at 8 weeks) Chu 2012 16.0 69 10.0 70 Gordon 2010a 18.0 73 6.0 65 Gordon 2010b 16.0 74 3.0 68 Kawamata 2019 3.0 62 3.0 68 Lin 2016 0.5 11 0.5 10 Random effects model 289 281 Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0.5477$, $p < 0.01$ Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Test for overall effect: $z = 3.81 (p < 0.001)$ Chu 2012 16.0 69 10.0 70 Lin 2016 2.24 [1.48; 3.38] 100.0% Lin 2016 2.25 [1.48] 1.26 [1.48] 1.26 [1.48] 1.26 [1.48] 1.26 [1.48] 1.26 [1.48] 1.26 [1	Heterogeneity: $I^2 = 77\%$, 1	$t^2 = 0.7251, p < 1$	0.01							
Follow-up <3 months (max. at 8 weeks)	Test for effect in subgroup	p: z = 3.16 (p = 0)	.002)							
Chu 2012 16.0 69 10.0 70 1.62 [0.79; 3.32] 6.5% Gordon 2010a 18.0 73 6.0 65 2.67 [1.13; 6.32] 6.0% Gordon 2010b 16.0 74 3.0 68 4.90 [1.49; 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 1.10 [0.02; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 2.15 [1.34; 3.44] 22.2% Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48; 3.38] 100.0% Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) 5.001 5.001 0.01 0.10 10.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) 0.01 0.51 2 10 110 Favors Opioids Favors Opioids Favors Placebo 5.001 10 10.51<	Follow-up <3 months	max. at 8 wee	(s)							
Gordon 2010a 18.0 73 6.0 65 2.67 [1.13] 6.32] 6.0% Gordon 2010b 16.0 74 3.0 68 4.90 [1.49] 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 1.10 [0.23] 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02] 4.168] 1.0% Random effects model 289 281 2.15 [1.34] 3.44] 22.2% Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48] 3.38] 100.0% Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) 5.12 10 110 110 Favors Opioids Favors Placebo	Chu 2012	16.0 69	10.0	70		1.62	[0.79: 3.32]	6.5%		
Gordon 2010b 16.0 74 3.0 68 4.90 [1.49; 16.08] 4.9% Kawamata 2019 3.0 62 3.0 68 1.10 [0.23; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 215 [1.34; 3.44] 22.2% Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48; 3.38] 100.0% Heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.51 2 10 110 Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Example 0.001 0.01 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) Favors Opioids Favors Placebo 10 10 10	Gordon 2010a	18.0 73	6.0	65	- <u></u> -	2.67	[1.13; 6.32]	6.0%		
Kawamata 2019 3.0 62 3.0 68 1.10 [0.23; 5.23] 3.8% Lin 2016 0.5 11 0.5 10 0.91 [0.02; 41.68] 1.0% Random effects model 289 281 2.15 [1.34; 3.44] 22.2% Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48; 3.38] 100.0% Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.1 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) 2260 2.24 [1.48; 3.38] 100.0%	Gordon 2010b	16.0 74	3.0	68		4.90	[1.49; 16.08]	4.9%		
Lin 2016 0.5 11 0.5 10 0.91 [0.02 ; 41.68] 1.0% Random effects model 289 281 2.15 [1.34 ; 3.44] 22.2% Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0.9000$ Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Test for overall effect: $z = 3.81$ ($p < 0.001$) 0.01 0.1 0.51 2 10 110 Favors Opioids Favors Placebo	Kawamata 2019	3.0 62	2 3.0	68		1.10	[0.23; 5.23]	3.8%		
Random effects model 289 281 2.15 [1.34; 3.44] 22.2% Heterogeneity: $l^2 = 0.6, \tau^2 = 0.9 = 0.46$ 2.15 [1.34; 3.44] 22.2% 2.15 [1.34; 3.44] 22.2% Random effects model 3048 2.24 [1.48; 3.38] 100.0% Heterogeneity: $l^2 = 70\%, \tau^2 = 0.5477, p < 0.01$ 0.01 0.1 0.51 2 10 110 Residual heterogeneity: $l^2 = 72\%, p < 0.01$ Favors Opioids Favors Placebo	Lin 2016	0.5 11	0.5	10		0.91	[0.02; 41.68]	1.0%		
Heterogeneity: $l^2 = 0\%$, $\tau^2 = 0$, $p = 0.46$ Test for effect in subgroup: $z = 3.17$ ($p = 0.002$) Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall effect: $z = 3.81$ ($p < 0.001$) Test for overall ef	Random effects mode	289)	281		2.15	[1.34; 3.44]	22.2%		
Test for effect in subgroup: $z = 3.17$ ($p = 0.002$) Random effects model 3048 Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ 2260 2.24 [1.48; 3.38] 100.0% Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.01 <th <="" colspan="2" td=""><td>Heterogeneity: $I^2 = 0\%$, τ^2</td><td>$^{2} = 0, p = 0.46$</td><td></td><td></td><td></td><td></td><td></td><td></td></th>	<td>Heterogeneity: $I^2 = 0\%$, τ^2</td> <td>$^{2} = 0, p = 0.46$</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>		Heterogeneity: $I^2 = 0\%$, τ^2	$^{2} = 0, p = 0.46$						
Random effects model 3048 2260 2.24 [1.48; 3.38] 100.0% Heterogeneity: $l^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ 0.01 0.1 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$) Favors Opioids Favors Placebo 10 10	Test for effect in subgroup	z = 3.17 (p = 0)	.002)							
Heterogeneity: $I^2 = 70\%$, $\tau^2 = 0.5477$, $p < 0.01$ Residual heterogeneity: $I^2 = 72\%$, $p < 0.01$ Test for overall effect: $z = 3.81$ ($p < 0.001$) Favors Opioids Favors Opioids	Random effects mode	I 3048	3	2260		2.24	[1.48; 3.38]	100.0%		
Residual heterogeneity: $l^2 = 72\%$, $p < 0.01$ 0.01 0.1 0.51 2 10 110 Test for overall effect: $z = 3.81$ ($p < 0.001$)Favors OpioidsFavors Placebo	Heterogeneity: I ² = 70%, 1	$t^2 = 0.5477, p < 0.5477$	0.01	Г						
	Residual heterogeneity: /2 Test for overall effect: z =	r = 72%, p < 0.07 3 81 (p < 0.001)		0.0 E	1 0.1 0.51.2 10 11 avors Opioids Eavors Placebo	0				
Test for subgroup differences: $\chi_{4}^{2} = 0.04$, df = 1 (p = 0.84)	Test for subgroup differen	ces: $\chi_{1}^{2} = 0.04$. c	f = 1 (p =)	0.84)	avois opiolos i avois Flacebo	,				

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

	O	pioids	Pla	icebo						
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight		
Follow-up ≥3 months (r	nax. at 1	5 wee	ks)							
Buynak 2010	9.0	328	33.0	160		0.13	[0.07; 0.27]	6.4%		
Buynak 2010	18.0	318	33.0	159		0.27	[0.16; 0.47]	8.5%		
Christoph 2017	8.0	126	4.0	32		0.51	[0.16; 1.58]	3.3%		
Christoph 2017	5.0	130	4.0	32		0.31	[0.09; 1.08]	2.8%		
Christoph 2017	4.0	128	3.0	31		0.32	[0.08; 1.37]	2.2%		
Christoph 2017	4.0	131	3.0	31		0.32	[0.07; 1.34]	2.2%		
Gimbel 2016	19.0	254	61.0	257	-	0.32	[0.19; 0.51]	9.4%		
Hale 2007	8.0	70	39.0	73		0.21	[0.11; 0.42]	6.6%		
Hale 2010	16.0	134	40.0	134		0.40	[0.24; 0.68]	8.7%		
Katz 2007	12.0	105	35.0	100		0.33	[0.18; 0.59]	7.8%		
Katz 2015	8.0	193	34.0	196		0.24	[0.11; 0.50]	6.0%		
Rauck 2016	8.0	229	23.0	232	- <u>-</u>	0.35	[0.16; 0.77]	5.6%		
Steiner 2011	22.0	256	36.0	283		0.68	[0.41; 1.12]	9.1%		
vonDrackova 2008	5.0	151	6.0	158		0.87	[0.27; 2.80]	3.2%		
Webster 2006	15.0	206	40.0	101		0.18	[0.11; 0.32]	8.5%		
Random effects model		2759		1979		0.31	[0.24; 0.40]	90.3%		
Heterogeneity: $I^2 = 43\%$, τ^2 Test for effect in subgroup:	z = 0.0964 z = -9.16	, p = 0 (p < 0	.04 .001)							
Follow-up <3 months (r	nax. at 8	week	s)							
Gordon 2010b	1.0	74	4.0	68		0.23	[0.03; 2.00]	1.1%		
Kawamata 2019	13.0	62	28.0	68	֥	0.51	[0.29; 0.89]	8.3%		
Lin 2016	0.5	11	0.5	10		0.91	[0.02; 41.68]	0.4%		
Random effects model		147		146	\diamond	0.49	[0.29; 0.84]	9.7%		
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.74$ Test for effect in subgroup: $z = -2.60$ ($p = 0.009$)										
Random effects model		2906		2125		0.33	[0.26; 0.41]	100.0%		
Heterogeneity: $I^2 = 39\%$, τ^2	= 0.0861	, p = 0	.05	ſ						
Residual heterogeneity: 12	= 37%, p	= 0.06		0.0	01 0.1 0.51 2 10 1	00				
Test for overall effect: z = -	9.54 (p <	0.001)		E E	avors Opioids Favors Placeb	0				
Test for subgroup difference	es: $\chi_1^2 = 2$	2.23, df	= 1 (p = 0	0.14)						

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

	O	bioids	Pla	acebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (r	nax. at '	15 wee	eks)					
Buynak 2010	11.0	646	6.0	319		0.91	[0.34; 2.43]	18.9%
Christoph 2017	41.0	511	1.0	126		10.11	[1.40; 72.79]	9.7%
Gimbel 2016	1.0	254	9.0	256		0.11	[0.01; 0.88]	9.2%
Hale 2007	0.5	70	5.0	72	<	0.10	[0.01; 1.85]	5.6%
Hale 2010	3.0	134	7.0	134		0.43	[0.11; 1.62]	15.0%
Katz 2007	1.0	105	2.0	100		0.48	[0.04; 5.17]	7.4%
Katz 2015	3.0	193	1.0	196		3.05	[0.32; 29.03]	8.1%
Rauck 2016	3.0	229	1.0	232		3.04	[0.32; 29.00]	8.1%
vonDrackova 2008	1.0	151	3.0	158		0.35	[0.04; 3.32]	8.1%
Random effects model		2293		1593	$ \rightarrow $	0.79	[0.32; 1.95]	90.0%
Heterogeneity: $I^2 = 51\%$, τ^2 Test for effect in subgroup:	= 0.9160 z = -0.51	p = 0. p = 0.	04 608)					
Follow-up <3 months (I	nax. at a	3 week	(s)					
Chu 2012	0.5	48	0.5	55		1.15	[0.02; 56.65]	3.3%
Gordon 2010b	0.5	78	0.5	78		1.00	[0.02; 49.77]	3.3%
Kawamata 2019	0.5	62	0.5	68		1.10	[0.02; 54.44]	3.3%
Random effects model Heterogeneity: $I^2 = 0\%$, $\tau^2 =$ Test for effect in subgroup:	= 0, p = 1. z = 0.07	188 00 (p = 0.9	47)	201		1.08	[0.11; 10.28]	10.0%
Random effects model Heterogeneity: $l^2 = 33\%$, τ^2 Residual heterogeneity: $l^2 = -1$ Test for overall effect: $z = -1$	= 0.5445 = 39%, p 0.52 (p =	2481 , <i>p</i> = 0. = 0.09 0.603)	12	1794 0.	01 0.1 0.51 2 10 100 Favors Opioids Favors Placebo	0.82	[0.38; 1.75]	100.0%
Test for subgroup difference	$es: \chi_1^2 = 0$.06, df =	= 1 (p = 0	.80)	,			

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

	Op	bioids	Pla	icebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 1	5 wee	ks)					
Buynak 2010	278	328	9 5	160		1.43	[1.25; 1.64]	8.9%
Buynak 2010	240	318	95	159		1.26	1.10: 1.46	8.6%
Christoph 2017	108	130	20	31		1.29	[0.98; 1.69]	4.0%
Christoph 2017	100	126	21	32		1.21	[0.93; 1.58]	4.1%
Christoph 2017	115	128	21	32		1.37	[1.06; 1.77]	4.3%
Christoph 2017	107	127	20	31		1.31	[1.00; 1.71]	4.0%
Gimbel 2016	125	254	120	256		1.05	[0.88; 1.26]	6.8%
Hale 2007	31	70	27	72		1.18	[0.79; 1.76]	2.2%
Hale 2010	73	134	64	134		1.14	[0.90; 1.44]	4.9%
Katz 2007	61	105	44	100		1.32	[1.00; 1.74]	4.0%
Katz 2015	125	193	95	196		1.34	[1.12; 1.60]	6.9%
Rauck 2016	94	229	101	232		0.94	[0.76; 1.17]	5.6%
Steiner 2011	140	256	146	283		1.06	[0.90; 1.24]	7.8%
vonDrackova 2008	80	151	83	158		1.01	[0.82; 1.25]	5.7%
Random effects model		2549		1876	🔶	1.20	[1.11; 1.29]	77.9%
Heterogeneity: $I^2 = 43\%$, τ^2	$^{2} = 0.0081$	p = 0	.04					
rescror encount subgroup.		(p - 0.	001)					
Follow-up <3 months (r	nax. at 8	week	s)					
Gordon 2010a	72	73	58	65		1.11	[1.01; 1.21]	11.7%
Gordon 2010b	65	73	44	68		1.38	[1.13; 1.67]	6.3%
Kawamata 2019	45	62	37	68		1.33	[1.02; 1.74]	4.1%
Random effects model		208		201		1.25	[1.01; 1.54]	22.1%
Heterogeneity: $I^2 = 76\%$, τ^2	= 0.0254	p = 0	.01					
l est for effect in subgroup:	z = 2.06	(p = 0.	039)					
Random effects model		2757		2077		1.20	[1.13: 1.28]	100.0%
Heterogeneity: $I^2 = 43\%$, τ^2	² = 0.0069	p = 0	.03		I	1	. ,	
Residual heterogeneity: /2	= 52%, p	< 0.01		0	.5 1 2	2		
Test for overall effect: z = {	5.69 (p <)	0.001)			Favors Opioids Favors Placebo	C		
Test for subgroup difference	es: $\chi_1^2 = 0$).13, df	= 1 (p = 0	0.72)	-			

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

	O	bioids	Pla	icebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 1	5 wee	ks)					
Buynak 2010	11.0	328	2.0	160		2.68	[0.60; 11.96]	11.2%
Buynak 2010	7.0	318	1.0	159		3.50	[0.43; 28.20]	5.7%
Christoph 2017	2.0	128	0.5	32		1.00	[0.05; 21.64]	2.6%
Christoph 2017	3.0	126	0.5	32		1.52	[0.08; 29.66]	2.8%
Christoph 2017	3.0	130	0.5	31		1.43	[0.07; 27.84]	2.8%
Christoph 2017	4.0	127	0.5	31		1.95	[0.11; 35.98]	2.9%
Gimbel 2016	4.0	254	4.0	256	· · · · · · · · · · · · · · · · ·	1.01	[0.25; 3.99]	13.2%
Hale 2007	2.0	70	0.5	72		4.11	[0.19; 89.65]	2.6%
Hale 2010	6.0	134	4.0	134	·	1.50	[0.43; 5.20]	16.2%
Katz 2007	2.0	105	3.0	100	(0.63	[0.11; 3.72]	8.0%
Katz 2015	2.0	193	2.0	196		1.02	[0.14; 7.14]	6.6%
Rauck 2016	3.0	229	1.0	232		3.04	[0.32; 29.00]	4.9%
Steiner 2011	3.0	256	2.0	283		1.66	[0.28; 9.84]	7.9%
vonDrackova 2008	0.5	151	1.0	158		0.52	[0.02; 15.48]	2.2%
Webster 2006	0.5	206	0.5	101		0.49	[0.01; 24.53]	1.6%
Random effects model		2755		1977	÷ '	1.49	[0.88; 2.51]	91.4%
Heterogeneity: $I^2 = 0\%$, τ^2	= 0, p = 0	.99						
Test for effect in subgroup	z = 1.48	(p = 0.	138)					
Follow-up <3 months (max. at 8	week	s)					
Chu 2012	1.0	69	0.5	70		2.03	[0.07; 59.50]	2.2%
Gordon 2010a	0.5	73	1.0	65).45	[0.02; 13.05]	2.2%
Gordon 2010b	0.5	73	0.5	68		0.93	[0.02; 46.29]	1.6%
Kawamata 2019	2.0	62	0.5	68		4.39	[0.20; 95.44]	2.6%
Random effects model		277		271		1.51	[0.28; 8.25]	8.6%
Heterogeneity: $I^2 = 0\%$, τ^2	= 0, p = 0	.79						
Test for effect in subgroup	z = 0.47	(p = 0.	635)					
Random effects model		3032		2248		1.49	[0.90; 2.45]	100.0%
Heterogeneity: $I^2 = 0\%$, τ^2	= 0, p = 1	.00					- / /	
Residual heterogeneity: 12	= 0%, p =	1.00		0.	01 0.1 1 10 100	1		
Test for overall effect: z =	1.56 (p =	0.120)			Favors Opioids Favors Placebo			
Test for subgroup difference	:es: χ ₁ ² = ().00, df	= 1 (p =)	0.99)				

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

	0	bioids	Pla	icebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (r	nax. at 1	5 wee	eks)					
Buynak 2010	177	646	29	319		3.01	[2.08; 4.36]	11.9%
Christoph 2017	146	511	8	126	4	1.50	[2.27; 8.92]	8.6%
Gimbel 2016	19	254	19	256	1	.01	[0.55; 1.86]	9.3%
Hale 2007	2	70	1	72		2.06	[0.19; 22.18]	1.6%
Hale 2010	12	134	10	134		.20	[0.54; 2.68]	7.4%
Katz 2007	7	105	1	100	++ e	6.67	[0.84; 53.22]	2.0%
Katz 2015	21	193	9	196	2	2.37	[1.11; 5.04]	7.9%
Rauck 2016	23	229	17	232	1	.37	[0.75; 2.50]	9.4%
Steiner 2011	32	256	31	283	1	.14	[0.72; 1.82]	10.9%
vonDrackova 2008	12	151	11	158	- * 1	.14	[0.52; 2.51]	7.6%
Random effects model		2549		1876	🔶 🔰 1	.80	[1.21; 2.65]	76.5%
Heterogeneity: $I^2 = 67\%$, τ^2 Test for effect in subgroup:	² = 0.2321 z = 2.93	, p < 0 (p = 0.	.01 003)					
Follow-up <3 months (r	nax. at 8	week	s)					
Chu 2012	4	69	໌ 5	70	— <u> </u>).81	[0.23; 2.90]	4.3%
Gordon 2010a	28	73	11	65	2	2.27	[1.23; 4.18]	9.3%
Gordon 2010b	39	73	12	68	+ 3	3.03	[1.74; 5.28]	9.9%
Random effects model		215		203	i 🗢 2	2.20	[1.26; 3.84]	23.5%
Heterogeneity: $I^2 = 43\%$, τ^2 Test for effect in subgroup:	² = 0.1034 z = 2.76	p = 0 p = 0.	.17 006)					
Random effects model		2764		2079		.86	[1.35; 2.56]	100.0%
Heterogeneity: $I^2 = 62\%$, τ^2	² = 0.1876	i, p < 0	.01	1	1 1 1 1 1 1			
Residual heterogeneity: 12	= 65%, <i>p</i>	< 0.01		0.0	01 0.1 0.51 2 10 100			
Test for overall effect: $z = 3$	3.82 (p <	0.001)		I	Favors Opioids Favors Placebo			
Test for subgroup difference	es: χ ₁ = ().34, df	= 1 (p =	0.56)				

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

	Op	bioids	Pla	acebo					
Study	Events	Total	Events	Total	Risk Ratio	RR	95	%-CI	Weight
Follow-up ≥3 months (r	nax. at 1	5 wee	eks)						
Buynak 2010	92.0	646	5	319		9.09	[3.73; 2	2.13]	13.3%
Christoph 2017	84.0	511	5	126		4.14	[1.72; 1	0.00]	13.5%
Gimbel 2016	14.0	254	6	256	+	2.35	[0.92;	6.02]	12.6%
Hale 2007	0.5	70	1	72		0.51	[0.02; 1	5.09]	1.6%
Hale 2010	8.0	134	6	134		1.33	[0.48;	3.74]	11.3%
Katz 2007	8.0	105	1	100		7.62	[0.97; 5	9.82]	4.0%
Rauck 2016	9.0	229	1	232		9.12	[1.16; 7	(1.39]	4.0%
Steiner 2011	11.0	256	5	283		2.43	[0.86;	6.91]	11.1%
vonDrackova 2008	7.0	151	5	158		1.46	[0.48;	4.52]	10.1%
Random effects model		2356		1680	\diamond	3.09	[1.78;	5.37]	81.6%
Heterogeneity: $I^2 = 46\%$, τ^2 Test for effect in subgroup:	z = 0.3018 z = 4.01	8, p = 0 (p < 0.	.06 001)						
Follow-up <3 months (r	nax. at 8	week	s)						
Gordon 2010a	11.0	73	3	65		3.26	[0.95; 1	1.19]	9.0%
Gordon 2010b	16.0	73	3	68		4.97	[1.51; 1	6.30]	9.4%
Random effects model		146		133		4.06	[1.73;	9.54]	18.4%
Heterogeneity: $I^2 = 0\%$, τ^2 : Test for effect in subgroup:	= 0, p = 0 z = 3.21	.63 (p = 0.	001)						
Random effects model		2502		1813	·	3.26	[2.08;	5.09]	100.0%
Heterogeneity: $I^2 = 34\%$, τ^2	= 0.1834	p = 0	.13			_			
Residual heterogeneity: 12	= 40%, p	= 0.09		0.	01 0.1 0.51 2 10 10	U			
Test for overall effect: z = t	0.18 (p <) 2	0.001)			Favors Opioids Favors Placebo				
Test for subgroup difference	es: χ ₁ = (J.27, df	= 1 (p =	0.60)					

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

	0	bioids	Pla	cebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 1	5 wee	ks)					
Buynak 2010	132	646	¹⁶	319		4.07	[2.47; 6.72]	20.5%
Christoph 2017	84	511	5	126		4.14	[1.72; 10.00]	10.0%
Gimbel 2016	7	254	2	256		3.53	[0.74; 16.82]	3.8%
Hale 2007	4	70	1	72		4.11	0.47; 35.91	2.1%
Hale 2010	10	134	5	134		2.00	[0.70; 5.70]	7.7%
Katz 2007	7	105	1	100		6.67	[0.84; 53.22]	2.3%
Katz 2015	10	193	1	196		10.16	1.31; 78.57	2.3%
Rauck 2016	9	229	6	232		1.52	[0.55; 4.20]	8.0%
Steiner 2011	9	256	3	283		3.32	[0.91; 12.12]	5.3%
vonDrackova 2008	18	151	8	158		2.35	[1.06; 5.25]	11.5%
Random effects model		2549		1876	🔶	3.27	[2.40; 4.46]	73.6%
Heterogeneity: $I^2 = 0\%$, τ^2 Test for effect in subgroup:	= 0, p = 0 z = 7.49	.69 (p < 0.	001)					
Follow-up <3 months (r	nax. at 8	week	s)					
Chu 2012	5	69	1	70		5.07	[0.61; 42.31]	2.2%
Gordon 2010a	20	73	14	65		1.27	[0.70; 2.31]	17.0%
Gordon 2010b	12	73	4	68		2.79	[0.95; 8.25]	7.2%
Random effects model		215		203		1.86	[0.91; 3.79]	26.4%
Heterogeneity: $I^2 = 29\%$, τ^2 Test for effect in subgroup:	f = 0.1294 z = 1.71	p = 0 p = 0.0	.24 088)					
Random effects model	2 0.005	2764		2079		2.73	[1.98; 3.77]	100.0%
Heterogeneity: $I = 20\%$, τ	= 0.0654	p = 0	.24	0.				
Test for overall effect: 7 = 6	– 0%, p = 3.14 (n <	0.09		0.0	UT U.T U.DTZ TU TUU			
Test for subgroup difference	$e^{2} = 2$	2.001) 2.04 df	= 1 (n = 1)	0 15)	Favors Opiolos Favors Placebo			
. set of ourgroup affording			· () - (

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

	O	bioids	Pla	acebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (I	nax. at '	15 wee	eks)					
Hale 2007	0.5	70	0.5	72		1.03	[0.02; 51.12]	0.6%
Gimbel 2016	2.0	254	2.0	256		1.01	[0.14; 7.10]	2.2%
vonDrackova 2008	9.0	151	6.0	158	- <u>+=</u> -	1.57	[0.57; 4.30]	8.4%
Katz 2007	5.0	105	3.0	100		1.59	[0.39; 6.47]	4.3%
Buynak 2010	94.0	646	18.0	319		2.58	[1.59; 4.19]	36.1%
Steiner 2011	10.0	256	3.0	283		3.68	[1.03; 13.24]	5.2%
Christoph 2017	174.0	511	11.0	126]= -	3.90	[2.19; 6.95]	25.6%
Rauck 2016	4.0	229	1.0	232		4.05	[0.46; 35.98]	1.8%
Random effects model		2222		1546	🔶	2.72	[1.98; 3.73]	84.1%
Heterogeneity: $I^2 = 0\%$, $\tau^2 =$	= 0, p = 0.	.67						
Test for effect in subgroup:	z = 6.15 ((p < 0.0	001)					
Follow-up <3 months (I	max. at 8	8 week	(s)					
Gordon 2010a	16.0	73	5.0	65	- <u>+</u>	2.85	[1.11; 7.34]	9.5%
Gordon 2010b	24.0	73	3.0	68		7.45	[2.35; 23.63]	6.4%
Random effects model		146		133		4.35	[1.68; 11.25]	15.9%
Heterogeneity: $I^2 = 39\%$, τ^2 Test for effect in subgroup:	= 0.1860 z = 3.04 (p = 0. p = 0.0	20)02)					
Random effects model		2368		1679	\$	2.91	[2.17; 3.90]	100.0%
Heterogeneity: $I^2 = 0\%$, $\tau^2 =$	= 0, p = 0.	57		Г				
Residual heterogeneity: I ² =	= 0%, p =	0.59		0.0	1 0.1 0.51 2 10	100		
Test for overall effect: $z = 7$.17 (p < 0).001)		F	avors Opioids Favors Place	bo		
Test for subgroup difference	$es: \chi_1^2 = 0$.85, df =	= 1 (p = 0	.36)				

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



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

	O	pioids	Pla	icebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 1	5 wee	ks)					
Buynak 2010	91	646	44.0	319		1.02	[0.73; 1.43]	46.0%
Christoph 2017	50	511	11.0	126		1.12	[0.60; 2.09]	13.2%
Gimbel 2016	6	254	8.0	256		0.76	[0.27; 2.15]	4.7%
Hale 2007	2	70	0.5	72		4.11	[0.19; 89.65]	0.5%
Hale 2010	7	134	10.0	134		0.70	[0.27; 1.78]	5.9%
Katz 2007	4	105	2.0	100	-	1.90	[0.36; 10.17]	1.8%
Rauck 2016	5	229	8.0	232		0.63	[0.21; 1.91]	4.2%
Steiner 2011	14	256	14.0	283		1.11	[0.54; 2.27]	9.9%
vonDrackova 2008	6	151	11.0	158		0.57	[0.22; 1.50]	5.5%
Random effects model		2356		1680	\$	0.97	[0.76; 1.23]	91.7%
Heterogeneity: $I^2 = 0\%$, τ^2 Test for effect in subgroup	= 0, p = 0 : z = -0.27	0.83 7 (p = 0	.787)					
Follow-up <3 months (max. at 8	week	s)					
Gordon 2010a	8	73	6.0	65		1.19	[0.43; 3.24]	5.1%
Gordon 2010b	9	73	3.0	68		2.79	[0.79; 9.89]	3.2%
Random effects mode		146		133		1.67	[0.73; 3.80]	8.3%
Heterogeneity: $I^2 = 8\%$, τ^2	= 0.0299,	p = 0.3	30					
Test for effect in subgroup	: z = 1.22	(p = 0.)	224)					
Random effects model		2502		1813		1.01	[0.81; 1.27]	100.0%
Heterogeneity: $I^2 = 0\%$, τ^2	= 0, p = 0	.72		Г		1		
Residual heterogeneity: 12	= 0%, p =	0.80		0.0	1 0.1 0.51 2 10	100		
Test for overall effect: z =	0.10 (p =	0.919)		F	avors Opioids Favors Pla	icebo		
Test for subgroup different	$ces: \chi_1^2 = 1$	1.54, df	= 1 (p = 0	0.21)				

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

Study	Total	Oj Mean	pioids SD	Total	Pla Mean	acebo SD	Standardised Mean Difference	SMD	95%-CI	Weight
Follow-up ≥3 months (r Katz 2015 Steiner 2011 Random effects model Heterogeneity: $I^2 = 92\%$, τ^2 Test for effect in subgroup:	nax. a 193 166 359 = 0.13 z = -0.	12 we -2.55 1.41 19, p < 12 (p =	eks) 10.42 9.22 0.01 0.907)	196 195 391	0.67 -0.85	11.17 9.77	-#-	-0.30 0.24 -0.03	[-0.50; -0.10] [0.03; 0.44] [-0.55; 0.49]	35.5% 35.2% 70.7%
Follow-up <3 months (a Kawamata 2019 Random effects model Heterogeneity: not applicab Test for effect in subgroup:	1t 5 we 62 62 le z = 0.3	eks) 2.85	14.09 0.756)	68 68	2.08	13.94		0.05 0.05	[-0.29; 0.40] [-0.29; 0.40]	29.3% 29.3%
Random effects model Heterogeneity: $I^2 = 85\%$, τ^2 Residual heterogeneity: I^2 Test for overall effect: $z = -$ Test for subgroup difference	421 = 0.08 = 92%, 0.03 (p es: χ ² ₁	55, p < p < 0.0 = 0.973 = 0.07, o	0.01 1 3) df = 1 (/	459	3)	-	1 -0.5 0 0.5 Favors Opioids Favors Placeb	1 0	[-0.37; 0.36]	100.0%

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

	Op	oioids	Pla	icebo				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 1	5 wee	ks)					
Buynak 2010	257	446	680	245	+	1.76	[1.45; 2.15]	22.6%
Christoph 2017	134	260	36	100		1.43	[1.07; 1.91]	13.5%
Gimbel 2016	96	242	49	248		2.01	[1.49; 2.70]	12.9%
Katz 2015	129	193	91	196		1.44	[1.20; 1.72]	25.0%
Steiner 2011	145	237	110	261		1.45	[1.22; 1.73]	25.9%
Random effects model		1378		1050	\	1.58	[1.40; 1.78]	100.0%
Heterogeneity: $I^2 = 36\%$, τ	² = 0.0068	p = 0	.18	I		I		
Test for overall effect: z =	7.37 (p < (0.001)		0.2	25 0.5 1 2	5		
				Fa	vors Placebo Favors Opioio	ls		

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

Study	Op Events	ioids Total E	Pla Events	cebo Total	Risk Ratio	RR	95%-CI	Weight
Follow-up ≥3 months (max. at 12	2 week	s)					
Hale 2007	55	69	22	69		2.50	[1.74; 3.60]	29.7%
Hale 2010	107	133	83	133		1.29	[1.10; 1.51]	36.7%
Katz 2007	84	103	38	90		1.93	[1.49; 2.50]	33.6%
Random effects model Heterogeneity: $I^2 = 88\%$, τ	l ² = 0.1120,	305 p < 0.0	1	292		1.80	[1.19; 2.70]	100.0%
Test for overall effect: z =	2.81 (p = 0	.005)		0.25 Favo	0.5 1 2 rs Placebo Favors Opioio	5 Is		

Figure S19. Patient assessed treatment effectiveness moderately or highly effective: treatment duration

at 8 weeks



Forest plots CNCP

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

Events	Total	Events	Total	Risk	Ratio R	R 95%-CI
35 le z = -1.95	106 106 (<i>p</i> = 0.0	53	115 115	+ ()	0.7 0.7	2 [0.51; 1.00] 2 [0.51; 1.00]
47 le z = -1.18	116 116 (<i>p</i> = 0.2	56 236)	116 116	+ ()	- 0.8 > 0.8	4 [0.63; 1.12] 4 [0.63; 1.12]
40 le z = -2.73	108 108 (<i>p</i> = 0.0	60	107 107	+	0.6 0.6	6 [0.49; 0.89] 6 [0.49; 0.89]
48 le z = -2.07	117 117 (<i>p</i> = 0.0	63	115 115 [0.2	25 0.5 1	0.7 0.7 1 2 4	5 [0.57; 0.98] 5 [0.57; 0.98]
	Events 35 $ e _{z} = -1.95$ 47 $ e _{z} = -1.18$ 40 $ e _{z} = -2.73$ 48 $ e _{z} = -2.07$	Events Total 35 106 106 106 107 108 108 108 108 108 108 108 108	Events Total Events $35 \ 106 \ 53 \ 106 \ z = -1.95 \ (p = 0.051)$ $47 \ 116 \ 116 \ 116 \ 116 \ 116 \ 108 \ 1$	Events Total Events Total Events Total 35 106 53 115 106 115 115 le 2 -1.95 ($p = 0.051$) 116 47 116 56 116 116 116 116 116 le z -1.18 ($p = 0.236$) 107 40 108 60 107 le z -2.73 ($p = 0.006$) 115 48 117 63 115 le z -2.07 ($p = 0.038$) 107	Events Total Events Total Risk 35 106 53 115 106 115 115 le $z = -1.95 (p = 0.051)$ 47 116 56 116 le $z = -1.95 (p = 0.051)$ 47 116 116 116 le $z = -1.18 (p = 0.236)$ 40 108 60 107 107 le $z = -2.73 (p = 0.006)$ 48 117 63 115 115 le $z = -2.07 (p = 0.038)$ 0.25 0.5 1	Events Total Events Total Risk Ratio R $35 \ 106 \ 53 \ 115 \ 106 \ 115 \ 106 \ 115 \ 106 \ 115 \ 116 $

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

Study	Op Events	ioids Total	Non-Op Events	ioids Total	Ri	sk Ratio	,	RR	95%-CI
High disability–modera Elsesser 2017 Random effects model Heterogeneity: not applicate Test for effect in subgroup:	tely limit 20 z = 0.12 (p	ting 137 137 0 = 0.9	23	163 163	-		-	1.03 1.03	[0.59; 1.80] [0.59; 1.80]
High disability–severel Elsesser 2017 Random effects model Heterogeneity: not applicat Test for effect in subgroup:	by limiting 111	137 137 137	127	163 163		+		1.04 1.04	[0.93; 1.17] [0.93; 1.17]
				0.25 Fa	0.5 vors Opioi	1 ds Favo	2 Drs Non-(⊣ 4 Opioids	

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

Study	Opi Events 1	ioids Total	Non-Op Events	oioids Total	Risk Ratio	RR	95%-CI	Weight
Adams 2006 Krebs 2018	208 4 11	4278 119	218 8	8589 119		1.92 1.38	[1.59; 2.31] [0.57; 3.30]	95.7% 4.3%
Random effects model Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0\%$ Test for overall effect: $z = 6$	= 0, p = 0.4 6.83 (p < 0.0	4397 7 001)		8708 [0.2 F	5 0.5 1 2 avors Opioids Favors Non-O	1.89 1 4 pioids	[1.57; 2.27]	100.0%

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

	Op	bioids	Non-Op	bioids				
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI	Weight
Krebs 2018	55	118	56	119		0.99	[0.76; 1.3]	100.0%
Random effects model Heterogeneity: $I^2 = NA\%$, τ	² = NA, p	118 = NA		119		0.99	[0.76; 1.3]	100.0%
Test for overall effect: z = -	0.07 (p =	0.94)		0.	.5 1	2		
					Favors Opioids Favors Non-	Opioids		

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

Study	O Events	pioids Total Ev	No /ents	Opioids Total		Risk I	Ratio	RR	95%-CI
Low dose, chronic opio Edlund 2014 Random effects model Heterogeneity: not applicabl Test for effect in subgroup: 2	e z = 17.73	6902 6902 (<i>p</i> < 0.001)	150	371371 371371				17.94 17.94	[13.03; 24.68] [13.03; 24.68]
Medium dose, chronic of Edlund 2014 Random effects model Heterogeneity: not applicabl Test for effect in subgroup: 2	e z = 20.81	se 3654 3654 (p < 0.001	150	371371 371371			\$	31.85 31.85	[22.99; 44.12] [22.99; 44.12]
High dose, chronic opid Edlund 2014 Random effects model Heterogeneity: not applicabl Test for effect in subgroup: 2	e z = 23.01	378 378 (p < 0.001)	150	371371 371371	r	1		➡ 150.64	[98.28; 230.92] [98.28; 230.92]
				Fa	0.01 (avours C).1 1 Opioids	10 1 Favours No	00 Opioids	

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

	Op	bioids	Non-Op	bioids			
Study	Events	Total	Events	Total	Risk Ratio	RR	95%-CI Weight
Elsesser 2017	111	170	73	165		1.48	[1.2; 1.81] 100.0%
Random effects model		170		165		1.48	[1.2; 1.81] 100.0%
Heterogeneity: $I^2 = NA\%$, τ	$^2 = NA, p$	= NA				7	
Test for overall effect: z = 3	3.75 (p <)	0.001)		0	.5 1	2	
					Favors Opioids Favors Non-O	pioids	

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

	0	pioids	Non-O	pioids					
Study	Events	Total	Events	Total		Risk Rat	io	RR	95%-CI
Time since drug started	d 31-180	days							
Ray 2016	70	12194	40	11752				1.69	[1.14; 2.49]
Random effects model		12194		11752		<	\bigcirc	1.69	[1.14; 2.49]
Heterogeneity: not applicab	le								
Test for effect in subgroup:	z = 2.64 (p = 0.00)8)						
Time since drug started	d >180 d	ays							
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 applicab	le								
Test for effect in subgroup:	z = 0.97 (p = 0.33	31)						
				0.	25 0.5	1	2	4	
				F	avours Or	pioids Fa	vours N	on-Opioid	S
								P	

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 judgemen t
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/ MODERAT E
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/impa	ict)	
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)	
Parallel (n=4)	3.65 (2.50, 5.31)	0.16
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)	
Parallel (n=3)	2.79 (1.83, 4.26)	0.56
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)	
Parallel (n=2)	4.75 (2.79, 8.08)	0.05
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)	
Parallel (n=3)	0.99 (0.75, 1.31)	0.44
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)	
Parallel (n=6)	1.27 (0.99, 1.63)	<0.0001
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)	
Parallel (n=6)	3.82 (1.87, 7.80)	0.0088
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)	
Parallel (n=5)	0.90 (0.34, 2.39)	0.21
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/impac	t)	
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-naive (n=3)	1.08 (0.38, 3.07)		
Opioid-experienced (n=4)	2.83 (1.20, 6.65)	0.03	
Opioid-naive and-experienced (n=3)	4.78 (2.98, 7.65)		
	4.04 (0.04.4.07)		
All comparisons (n=11)	1.01 (0.81, 1.27)		
	1.02 (0.58, 1.79)		
Opioid-experienced (n=5)	0.92 (0.50, 1.69)	0.93	
Opioid-naive 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)	
Opioid-experienced (n=5)	0.66 (0.44, 0.98)	0.03
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)	
Opioid-experienced (n=5)	0.34 (0.24, 0.48)	0.21
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)							
Opioid-experienced (n=1)	2.01 (1.49, 2.70)	0.18						
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 effectivene	ess							
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	
№ of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Opioids	Placebo	Relative (95% CI)	Absolute (95% Cl)	Certainty
Pain intensity (follow up: range	4 weeks to 15 wee	ks; assessed with: s	elf-reported NRS [(0–10]; lower is bette	er; 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)	$\bigoplus_{LOW} \bigcirc$
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)	$\bigoplus_{LOW} \bigcirc$
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)	$\oplus \bigoplus_{\text{Low}} \bigcirc$
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Sleep quality (f	ollow up: range 8	3 weeks to 14 weel	ks; assessed with: se	elf-reported VAS [0-	–100]; higher is bet	tter; 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)	
Sleep quality: p	bain interference/	impact on sleep (fo	ollow up: range 5 we	eks to 8 weeks; as	sessed with: self-re	ported NRS [0-10]; lower is be	etter; the MID = 1-poir	nt)		• • • •	
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)	
Trial discontinu	ations (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)	
Trial discontinu	ations (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)	
Trial discontinu	ations (efficacy l	ack) (follow up: rar	nge 4 weeks to 15 we	eeks)							
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Opioid withdray	wal symptoms (fo	bllow 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)	
Opioid depende	ency (follow up:	5 weeks)									
1	randomised trials	serious ^a	not serious	serious ⁱ	not serious	none	Kawamata et al. re by the Data and Sa 68).	ported that "no patient fety Monitoring Board"	s were judged to have 'in either the opioid (e developed drug dependency n = 62) or placebo group (n =	$\bigoplus_{LOW} \bigcirc \bigcirc$
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 a (total n = 607).	$\bigoplus_{LOW} \bigcirc \bigcirc$			
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Adverse events	(serious) (follow	v up: range 4 week	s 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)	
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 intervention arm.	addressing mortality	reported any treatme	nt-related deaths in either	
Nausea (follow	up: range 4 wee	ks 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)	
Vomitting (follow	w up: range 4 we	eeks 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)	
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Dizziness (follo	w up: range 8 w	eeks 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)	$\bigoplus_{\rm LOW} \bigcirc$
Somnolence (fo	ollow up: range 8	weeks to 15 week	s)								
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)	$\bigoplus_{\rm LOW}$
Headache (follo	w up: range 8 w	eeks 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)	
Depression and	Anxiety (follow	up: range 5 weeks	to 12 weeks; assess	sed with: self-report	ed SF-36v2 MCS [[0-100]; higher is better; surroo	gate 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)	
Suicidal ideatio	n or behavior (fo	llow up: range 12 v	veeks to 14 weeks)								
2	randomised trials	serious ^a	not serious	serious ^b	serious ¹	none	Christoph et al. repo group (n = 511). Ste placebo group (n = 2 one in the interventi	orted that no events oc iner et al. reported tha 283) compared to non- on group (n = 134) and	ccurred in either the p at only 1 event of suic e in the opioids group d none in the placebo	lacebo (n = 126) or opioid idal ideation occurred in the p (n = 256). Hale et al. reported p group (n = 134).	
PGIC: much im	proved or very n	nuch improved (folle	ow 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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
PGR study med	lication: good, ve	ery good, or excelle	ent (follow up: 12 we	eks)							
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)	

Patient assessed treatment effect	tiveness: moderately o	r highly effective	(follow up: 8 weeks)
		J J	

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)	
---	----------------------	----------------------	-------------	-------------	----------------------	------	----------------	----------------	-------------------------------	--	--

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: I² >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: I² >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.

I. 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

Nº of studies	Study design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Opioids	Non-Opioids	Relative (95% Cl)	Absolute (95% Cl)	
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)	$\oplus \bigoplus_{\text{LOW}} \bigcirc$
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)	
30% reduction	in BPI pain severi	ty 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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
30% reduction	in BPI interference	e 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)	
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Drug abuse (fo	llow up: 12 month	s)									
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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
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)	
Pain Severity a	nd Disability (ther	apy duration ≥6 mo	onths)								
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)	
Opioid Abuse o	or Dependence (fo	llow up: 18 months	5)								
1	observational studies	very serious ^f	not serious	not serious	serious ⁹	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)	
Any adverse ev	vents (therapy dur	ation ≥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)	$\bigoplus_{LOW} \bigcirc \bigcirc$
Deaths (time si	nce 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)	

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. Optimal information size criterium met (87 per group; α = 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 \leq 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 l^2 statistic was used to quantify the proportion of variation in point estimates due to differences across studies. When heterogeneity was large (e.g. $l^2 > 75\%$) 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 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



Risk Ratio