Supplementary Material*

Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Systemic pharmacologic therapies for low back pain: a systematic review for an American College of Physicians clinical practice guideline. Ann Intern Med. doi:10.7326/M16-2458

* This supplementary material was provided by the authors to give readers further details on their article. The material was reviewed but not copyedited.

Supplement Table 1. Quality assessments of randomized controlled trials not included in a systematic review

Author, Year	Was randomization adequate?	Was treatment allocation concealed?	Were treatment groups similar at baseline?	Were patients blinded?	Were care providers blinded?	Were outcome assessors/data analysts blinded?	Were cointerventions avoided or similar among groups?
Baron, 2010 (108)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baron, 2014 (109)	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes
Brotz, 2010 (88)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cloutier, 2013 (50)	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Eskin, 2014 (120)	Yes	Unclear	Yes	Yes	Yes	Yes	Yes
Farajirad, 2013 (100)	Unclear	Unclear	Yes	Unclear	No	No	Unclear
Friedman, 2008 (121)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Friedman, 2015 (54)	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear
Goldberg, 2015 (125)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hedeboe, 1982 (122)	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Yes
Herrmann, 2009 (29)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes
Holve, 2008 (123)	No (sequential allocation)	No	Unclear	Yes	Yes	Yes	Yes
Hyup Lee, 2013 (51)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes
Kalita, 2014 (110)	Yes	Unclear	Yes	No	No	No	Yes
Katz, 2011 (32)	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes
Kivitz, 2013 (33)	Unclear	Unclear	Yes	Yes	Yes	Unclear	Yes
Majchrzycki, 2014 (30)	Yes	No	Yes	No	No	Unclear	Unclear
Markman, 2014 (111)	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Markman, 2015 (55)	Yes	Yes	Unclear	Yes	Yes	Unclear	Unclear
Mazza, 2010 (99)	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Pareek, 2009 (80)	Unclear	Unclear	Yes	Yes	Unclear	Yes	Unclear
Pota, 2012 (2012)	Unclear	No	Yes	Yes	Unclear	Unclear	Unclear
Ralph, 2008 (81)	Unclear	Unclear	Yes	Yes	Unclear	Yes	Yes
Rauck, 2014 (52)	Unclear	Unclear	No; not sex	Yes	Yes	Unclear	Yes
Rauck, 2016 (57)	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Yes
Rodrigues, 2014 (124)	Yes	Yes	Yes	Yes	Unclear	Yes	Yes
Romano, 2009 (113)	Unclear	Unclear	Unclear	Yes	Yes	Unclear	Yes
Sakai, 2015 (115)	Unclear	Unclear	Yes	No	No	Yes	Unclear
Schiphorst Preuper, 2014 (53)	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear
Schukro, 2016 (101)	Yes	Unclear	Unclear	Yes	Yes	Yes	Yes
Shirado, 2010 (31)	Yes	No	Yes	No	No	Yes	Yes
Skljarevski, 2009 (96)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Skljarevski, 2010 (97)	Unclear	Unclear	Unclear	Yes	Unclear	Yes	Yes
Skljarevski, 2010 (98)	Unclear	Unclear	Unclear	Yes	Unclear	Yes	Yes
Wen, 2015 (56)	Unclear	Unclear	Yes	Yes	Yes	Unclear	Unclear
Williams, 2014 (20)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yaksi, 2007 (114)	Unclear	Unclear	Yes	Unclear	Unclear	Unclear	Yes

Author, Year	Was compliance acceptable in all groups?	Was attrition reported?	Was attrition level an acceptable level?	Was the timing of outcome assessment similar for all groups?	Was there an intention-to-treat analysis?	Is there a registered or published protocol?	Was there avoidance of selective outcome reporting?	Quality Rating
Baron, 2010 (108)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Fair
Baron, 2014 (109)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Fair
Brotz, 2010 (88)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Cloutier, 2013 (50)	Unclear	Yes	No; <20%	Yes	Yes	Unclear	Unclear	Good
Eskin, 2014 (120)	Yes	Yes	Yes	Yes	No	Unclear	Yes	Fair
Farajirad, 2013 (100)	Unclear	No	Unclear	Unclear	Unclear	Unclear	Unclear	Poor
Friedman, 2008 (121)	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Good
Friedman, 2015 (54)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair
Goldberg, 2015 (125)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Hedeboe, 1982 (122)	Unclear	No	Unclear	Yes	Yes	Unclear	Unclear	Fair
Herrmann, 2009 (29)	Yes	Yes	Yes	Yes	Yes	No	Yes	Fair
Holve, 2008 (123)	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Poor
Hyup Lee, 2013 (51)	Yes	Yes	No; 21%	Yes	Yes	Yes	Yes	Good
Kalita, 2014 (110)	Unclear	Yes	No	Yes	Yes	Yes	Yes	Poor
Katz, 2011 (32)	Unclear	Yes	No; 32%	Yes	Yes	Yes	Yes	Fair
Kivitz, 2013 (33)	Unclear	Yes	No; 37%	Yes	Yes	Yes	Yes	Fair
Majchrzycki, 2014 (30)	Yes	Yes	Yes	Yes	Yes	No	Yes	Fair
Markman, 2014 (111)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair
Markman, 2015 (55)	Unclear	Yes	Yes	Yes	Yes	Yes	Unclear	Fair
Mazza, 2010 (99)	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Fair
Pareek, 2009 (80)	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Fair
Pota, 2012 (2012)	Unclear	Yes	Yes	Yes	Yes	No	Yes	Fair
Ralph, 2008 (81)	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Fair
Rauck, 2014 (52)	Yes	Yes	No; 39%	Yes	Yes	No	Yes	Poor
Rauck, 2016 (57)	Unclear	Yes	Yes	Yes	Yes	Yes	Yes	Fair
Rodrigues, 2014 (124)	Unclear	Yes	Yes	Yes	Yes	Unclear	Yes	Fair
Romano, 2009 (113)	Unclear	Yes	Yes	Yes	No	Unclear	Yes	Fair
Sakai, 2015 (115)	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Poor
Schiphorst Preuper, 2014 (53)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Fair
Schukro, 2016 (101)	Unclear	Yes	No	Yes	No (partial)	Yes	Yes	Poor
Shirado, 2010 (31)	Yes	Yes	Yes	Yes	Yes	No	Yes	Good
Skljarevski, 2009 (96)	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Good
Skljarevski, 2010 (97)	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Fair
Skljarevski, 2010 (98)	Yes	Yes	Yes	Yes	No	Unclear	Unclear	Fair
Wen, 2015 (56)	Unclear	Yes	No; 25%	Yes	Yes	Unclear	Unclear	Fair
Williams, 2014 (20)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Good
Yaksi, 2007 (114)	Unclear	No	Unclear	Yes	Unclear	Unclear	Yes	Poor

Supplement Table 2. Quality assessment of systematic reviews

Author, Year	'A priori' design provided?	Duplicate study selection and data extraction? a. Study selection b. Data extraction	Comprehensive literature search performed?	Status of publication used as an inclusion criteria?	List of studies (included and excluded) provided?	Characteristics of the included studies provided?
Chaparro, 2013 (46)	Yes	Yes to both	Yes	Yes	No	Yes
Roelofs, 2008 (21)	Yes	a. Yes b. Yes	Yes	Unclear	Yes	Yes
Urquhart, 2010 (95)	Yes	a. Yes b. No	Yes	Unclear	Yes	Yes
Van Tulder, 2003 (79)	Yes	a. Yes b. Yes	Yes	Unclear	Yes	Yes

Author, Year	Scientific quality of included studies assessed and documented?	Scientific quality of the included studies used appropriately in formulating conclusions?	Methods used to synthesize the findings of studies appropriate?	Likelihood of publication bias assessed?	Conflict of interest stated? a) Systematic Review b) Individual Studies	Quality Rating
Chaparro, 2013 (46)	Yes	Yes	Yes	Yes	Yes	Good
Roelofs, 2008 (21)	Yes	Yes	Yes	Yes	a. Yes b. No	Good
Urquhart, 2010 (95)	Yes	Yes	Yes	Yes	a. Yes b. No	Good
Van Tulder, 2009 (79)	Yes	Yes	Yes	Yes	a. Yes b. No	Good

Supplement Table 3. Summary of systematic reviews of pharmacologic treatments for low back pain

Treatment	Author, year	Number and Type of Studies	Interventions and Number of Patients	Conclusions
Acetaminophen	Roelofs, 2008 (21)	65 RCT and controlled clinical trials Acute low back pain (25 trials), chronic low back pain (9 trials) mixed or unclear low back pain population (31 trials) 6 trials NSAIDs versus paracetamol or acetaminophen	A. NSAIDs (nonselective and selective) B. Other medications C. Other active interventions (i.e., passive physical modalities) D. Placebo Total n=11,237	For acute LBP, NSAIDs were no different for improvement in pain intensity vs. paracetamol/acetaminophen (3 studies; SMD -0.21, 95% CI -0.43 to 0.02) One study found limited evidence that paracetamol was less effective than NSAIDs for chronic low back pain. Other comparisons of NSAIDs are discussed in the NSAIDs or opioids section. NSAIDs were associated with more side effects than paracetamol (4 trials, RR 1.76, 95% CI 1.12 to 2.76)
Antidepressants	Urquhart, 2010 (95)	10 RCTs; 9 trials conducted in pts with chronic low back pain; 1 trial duration of low back pain not reported. Duration of followup 10 days to 12 weeks.	A. Antidepressants (n=315): paroxetine (3 studies); desipramine (3 studies); imipramine (2 studies); maprotiline (2 studies); fluoxetine (2 studies); bupropion, trazodone, amitriptyline, nortriptyline and clomipramine IV (1 study each) B. Placebo (n=252)	There were no significant differences between antidepressants and placebo for pain relief (6 trials; SMD -0.04, 95% CI -0.25 to 0.17) or depression (2 trials; SMD 0.06 (95% CI -0.29 to 0.40) in patients with chronic low back pain.
NSAIDs	Roelofs, 2008 (21)	65 RCTs and controlled clinical trials Acute low back pain (25 trials), chronic low back pain (9 trials) mixed or unclear low back pain population (31 trials)	A. NSAIDs (nonselective and selective) B. Other medications C. Other active interventions (i.e., passive physical modalities) D. Placebo Total n=11,237	For acute LBP, NSAIDs associated with greater improvement in pain intensity vs. placebo (4 studies; WMD -8.39, 95% CI -12.68 to -4.10), but no clear effects on pain relief. For chronic LBP, NSAIDs associated with greater improvement in pain vs. placebo (4 trials, WMD -12.40, 95% CI -15.53 to -9.26). For radicular LBP, there was no difference in pain intensity between NSAIDs versus placebo. Studies of NSAIDs vs. acetaminophen or opioids are discussed in those sections. NSAIDs were associated with more side effects than placebo (10 trials, RR 1.35, 95% CI 1.09 to 1.68) COX-2-selective NSAIDs were associated with lower risk of side effects versus nonselective NSAIDs (4 trials; RR 0.83, 95% CI 0.70 to 0.99). Serious harms were rare.

Treatment	Author, year	Number and Type of Studies	Interventions and Number of Patients	Conclusions
Opioids	Carson, 2011 (71)	41 RCTs: 10 comparing long-acting with another long-acting opioid; 3 were for low back pain. 27 trials comparing long-acting opioid to placebo (for indirect comparisons); 4 for back pain 7 trials comparing long-acting vs. short-acting opioids; 5 for back pain	Comparisons of long-acting opioids: total 1310 patients in trials for LBP 4 trials for low back pain comparing long-acting opioid to placebo are all summarized elsewhere Comparisons of long vs. short acting opioids: 284 total patients in trials for LBP	Insufficient evidence from 10 head-to-head trials to suggest that a long-acting opioid is superior to another in terms of efficacy in adult patients with chronic noncancer pain. No useful indirect evidence for determining the comparative efficacy of long-acting opioids was found in 27 placebo-controlled trials In 7 fair-quality trials directly comparing a long-acting opioid to a short-acting opioid there was no good quality evidence to suggest superior efficacy of long-acting opioids as a class over short-acting opioids. Insufficient evidence from 10 head-to-head trials of long acting opioids that any drug safer than others. No trials adequately assessed addiction or abuse. There was insufficient evidence from 27 placebo-controlled trials to suggest that a long-acting opioid was superior in terms of adverse events to any other. No convincing evidence from 7 RCTs to suggest lower adverse event rates with long-acting opioids as a class compared with short-acting opioids for all assessed adverse events. No data compared rates of addiction or abuse of long-acting and short-acting opioids.

Treatment	Author, year	Number and Type of Studies	Interventions and Number of Patients	Conclusions
	Chaparro, 2013 (46)	A. Strong opioids vs. placebo: 7 trials B. Tramadol vs. placebo: 5 trials C. Buprenorphine vs. placebo: 2 trials D. Opioids vs. NSAIDs: 2 trials in 1 article all subacute or chronic low back pain Duration of followup 4 weeks to 13 weeks	A. Strong opioids, n=1154, placebo, n=733 B. Tramadol, n=689, placebo, n=689 C. Buprenorphine, n=312, placebo, n=341 D. Opioids n=785 celecoxib, n=798	A. Pain: moderate-quality evidence that strong opioids are better than placebo; SMD 0.43 lower (95% CI 0.52 to 0.33); Function: Moderate-quality evidence better than placebo in improving function (SMD 0.26 lower disability score [95% CI 0.37 to 0.15]) B. Pain: low-quality evidence tramadol is better than placebo, SMD 0.55 lower, 95% CI 0.66 to 0.44; Function: Moderate evidence tramadol is better than placebo, SMD 0.18 lower (95% CI 0.29 to 0.07) C. Pain: very low-quality evidence that transdermal buprenorphine is better than placebo (MD 0.58 lower, 95% CI 0.61 to 0.55; Function: very low-quality evidence of no difference in function (MD 3 lower (95% CI 11.44 lower to 5.44 higher) D. Pain: very low-quality evidence that tramadol is better than celecoxib; Note: this seems to be a misprint; in fact, celecoxib appeared to be better than tramadol (at least 30% pain reduction: 63.7% with celecoxib; 52.5% with tramadol, OR 0.63 [95% CI 0.52, 0.77])
Skeletal muscle relaxants and benzodiazepines	Van Tulder, 2009 (79)	A. Skeletal muscle relaxants vs. placebo: 11 trials B. Antispasticity medications vs. placebo: 2 trials C. Benzodiazepines vs. placebo: 4 trials D. Muscle relaxants vs. muscle relaxants: 8 trials E. Muscle relaxants + analgesics vs. placebo + analgesics: 6 trials	A. Skeletal muscle relaxants, n=527, placebo, n=421 B. Antispasticity medications, n=110, placebo, n=110 C. Benzodiazepines, n=173, placebo, n=167 D. Muscle relaxants, n=615 E. Muscle relaxants + analgesics, n=332, placebo + analgesics, n=324	A. Pain relief: 2-4 days, 4 trials, RR 0.80 (95% CI 0.71 to 0.89) and 5-7 days, 3 trials, RR 0.58 (95% CI 0.45 to 0.76); Global Efficacy: 2-4 days, 4 trials, 0.49 (95% CI 0.25 to 0.95) and 5-7days, 4 trials, RR 0.68 (95% CI 0.41 to 1.13); acute low back pain B. Pain relief: Two high quality trials found effectiveness at 4 days; acute low back pain C. Pain relief: One low quality trial found benzodiazepine more effective than placebo at 5 days for acute low back pain; 5-7 days, 2 trials, RR 0.82 (95% CI 0.72 to 0.94) and 10-14 days, 2 trials, RR 0.53, 95% CI 0.42 to 0.97) for chronic low back pain D. No clear differences between muscle relaxants E. Pain relief and decrease of muscle spasm: 3 high quality trials found tizanidine plus analgesic more effective than placebo plus analagesic at 3-4 days and 7-8 days, acute low back pain

COX-2= cyclooxygenase-2, CI=confidence interval, LBP=low back pain, NSAIDS=nonsteroidal anti-inflammatory drug, RCT=randomized controlled trial, RR=relative risk, SMD=standard mean difference. WMD=weighted mean difference

Supplement Table 4. Characteristics and conclusions of acetaminophen trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Williams, 2014 (20) 12 weeks Acute *Good*	A: Acetaminophen: 665 mg 2 tablets orally every 6-8 hours (6 tabs/day) + placebo 1-2 tabs orally every 4-6 hours as needed (up to 8 tabs/day) (n=550) B: Acetaminophen: Placebo 2 tablets orally every 6-8 hours (6 tabs/day) + 500 mg 1-2 tablets orally every 4-6 hours as needed (up to 8 tablets/day) (n=546) C: Placebo: Placebo 2 tablets orally every 6-8 hours (6 tablets/day) + placebo 1-2 tablets orally every 4-6 hours as needed (up to 8 tablets/day) (n=547) Medications taken until recovery or for 4 weeks	A. vs. B. vs. C. Mean age: 44 vs. 45 vs. 45 years Female: 48% vs. 47% vs. 45% Baseline pain (mean, 0-10 NRS): 6.3 vs. 6.3 vs. 6.2 Baseline RDQ (mean, 0-24): 3.5 vs. 3.6 vs. 3.7 Pain below knee: 20% vs. 21% vs. 18	A. vs. B. vs. C. Pain (mean, 0-10): 3.7 vs. 3.8 vs. 3.6 at week 1, 2.6 vs. 2.6 vs. 2.5 at week 2, 1.7 vs. 1.8 vs. 1.7 at week 4, 1.2 vs. 1.3 vs. 1.3 at w 12 RDQ (mean, 0-24): 7.7 vs. 8.0 vs. 8.3 at week 1, 5.2 vs. 5.4 vs. 5.3 at week 2, 3.2 vs. 3.5 vs. 3.3 at week 4, 2.4 vs. 2.6 vs. 2.4 at week 12 Patient Specific Functional Scale (mean, 0-10): 6.2 vs. 6.1 vs. 6.2 at week 1, 7.3 vs. 7.2 vs. 7.4 at week 2, 8.2 vs. 8.1 vs. 8.2 at week 4, 8.7 vs. 8.7 vs. 8.7 at week 12 Global change (mean, -5 to +5): 2.1 vs. 2.0 vs. 2.1 at week 1, 2.8 vs. 2.7 vs. 2.8 at week 2, 3.4 vs. 3.4 vs. 3.5 at week 4, 3.8 vs. 3.7 vs. 3.8 at week 12 SF12 Physical score (mean, 0-100): 50 vs. 50 vs. 51 at week 4, 55 vs. 55 vs. 55 at week 12 SF12 Mental score (mean, 0-100): 44 vs. 44 vs. 44 at week 4, 46 vs. 46 vs. 45 at week 12	A. vs. B. vs. C. Sleep quality "fairly bad" or "very bad": 28% (143/514) vs. 26% (129/501) vs. 26% (127/496) at week 1, 17% (85/508) vs. 18% (88/495) vs. 17% (85/497) at week 2, 12% (59/507) vs. 11% (57/500) vs. 10% (52/503) at week 4, 11% (54/506) vs. 11% (55/503) vs. 8.6% (44/514) at week 12 No differences in use of concomitant medications or health services or hours absent from work Days to recovery (median, days): 17 vs. 17 vs. 16 Satisfied with treatment: 76% (365/478) vs. 72% (342/472) vs. 73% (335/458)

LBP=low back pain, NRS=numeric rating scale, RDQ=Roland-Morris Disability Questionnaire

Supplement Table 5. Characteristics and conclusions of NSAID trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Herrmann, 2009 (29) 5 days Acute *Fair*	A: Lornoxicam 8mg tablets, with 16 mg loading dose on day 1, then 8mg after 8 hours; 8 mg twice per day on days 2-4; 8 mg on day 5 B: Diclofenac: 50 mg twice per day on days 1 and 5; 50mg three times per day on days 2-4. C: Placebo capsules in lornoxicam or diclofenac blister packs Day 5 treatment was optional	A. vs. B. vs. C. Mean age: 51.8 vs. 48.9 vs. 48.4 Female: 44% vs. 47% vs. 42% Pain etiology: Sciatica or lumbo-sciatica	A. vs. B. vs. C. Pain intensity difference, mm: 3 hours: -21.0 vs. -18.7 vs. -15.3, p≤0.05 for A. vs. C. 4 hours: -22.0 vs. -21.5 vs. -14.8, p≤0.05 for A. vs. C. 6 hours: -20.5 vs. -22.4 vs. -14.9, p≤0.05 for A. vs. C. 8 hours: -22.0 vs. -24.1 vs. -13.7, p≤0.05 for A. vs. C. Sum of time-weighted pain intensity difference, mm x minute: 0-4 hours: -4020 vs. -3879 vs. -2901, p≤0.05 for A. vs. C. 0-6 hours: -6486 vs. -6358 vs. -4713, p≤0.05 for A. vs. C. 0-8 hours: -9125 vs. -8833 vs. -6257, p≤0.05 for A. vs. C. Pain Relief (mm): 3 hours: 30.1 vs. 30.8 vs. 26.6 4 hours: 31.7 vs. 33.9 vs. 26.6 6 hours: 31.1 vs. 34.3 vs. 26.1 8 hours: 31.9 vs. 35.6 vs. 23.9, p≤0.05 for A. vs. C. Peak pain intensity difference, A. vs. C: -27.9 mm vs. -19.9 mm, p=0.01 Time to peak pain intensity difference, A. vs. C: 243 vs. 240 minutes, no difference Peak pain relief, A. vs. C. : 38.0 mm vs. 31.1 mm, p=0.05 Time to peak pain relief: no difference Start of peak pain relief: no difference End of peak pain relief: no difference Duration of peak pain relief: no difference	

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Katz, 2011 (32) 12 weeks Chronic *Fair*	A. Naproxen 1000 mg/day + placebo (n=88) B. Placebo (n=41)	A vs. B Mean age: 52.1 vs. 52.2 Female: 47.7% vs. 56.1% BMI, mean: 28.6 vs. 28.6 Duration of LBP, mean years: 13.0 vs. 9.7 RDQ, mean: 12.4 vs. 13.7	A vs. B at 6 weeks Change in Average LBPI: -2.54 vs. -1.96; p=0.68 ≥30% reduction in LBPI: 56.8% vs. 31.7%, p= 0.006 ≥50% reduction in LBPI: 34.1% vs. 19.5%, p= 0.067 Change from baseline RDQ: -2.43 vs. -2.41; p=0.482	
Kivitz, 2013 (33) 16 weeks Chronic *Fair*	A. Naproxen 1000 mg/day (n=295) B. Placebo (n=230)	A vs. B Mean age: 52.6 vs. 51.2 Female: 51.5% vs. 54.3% BMI, mean: 30.3% vs. 29.1% Duration of LBP, mean years: 11.2 vs. 11.3 LBPI: 6.77 vs. 6.71 RDQ: 12.86 vs. 12.79	A vs. B change from baseline at week 16: LBPI: -1.66 vs. -1.25, p=0.405 RDQ: -2.07 vs. -1.75, p=0.037 Global assessment of pain: -0.50 vs. -0.40, p=0.405 ≥30% reduction in LBPI: 37.6% vs. 27.0%, p=0.009 ≥50% reduction in LBPI: 26.4% vs. 17.0%, p=0.009 ≥70% reduction in LBPI: 12.5% vs. 9.6%, p=0.278 ≥90% reduction in LBPI: 5.4% vs 3.5%, p=0.286	
Majchrzycki, 2014 (30) 2 weeks Acute, subacute *Fair*	A. Deep tissue massage + NSAID (n=26) B. Deep tissue massage (n=28)	A. vs. B. Mean age: 50.8 vs. 52.6 Female: 50.0% vs. 46.4% Chronic pain: 100% Baseline pain: not reported Baseline function: not reported QOL: not reported	A. vs. B. VAS1 (0-100): pain intensity during resting: 16.5 vs. 13.9 VAS2 (0-100): pain intensity during motion: 3.2 vs. 3.4 VAS3 (0-100): pain intensity during mobility of the aching area of the spine: 4.8 vs. 8.2	A. vs. B. Difference scores, no significantly different results between groups on: RDQ: 21.2 vs. 16.1 ODI: 24.7 vs. 19.6

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Shirado, 2010 (31) 12 months Subacute *Good*	A: NSAIDs: loxoprofen sodium, 60 mg tablet 3 times daily; diclofenac sodium, 25 mg tablet 3 times daily; or zaltoprofen, 80 mg tablet 3 times daily B: Exercise: medical professionals at each clinic gave instruction of the exercise. 2 types of exercise: trunk strengthening and stretching. 2 sets of 10 repetitions of each exercise per day were encouraged.	A. vs. B. Mean age: 42.5 vs. 42.0 Female: 59% vs. 52% Pain type: All chronic pain Baseline pain: VAS (0-10): 3.8 vs. 3.5 QOL scores: RDQ (0-24): 3.7 vs. 3.0 JLEQ score (0-120): 21.8 vs. 20.5	A. vs. B. Baseline to 8 week change ratio: Pain: VAS (0-10): -0.35 vs. -0.44, p=0.332	A. vs. B. Baseline to 8 week change ratio: Function: Finger-floor distance: 0.00 vs. -0.09, p=0.112 RDQ: -0.47 vs. -0.72, p=0.023 JLEQ: -0.44 vs. -0.58, p=0.021

BMI=body mass index, JLEQ=Japan Low Back Pain Evaluation Questionnaire, LBPI=low back pain intensity, NSAIDS=nonsteroidal anti-inflammatory drug, ODI=Oswestry Disability Index, RDQ=Roland Morris Disability Questionnaire, VAS=visual analog scale, QOL=quality of life

Supplement Table 6. Characteristics and conclusions of opioid trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Cloutier, 2013 (50) 4 weeks Subacute, chronic *Good*	A: Oxycodone/Naloxone, both controlled release, titrated dose of 10mg/5mg q 12h up to 40mg/20mg q 12 hour B: Placebo Crossover design: 4 weeks of each intervention	Due to crossover design, all patients received both A and B. Among the 54 analyzed: Mean age: 50.6 years Female: 50% Baseline score on Pain and Disability Index: 42 on a 0-70 scale (70 worst) Among the full 83 enrolled: Mean age: 51.3 years Female: 53%	A vs. B ITT Analysis (n=83): Pain VAS (0-100): A. 52.2 mm (SD 23.0; B: 57.8 mm (SD 24.2) (p=0.053) Ordinal pain score: A: 2.3 (SD 0.8); B: 2.5 (SD 0.9), (p=0.086) No other results for ITT analysis Per protocol analysis: Pain VAS (0-100): A. 48.6 mm (SD 23.1); B: 55.9 mm (SD 25.4) (p=0.03) Ordinal pain score: A: 2.1 (SD 0.8); B: 2.4 (SD 0.9), (p=0.042)	A vs. B Pain Disability Index: 34 vs. 38, p=0.05 (per protocol analysis) SF-36 General Health: "no difference" Quebec Back Pain Disability: "no difference"
Friedman, 2015 (54) 3 months Acute *Fair*	All arms received Naproxen, 500 mg every 12 hours, plus: A: Oxycodone, 5mg; Acetaminophen, 325 mg 1 or 2 tablets every 8 hours (n=108) B. Cyclobenzaprine, 5mg 1 or 2 tablets every 8 hours (n=108) C. Placebo (n=107)	A vs. B vs. C Mean age: 39 vs. 38 vs. 39 Female sex: 60 vs. 45 vs. 53 Race: Not reported Mean RDQ score at end of ED discharge: 18.9 vs. 18.4 vs. 18.7	Not reported	A vs. B vs. C Mean improvement on Roland Morris Disability Questionnaire at 1-week: 11.1 vs. 10.1 vs. 9.8, p=0.28 for A vs. C, p=0.77 for B vs. C, p=0.45 for A vs. B Any adverse events: 43/108 vs. 36/10/8 vs. 22/107 Drowsiness: 16/108 vs. 7/108 vs. 4/107 Dizziness: 16/108 vs. 3/108 vs. 3/107 Stomach irritation: 7/108 vs. 7/108 vs. 5/107 Nausea or vomiting: 19/108 vs. 4/108 vs. 6/107

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Hyup Lee, 2013 (51) 29 days Subacute, chronic *Good*	A. Extended-release tramadol HCl 75 mg/acetaminophen 650 mg fixed-combination tablet (n=125) Max dose=4 tabs/d=300 mg tramadol B. Placebo (n=120)	A vs. B Mean age: 59.9 vs. 60.4 years Female sex: 75% vs. 74%	A vs. B Pain intensity change ≥30%, full analysis set: 57.7% (49/85) vs. 41.1% (37/90); p=0.037 Pain intensity change ≥30%, per protocol: 63% (46/73) vs. 44.9% (35/78); p=0.027 Pain intensity change ≥50%, full analysis set: 31.8% vs. 20.0%; p=0.075 Pain intensity change ≥50%, per protocol: 34.3% vs. 21.8%; p=0.088	A vs. B Korean SF-36: patients in the intervention group had significant improvements in role-physical, general health, and reported health transition domains, and a tendency (p=0.052) toward improvement in vitality Korean ODI: patients in the intervention group had significant functional improvement in the personal care section (p=0.045) and a tendency (p=0.053) toward improvement in total ODI scores
Markman, 2015 (55) 3 days Chronic *Fair*	A: Oxymorphone hydrochloride, 5mg (n=8) B. Propoxyphene/acetaminophen, 100mg/650mg (n=8) C. Placebo (n=8) All participants received single doses of the drugs at 3 separate visits in a random order after a washout period of at least 3 days.	Overall population: Age, mean: 71.8 years Male: 12/24 (50%) Race: Caucasian: 23/24 (96%) Duration of symptoms: >12months: 23/24 (96%) BMI, mean: 31.52	A vs. C Difference in median time to first moderate pain symptom on treadmill (≥4/10 on NRS): -0.25, 98.3% CI -6.54 to 5.00) Pain at rest (NRS): 1.59 vs. 1.63, Treatment effect -0.04, 98.3%CI -0.72 to 0.65 Final pain rating (NRS): 5.87 vs. 5.67, Treatment effect 0.20, 98.3% CI -0.74 to 1.14) Modified BPI-SF, interference score: 3.87 vs. 4.06, Treatment effect -0.19, 98.3% -1.03 to 0.65 Modified BPI-SF, pain intensity score: 4.28 vs. 4.45, Treatment effect -0.17, 98.3% -0.92 to 0.58 Swiss Spinal Stenosis Questionnaire, symptom severity: 3.03 vs. 3.06, Treatment effect -0.03, 98.3% CI -0.19 to 0.13 Patient Global Assessment of Pain: 2.47 vs. 2.76, Treatment effect -0.03, 98.3% -0.52 to 0.47	A vs. C Swiss Spinal Stenosis Questionnaire, physical function: 2.41 vs. 2.45, Treatment effect -0.04, 98.3% CI -0.16 to 0.09 RDQ: 13.01 vs. 13.19, Treatment effect -0.18, 98.3% -1.37 to 1.02 ODI: 37.36 vs. 37.34, Treatment effect 0.02, 98.3% CI -3.46 to 3.51

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Rauck, 2014 (52) 12 weeks Chronic *Poor*	A. Extended-release hydrocodone in 10, 20, 30, 40, and 50 mg capsules (n=151) Mean dose=119 mg/d Max dose=200 mg/d B. Placebo (n=151)	A vs. B Mean age: 50.4 vs. 50.8 years Female sex: 62% vs. 49%; p=0.028 Mean pain score before titration (NRS 0-10): 6.9 vs. 6.9 Mean pain score after titration (NRS 0-10): 3.1 vs. 3.1	A vs. B Change from baseline in mean daily pain intensity score: 0.48 vs. 0.96; p=0.008	
Rauck, 2016 (57) 12 weeks Chronic *Fair*	A: Buccal buprenorphine 150-450 µg bid based on open-label titrated dose (n=229) B: Placebo (n=232)	A vs. B Mean age: 51 vs. 49 years Female sex: 54% vs. 59% Mean pain score before titration (NRS 0-10): 7.2 vs. 7.3 Mean pain score at randomization: 2.8 vs. 2.8	A vs. B Pain, NRS (0-10), mean increase from baseline: 0.94 vs. 1.59, difference -0.67 (95% CI -1.07 to -0.26) Pain improved ≥30%: 63% (132/209) vs. 47% (99/211); p=0.001 Pain improved ≥50%: 41% (86/209) vs. 33% (70/211)	A vs. B Roland Morris Disability Questionnaire (0-24, mean change from baseline to follow-up: 0.6 vs. 1.2, difference -0.75 (95% CI -1.77 to 0.27) Medical Outcomes Score Sleep Subscale: No differences, data not reported Patient Global Impression of Change (0 to 7), mean change from baseline to follow-up: 4.5 vs. 3.9, difference 0.6 (95% CI 0.2 to 1.0)
Schiphorst Preuper, 2014 (53) 2 weeks Chronic *Fair*	A. Tramadol 37.5 mg/acetaminophen 325 mg fixed-combination capsule (n=25) Max dose tramadol=225 mg/d B. Placebo (n=25)	A vs. B Mean age: 42 vs. 44 years Female sex: 72% vs. 64% Mean duration of pain: 18 vs. 24 months Mean pain score (VAS 0-10): 6.1 vs. 4.7	A vs. B VAS (0-10) current pain, baseline-followup: 6.1-5.1 vs. 4.7-4.5; change -1 vs. -0.2 VAS (0-10), maximum pain, baseline-followup: 7.3-7.4 vs. 7.1-7.7; change 0.1 vs. 0.6 VAS (0-10), minimum pain, baseline-followup: 4.4-3.8 vs. 2.0-2.6; change -0.6 vs. 0.6 Pain relief: 42% (10/24) vs. 4% (1/25); RR 10.42 (95% CI 1.44 to 75.29) Same pain or worsened: 58% (14/24) vs. 96% (24/25); RR 0.61 (95% CI 0.43 to 0.86)	A vs. B Lifting (kg), baseline-followup: 18-19 vs. 20-17 kg; change 1 vs. -3 kg Carrying (kg), baseline-followup: 24-20 vs. 24-21 kg; change -4 vs. -3 Static bending (s), baseline-followup: 119-143 vs. 158-192.5; change 24 vs. 34.5 s Dynamic bending (s/rep), baseline-followup: 2.7-2.8 vs. 2.7-3.0; change 0.1 vs. 0.3 Roland Morris Disability Questionnaire (0-24), baseline-followup: 13.0-11.5 vs. 13.0-13.0; change -1.5 vs. 0

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Wen, 2015 (56) 12 weeks Chronic *Fair*	A: Hydrocodone, once daily, dose determined in open-label run-in phase (mean 57 mg) (n=296) B: Placebo (n=292)	A vs. B Age, mean: 49.2 vs. 47.9 Male: 124/296 (42%) vs. 126/292 (43%) Race: White: 195/296 (66%) vs. 207/292 (71%); Black: 67/296 (23%) vs. 21/292 (17%)	A vs. B Average pain over the last 24 hours, assessed weekly (least squares mean, 0-10): 3.7 vs. 4.23, mean difference -0.53, p=0.0016 Reduction in pain intensity >=30%: 65% vs. 53%, p=0.0033 Reduction in pain intensity >=50%: 48% vs. 39%, p=0.02	A vs. B Sleep disturbance: No statistically significant difference ODI, BPI-SF, SF-36: No statistically significant differences Supplemental medication use: 22% vs. 17%, p=0.17 Withdrawal due to treatment emergent adverse effects: A vs. B: 4% vs. 3% Any treatment emergent adverse event: 136/296 (46%) vs. 103/292 (35%) Nausea: 24/296 (8%) vs. 16/292 (5%) Constipation: 10/296 (3%) vs. 7/292 (2%) Vomiting: 18/296 (6%) vs. 9/292 (3%) Dizziness: 9/296 (3%) vs. 5/292 (2%) Headache: 6/292 (2%) vs. 5/292 (2%) Somnolence: 3/296 (1%) vs. 2/292 (1%)

BMI=body mass index, BPI=Brief Pain Inventory, CI=confidence interval, ED=emergency department, ITT=intention to treat, LBP=low back pain, NRS=numeric rating scale, ODI= Oswestry Disability Index, RDQ=Roland Morris Disability Questionnaire, RR=relative risk, SD=standard deviation, VAS=visual analogue scale

Supplement Table 7. Characteristics and conclusions of skeletal muscle relaxant trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Friedman, 2015 (54) 3 months Acute *Fair*	All arms received Naproxen, 500 mg every 12 hours, plus: A: Oxycodone, 5mg; Acetaminophen, 325 mg 1 or 2 tablets every 8 hours (n=108) B. Cyclobenzaprine, 5mg 1 or 2 tablets every 8 hours (n=108) C. Placebo (n=107)	A vs. B vs. C Mean age: 39 vs. 38 vs. 39 Female sex: 60 vs. 45 vs. 53 Race: Not reported Mean RDQ score at end of ED discharge: 18.9 vs. 18.4 vs. 18.7	Not reported	A vs. B vs. C Mean improvement on RDQ at 1-week: 11.1 vs. 10.1 vs. 9.8, p=0.28 for A vs. C, p=0.77 for B vs. C, p=0.45 for A vs. B Any adverse events: 43/108 vs. 36/10/8 vs. 22/107 Drowsiness: 16/108 vs. 7/108 vs. 4/107 Dizziness: 16/108 vs. 3/108 vs. 3/107 Stomach irritation: 7/108 vs. 7/108 vs. 5/107 Nausea or vomiting: 19/108 vs. 4/108 vs. 6/107
Pareek, 2009 (80) 7 days Acute *Fair*	A. Tizanidine 2 mg + aceclofenac 100 mg twice daily for 7 days (n=101) B. Aceclofenac 100 mg twice daily for 7 days (n=96)	A. vs. B. Mean age: 62 vs. 58 years Female:39% vs. 40% Baseline pain, function not reported	A. vs. B. Pain at rest, mean change from baseline day 3: -3.01 vs. -1.90, p=0.0001; day 7 -5.88 vs. -4.35, p=0.0001 Pain with movement, mean change from baseline day 3: -2.94 vs. -1.81, p=0.0001; day 7 -6.09 vs. -3.98, p=0.0001	A. vs. B. Global improvement, proportion of patients reporting good or excellent response: 75% (71/94) vs. 34% (31/94); RR 1.28 (95% CI 1.07 to 1.52)
Ralph, 2008 (81) 7 days Acute *Fair*	A. Carisoprodol 250 mg three times daily for 7 days (n=277) B. Placebo three times daily for 7 days (n=285	A. vs. B. Mean age: 39 vs. 42 years Female:49% vs. 55% Baseline pain severity: mild 0.4% vs. 0.4%; moderate 74% vs. 74%; severe 25% vs. 26% Baseline RDQ 10 vs. 10	A. vs. B. Pain, patient-rated impression of pain relief, mean change from baseline day 3 (scale 0-4; higher score = greater pain relief): 1.8 vs. 1.1, p<0.0001; day 7 between-group difference p<0.0001 (data not shown)	A. vs. B. Global improvement, patient-rated impression of change, mean change from baseline at day 3 (scale 0-4; higher score = greater improvement); 2.3 vs. 1.7, p<0.0001; day 7 between-group difference p<0.0001 (data not shown)

ED=emergency department, RDQ=Roland Morris Disability Questionnaire, RR=relative risk

Supplement Table 8. Characteristics and conclusions of benzodiazepine trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Brotz, 2010 (88) 1 year LBP duration not specified *Good*	A: Diazepam: 5 mg po twice daily x 5 d, then tapered (tapering regimen not specified) (n=30) B: Placebo (n=30)	A. vs. B. Mean age: 43 vs. 42 years Female: 37% vs. 50% Baseline pain (median, 0-10 VAS): 8 vs. 8 Baseline RDQ (median, 0-24): 14 vs. 14	A. vs. B. Pain improved ≥50%: 41% (12/29) vs. 79% (23/29) at 1 w, RR 0.5 (95% CI 0.3 to 0.8);	A. vs. B. Duration of inability to work (median, days): 26 vs. 15 (p=0.73) RDQ (median improvement, 0-24): 3.0 vs. 5.0 at 1 week (p=0.67) RDQ (median, 0-24): 2 vs. 1 at 1 year Diclofenac consumption (median, mg): 750 vs. 750 at 1 week (p=0.78) Sensory loss improved: 83% (15/18) vs. 86% (19/22) at 1 week, RR 1.0 (95% 0.7 to 1.3) Sensory loss: 43% (9/21) vs. 44% (10/23) at 1 year Reduction of paresis: 22% (6/27) vs. 28% (8/28) at 1 week, RR 0.8 (95% CI 0.3 to 2.0) Paresis: 14% (3/21) vs. 13% (3/23) at 1 year Inability to work beyond day 28: 55% (16/29) vs. 41% (12/29) at 1 week, RR 1.3 (95% CI 0.7 to 2.2) Request for additional analgesics: 51% (15/29) vs. 41% (12/29) at 1 week, RR 1.3 (95% CI 0.7 to 2.3) Underwent surgery: 7 vs. 6 at 6 weeks, 8 vs. 7 at 1 year

CI=confidence interval, LBP=low back pain, RDQ= Roland-Morris Disability Questionnaire, RR=relative risk, VAS=visual analogue scale

Supplement Table 9. Characteristics and conclusions of antidepressant trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Farajirad, 2013 (100) 8 weeks Chronic *Poor*	A. Amitriptyline 25 mg daily titrated to 150 mg daily (maximum) by week 2 (n= not reported) B. Sustained-release bupropion 150 mg daily titrated to 300 mg daily by week 2 (n= not reported)	A. vs. B. Mean age 37 vs. 34 years No other demographic or clinical characteristics reported	A. vs. B. No data shown Pain: No significant difference between groups	Not reported
Mazza, 2010 (99) 13 weeks Chronic *Fair*	A. Escitalopram 20 mg daily (n=41) B. Duloxetine 60 mg daily (n=44)	A. vs. B. Mean age 52 vs. 54 years 56% vs. 57% female Race not reported Pain, mean VAS (scale 0-10) 6.3 vs. 6.4 Function, mean CGI-S score (scale 0-10) 3.6 vs. 3.5	A. vs. B. Pain, VAS (0-10) mean change from baseline: -2.3 vs. -2.45; p=0.74	A. vs. B. Function, CGI-S mean change from baseline: -0.92 vs. -0.69; p=0.21 Quality of life, mean change SF-36 subscales: no significant difference between groups for any subscale
Schukro, 2016 (101) 4 weeks Chronic *Poor*	A: Duloxetine 30 mg/day titrated to 60 mg/day in week 1 and 60 mg/day titrated to 120 mg/day in week 2, maintained on 120 mg/day weeks 3 and 4 B: Placebo Crossover design with 2 week washout, 4 weeks initial treatment, 2 week washout, 4 weeks crossover (n=25 for intention-to-treat population)	Baseline characteristics reported overall Mean age: 58 years 51% female Race not reported Pain, mean VAS (scale 0-10) 6.8 SF-36 Physical Component Summary 28	A vs. B Pain improved >50%: 40% (10/25) vs. 8.0% (2/25); p=0.04 Pain, VAS (0-10), mean in week 4: 3.7 vs. 5.7; p<0.05 (per-protocol analysis, n=21) painDETECT (0-38), mean at 4 weeks: 18 vs. 21, p=0.002	A vs. B SF-36 Mental Component Summary. mean at 4 weeks: 50 vs. 46; p=0.02 SF-36 Physical Component Summary, mean at 4 weeks: 36 vs. 31; p=0.01 Tramadol rescue medication use: 20% (5/25) vs. 28% (7/25); p>0.05

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Skljarevski, 2009 (96) 13 weeks Chronic *Good*	A. Duloxetine 20 mg daily (n=59) B. Duloxetine 60 mg daily (n=116) C. Duloxetine 120 mg daily (n=112) D. Placebo (n=117)	A. vs. B. vs. C. vs. D. Mean age 53 vs. 53 vs. 55 vs. 54 years 61% vs. 58% vs. 58% vs. 55% female Race: 78% vs. 78% vs. 82% vs. 80% white; 22% vs. 22% vs. 18% vs. 20% other Pain, mean BPI 6.4 vs. 6.2 vs. 6.1 vs. 6.2 Function, mean CGI-S score 4.1 vs. 3.5 vs. 3.6 vs. 3.7	A. vs. B. vs. C. vs. D. Pain, mean change from baseline: -1.77 vs. -2.46 vs. -2.40 vs. -2.10; no significant differences among groups Pain, BPI-S mean change from baseline: -1.79 vs. -2.50 vs. -2.45 vs. -1.87; B vs. D: p<0.05	A. vs. B. vs. C. vs. D. Function, BPI-I average mean change from baseline: -1.84 vs. -2.40 vs. -1.92 vs. -1.61; B vs. D: p<0.05 Quality of life, mean change SF-36 subscales: -Bodily pain: 1.51 vs. 1.95 vs. 2.11 vs. 1.36; B vs. D, C vs. D: p<0.05 No significant difference among groups for other subscales Quality of life, EuroQoL (EQ) 5D US Index score mean change from baseline: 0.04 vs. 0.07 vs. 0.08 vs. 0.05; no significant differences among groups Global improvement, CGI-S mean change from baseline: -0.53 vs. -0.94 vs. -1.06 vs. -0.53; B vs. D, C vs. D: p<0.05
Skljarevski, 2010 (97) 12 weeks Chronic *Fair*	A. Duloxetine 60 mg daily (n=198) B. Placebo (n=203)	A. vs. B. Mean age 55 vs. 53 years 60% vs. 63% female Race: 96% vs. 95% white, 3% vs. 3% African, 2% vs. 3% other Pain, mean BPI 5.8 vs. 5.8 Function, mean CGI-S 3.5 vs. 3.3 Function, mean RDQ 9.6 vs. 9.3	A. vs. B. Pain, BPI-S mean change from baseline: -2.25 vs. -1.65; p=0.002 Pain, BPI 24-hour Average Pain Score, proportion of patients with 30% improvement in score: 57% (111/195) vs. 49% (97/199); p=0.11; 50% improvement in score: 49% (95/195) vs. 35% (69/199); p=0.005	A. vs. B. Function, BPI-I scale, mean change from baseline: -2.01 vs. -1.43; p≤0.001 Function, RDQ mean change from baseline: -2.69 vs. -2.22; p=0.26 Quality of life, Profile of Mood states total mood disturbance mean change from baseline: -6.77 vs. -2.77; p≤0.001 Global improvement, CGI-S mean change from baseline: -0.95 vs. -0.79; p=0.08 Global improvement, Patients' Global Impressions score, mean change from baseline: 2.88 vs. 3.19; p=0.01

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Skljarevski, 2010 (98) 13 weeks Chronic *Fair*	A. Duloxetine 60 mg daily; titrated to 120 mg daily in nonresponders after week 7 (n=115) B. Placebo; sham titration in nonresponders after week 7 (n=121)	A. vs. B. Mean age 52 vs. 51 years 62% vs. 60% female Race: 74% vs. 75% white, 20% vs. 17% Hispanic, 6% vs. 7% other Pain, mean BPI 5.9 vs. 6.0 Function, mean CGI-S 3.2 vs. 3.2	A. vs. B. Pain, BPI-S mean change from baseline: -2.66 vs. -1.90; p<0.05 Pain, BPI 24-hour Average Pain Score mean change from baseline: -2.08 vs. -1.30; p≤0.01	A. vs. B. Function, BPI-I, mean change from baseline: -1.92 vs. -1.18; p≤0.01 Quality of life, Athens Insomnia Scale mean change from baseline: -2.07 vs. -1.49; p=0.38 Quality of life, SF-36 mean between group difference significant for bodily pain (p=0.04), general health (p=0.04) and vitality (p=0.04) subscales favoring duloxetine; no difference for other subscales (data not shown) Return to work, mean between-group difference significant for WPAI measure of health outcomes subscale (p=0.002) favoring duloxetine; no difference for other subscales (data not shown) Global improvement, CGI-S mean change from baseline: -0.98 vs. -0.77; p=0.14

BPI=Brief Pain Inventory; BPI-I=Brief Pain Inventory Interference scale; BPI-S=Brief Pain Inventory Severity scale; CGI-S=Clinical Global Impressions of Severity scale; RDQ=Roland Morris Disability Questionnaire; VAS=visual analogue scale; WPAI=work productivity and activity impairment.

Supplement Table 10. Characteristics and conclusions of antiseizure medication trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Baron, 2010 (108) 5 weeks Subacute, chronic *Fair*	Placebo run-in period for 7 days, then pregabalin run-in for 28 days, then: A: Pregabalin: Optimal dose from run-in period (mean 410 mg) x 5 weeks, then 1 week taper (n=110) B: Placebo: Pregabalin taper x 1 week, then placebo x 4 week, then taper x 1 week (n=108)	A. vs. B. Mean age: 52 vs.53 years Female: 49% vs. 55% Baseline pain (mean, 0-10 VAS): 6.36 vs. 6.39 Baseline function: Not reported	A. vs. B. Pain (mean change from baseline, 0-10 VAS): -0.16 vs. 0.05 (p=0.33) Pain ≥7/10 (days): 7.1% (8/108) vs. 6.4% (7/107) at 5 weeks	A. vs. B. Loss of response (≥1 point increase in weekly mean pain score or use of rescue medication): 27.8% vs. 28.0% at 5 weeks, HR 0.87 (95% CI 0.52 to 1.47) Medical Outcome Study Sleep Scale sleep disturbance (mean change, 0-100): 2.26 vs. 6.86 (p=0.03) Medical Outcome Study Sleep Scale sleep quantity (mean change, hours): 0 vs. -0.43 (p=0.004) No differences on other MOS Sleep Scale subscales HADS anxiety (mean change, 0-21): -0.19 vs. 0.82 at 5 weeks (p=0.01) HADS depression (mean change, 0-21): -0.57 vs. 0.56 at 5 weeks (p=0.0006) EQ-5D, RDQ: No differences, data not reported
Baron, 2014 (109) 9-10 weeks Subacute, chronic *Fair*	Washout for 3-14 days, then tapentadol PR run-in for 3 weeks, then: A: Pregabalin + tapentadol PR: Pregabalin 150 mg/day x 1 week, 300 mg/day x 7 week + tapentadol PR 300 mg/day (n=157) B: Tapentadol PR: Tapentadol 300 mg/day + 100 mg/day x 1 week, tapentadol 300 mg/day + 200 mg/day x 7 week (n=152)	A. vs. B. Mean age: 56 vs.58 years Female: 54% vs. 62% Baseline pain: 5.9 vs. 5.9 (at randomization) Baseline function: Not reported	A. vs. B. Pain (mean change from baseline, 0-10 VAS): -1.6 vs. -1.7 at 9-10 weeks (p>0.05)	A. vs. B. Leg pain (mean change from baseline, 0-10 VAS): -1.6 vs. -1.9 at 9-10 weeks Patient satisfaction good, very good, or excellent: 73% (114/157) vs. 67% (102/152) at 9-10 weeks "Minimally", "much", or "very much" improved: 82% (129/157) vs. 81% (123/152) at 9-10 weeks SF-12: No difference on any subscale at 9-10 weeks EQ-5D (mean, 0-10): 0.60 vs. 0.61 at 9-10 weeks HADS anxiety (mean): 5.8 vs. 6.0 at 9-10 weeks HADS depression (mean): 5.4 vs. 6.2 at 9-10 weeks

Author, Year

Duration of Followup
LBP Duration

Quality	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Kalita, 2014 (110) 14 weeks Chronic *Poor*	A: Pregabalin: 75 mg bid x 2 weeks, 150 mg bid x 2 weeks, 300 mg bid, then increased if tolerated and needed (mean dose ~430 mg/day) (n=97) B: Amitriptyline: 12.5 nightly x 2 weeks, 25 mg nightly x 4 weeks, then 50 mg nightly, then increased if tolerated and needed (mean dose 38 mg/day) (n=103)	A. vs. B. Mean age: 42 vs.42 years Female: Not reported Baseline pain: 6.7 vs. 6.7 Baseline ODI: 42 vs. 42 Radiculopathy: 47% Spinal stenosis: 6%	A. vs. B. Pain (mean, 0-10 VAS): 6.7 vs. 6.7 at baseline, 4.2 vs. 3.9 at 4 weeks, 3.8 vs. 2.8 at 16 weeks (estimated from graph; p>0.05 at all-time points) Pain improved by ≥50%: 39% (38/97) vs. 57% (59/103), RR 0.68 (95% CI 0.51 to 0.92) Findings for dichotomous outcomes similar for patients with nonradicular back pain and radiculopathy; with or without neurological deficit	A. vs. B. ODI (mean, 0-100): 42 vs. 42 at baseline, 30 vs. 26 at 4 weeks, 22 vs. 17 at 16 weeks (estimated from graph; p>0.05 at all-time points) ODI improved >20%: 50% (48/97) vs. 65% (67/103), RR 0.76 (955 CI 0.59 to 0.97) Findings for dichotomous outcomes similar for patients with nonradicular back pain and radiculopathy; with or without neurological deficit
Markman, 2014 (111) 10 days Subacute, chronic *Fair*	A: Pregabalin: 75 mg by mouth twice daily x 3 days, 150 mg twice daily x 7 days, 75 mg twice daily x 4 days (n=14) B: Placebo: Diphenhydramine 6.25 mg po twice daily x 3 days, 12.5 mg twice daily x 7 days, 6.25 mg twice daily x 4 days (n=12) Each treatment for 2 weeks, with 1 week washout	A. vs. B. Mean age: 71 vs.69 years Female: 29% vs. 33% Baseline pain with ambulation (mean, 0-10 NRS): 7.7 vs. 7.1 Baseline RDQ (mean, 0-24): 13 vs. 14	A. vs. B. Pain with ambulation (mean, 0-10 NRS): 7.22 vs. 6.97 at 2 weeks (p=0.46) Brief Pain Inventory-Short Form, interference (mean, 0-10): 3.7 vs. 3.58 at 2 weeks (p=0.68) BPI-SF, pain intensity (mean, 0-10): 4.4 vs. 4.5 at 2 weeks (p=0.68)	A. vs. B. Walking distance (mean, m): 237 vs. 261 at 2 weeks (p=0.35) RDQ (mean, 0-24): 13 vs. 11 at 2 weeks (p=0.01) ODI (mean, 0-100): 38 vs. 36 at 2 weeks (p=0.36) Swiss Spinal Stenosis Questionnaire, symptom severity (mean): 3.09 vs. 2.94 at 2 weeks (p=0.07) Swiss Spinal Stenosis Questionnaire, physical function (mean): 2.40 vs. 2.45 at 2 weeks (p=0.57)

Author, Year

Duration of Followup
LBP Duration

Quality	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Pota, 2012 (112) 3 weeks Chronic *Fair*	Buprenorphine run-in period for 3 weeks, then: A: Pregabalin 300 mg/day + transdermal buprenorphine 35 mcg/h x 3 weeks (n=22) B: Placebo + transdermal buprenorphine 35 mcg/h x 3 weeks (n=22)	A. vs. B. Mean age: 56 years (overall) Female: 50% (overall) Baseline pain (mean, 0-100 VAS): 35 vs. 32 Baseline function: Not reported	A. vs. B. Pain (mean, 0-100 VAS): 9.5 vs. 32.8 at 1 week, 6.1 vs. 32.8 at 2 weeks, 5.7 vs. 33.3 ($p<0.05$) at 3 weeks Short-Form McGill Pain Questionnaire Pain Rating Index (mean, 0-15): 9.2 vs. 16.5 at 1 week, 4.6 vs. 16.6 at 2 weeks, 3.7 vs. 16.2 at 3 weeks ($p<0.05$) SF-MPQ Present Pain Intensity (mean, 0-5): 0.4 vs. 1.7 at 1 weeks, 0.3 vs. 1.8 at 2 weeks, 0.3 vs. 2.0 at 3 weeks	A. vs. B. Sleep interference (mean, 0-10): 0.2 vs. 2.3 at 1 week, 0.7 vs. 1.8 at 2 weeks, 0.6 vs. 1.9 at 3 weeks ($p>0.05$) Acetaminophen use (mean, mg): 46 vs. 636 at week 3 ($p<0.05$)
Romano, 2009 (113) 4 weeks Chronic *Fair*	A: Pregabalin ~1 mg/kg/d x 1 week, then 2-4 mg/kg/d (mean 2.1 mg/kg/d) (n=12) B: Celecoxib ~3-6 mg/kg/d (mean 4.2 mg/kg/d) (n=12) C: Pregabalin + celecoxib (mean 1.78 and 3.75 mg/kg/d) (n=12) Each treatment for 4 weeks, with 1 week washout prior to crossover	A. vs. B. vs. C. Mean age: 53 years (overall) Female: 56% (overall) Baseline pain: Not reported for initial intervention (mean 45-48) Baseline function: Not reported for initial intervention Disc prolapse: 47% Lumbar spondylosis: 39% Spinal stenosis: 19%	A. vs. B. vs. C. Pain (mean, 0-100 VAS): 43 vs. 40 vs. 29 at 4 weeks ($p=0.0001$ for A. vs. C. and $p=0.001$ for B vs. C) Pain reduction: 10% vs. 12% vs. 38% at 4 weeks LANSS score <12 Pain (mean, 0-100 VAS): 50.7 vs. 32.5 vs. 32.9 at 4 weeks ($p=0.0002$ for A. vs. C. and $p=0.9$ for B vs. C) Pain reduction (estimated from graph): -2.5% vs. 26% vs. 27% at 4 weeks LANSS score >12 Pain (mean, 0-100 VAS): 36.3 vs. 32.5 vs. 23.1 ($p=0.01$ for A. vs. C. and $p=0.0001$ for B vs. C) Pain reduction (estimated from graph): 23% vs. 2% vs. 52%	

Author, Year

Duration of Followup
LBP Duration

Quality	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Sakai, 2015 (115) 4 weeks Chronic *Poor*	A: Pregabalin 75 mg before bedtime (n=30) B: Tramadol 75 mg/acetaminophen 650 mg in twice daily divided doses (n=30)	A vs. B Mean age: 72 vs. 73 years Female: 30% vs. 37% Baseline low back pain (0-10 VAS): 6.0 vs. 6.7 Baseline leg pain (0-10 VAS): 4.1 vs. 3.1 Baseline RDQ: 9.7 vs. 11.5 Neuropathic pain (Neuropathic Pain Screening Questionnaire >6): 43% vs. 30%		Effective or remarkably effective: 73% vs. 83% Time to positive effects (mean, days): 10.2 vs. 6.1 ($p<0.05$)
Yaksi, 2007 (114) 4 months LBP duration not specified *Poor*	A: Gabapentin: initial dose 300 mg/day, titrated up to 2400 mg/day (mean not reported) (n=28) B: No gabapentin (n=27) Both groups also received exercise, lumbar corset, and NSAIDS; duration of treatment 4 months	A. vs. B. Mean age: 51 vs.51 years Female: 79% vs. 56% Baseline pain (mean, 0-10 VAS): 7.0 vs. 6.7 Baseline function: Not reported	A. vs. B. Pain (mean, 0-10 VAS): 5.1 vs. 5.6 at 1 month (p=0.40), 4.3 vs. 5.0 at 2 months (p=0.12), 3.6 vs. 4.8 at 3 months (p=0.04), 2.9 vs. 4.7 at 4 months (p=0.006)	A. vs. B. Walking distance >1000 m (estimated from graph): 65% vs. 21% at 4 months (p=0.001) Sensory deficit: 32% (9/28) vs. 63% (17/27)

CI=confidence interval, HADS=Hospital Anxiety and Depression Scale, LANSS=Leeds Assessment of Neuropathic Symptoms and Signs, LBP=low back pain, MPQ=McGill Pain Questionnaire, NSAIDS=nonsteroidal anti-inflammatory drug, ODI=Oswestry Disability Index, RDQ= Roland-Morris Disability Questionnaire, RR=relative risk, VAS=visual analogue scale

Supplement Table 11. Characteristics and conclusions of systemic corticosteroid trials

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Eskin, 2014 (120) 5-7 days Acute *Fair*	A: Prednisone: 50 mg by mouth once daily x 5 days (n=32) B: Placebo (n=35)	A. vs. B. Mean age: 39 vs. 41 years Female: 33% vs. 27% Baseline pain (mean, 0-10 VAS): 8.0 vs. 8.0 Baseline function: Not reported	A. vs. B. Pain (mean, 0-3 VRS): 1.3 vs. 1.1 at 5-7 days (difference 0.2, 95% CI -0.2 to 0.6) No or mild pain: 56% vs. 69% (difference -13%, 95% -36% to 10%)	A. vs. B. Days of work lost (mean): 2.1 vs. 1.3 (p=0.06) Sought further care: 40% vs. 18% (difference 22%, 95% CI 0% to 43%)
Friedman, 2008 (121) 1 month Acute *Good*	A: Methylprednisolone: 160 mg IM x 1 (n=37) B: Placebo (n=41)	A. vs. B. Mean age: 39 vs. 37 years Female: 54% vs. 51% Baseline pain (0-10 VAS): 8.9 vs. 9.1 Baseline function: Not reported	A. vs. B. Improvement in pain (mean, 0-10 VAS): difference 1.1 (95% CI -0.5 to 2.8) at 1 week; 7.1 vs. 5.8 at 1 month, difference 1.3 (95% CI -0.2 to 2.7) Back pain in prior 24 hours: 46% vs. 61% at 1 month, OR 0.54 (95% CI 0.22 to 1.3)	A. vs. B. Analgesic use in past 24 hours: 22% vs. 43% at 1 month, OR 0.39 (95% CI 0.14 to 1.1) RDQ18 (median, 0-18): 0 vs. 0 (p=0.009) RDQ18 1 or higher: 42% vs. 46% at 1 week; 19% vs. 49% at 1 m, OR 0.25 (95 5CI 0.09 to 0.7) Not resumed usual activities: 14% vs. 23% at 1 month, OR 0.56 (95% CI 0.17 to 1.9) Not resumed work (among full-time workers): 8% (2/24) vs. 13% (3/24) at 1 month, OR 0.64 (95% CI 0.10 to 4.2) Did not seek additional health care: 67% vs. 59% at 1 month, difference 8% (95% CI -14% to 30%)

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Goldberg, 2015 (125) 1 year Acute *Good*	A: Prednisone 60 mg po once daily x 5 days, 40 mg by mouth once daily x 5 days, 20 mg by mouth once daily x 5 days (n=181) B: Placebo (n=88)	A vs. B Mean age: 46 vs. 47 years Female: 46% vs. 42% Baseline pain (0-10 NRS): 6.6 vs. 6.9 Baseline ODI: 51.2 vs. 51.1	A vs. B Improvement in pain (mean, 0-10 NRS): -3.0 vs. -2.8 at 3 w, adjusted difference -0.3 (95% CI -1.0 to 0.4); -5.2 vs. -4.6 at 52 weeks, adjusted difference -0.6 (95% CI -1.3 to 0.2) Pain improved ≥3 points: 51% vs. 51% at 3 w, RR 1.0 (95% 0.8 to 1.3); 83% vs. 78% at 52 w, RR 1.1 (95% 0.9 to 1.2) Pain improved ≥5 points: 28% vs. 26% at 3 w, RR 1.1 (95% CI 0.8 to 1.7); 68% vs. 57% at 52 w, RR 1.2 (95% CI 0.9 to 1.4)	A vs. B Improvement in ODI: -19 vs. -13 at 3 weeks, adjusted difference -6.4 (95% CI -11 to -1.9); -38 vs. -30 at 52 weeks, adjusted difference -7.4 (95% CI -12 to -2.2) Improvement in SF-36 Physical Component Summary: 5.8 vs. 3.8 at 3 weeks, adjusted difference 3.3 (95% CI 1.3 to 5.2); 18 vs. 16 at 52 weeks, adjusted difference 2.5 (95% CI -0.3 to 5.4) Improvement in SF-36 Mental Component Summary: 1.2 vs. -0.7 at 3 weeks, adjusted difference 2.2 (95% CI -0.4 to 4.8); 6.9 vs. 3.1 at 52 weeks, adjusted difference 3.6 (95% CI 0.6 to 6.7) ODI improved ≥30 points: 27% vs. 17% at 3 weeks, RR 1.7 (95% CI 1.1 to 2.9); 71% vs. 57% at 52 weeks, RR 1.3 (95% CI 1.0 to 1.6) ODI improved ≥50%: 33% vs. 20% at 3 weeks, RR 1.8 (95% CI 1.1 to 2.9); 87% vs. 68% at 52 weeks, RR 1.2 (95% CI 1.1 to 1.5) Back surgery: 9.9% vs. 9.1% at 52 weeks, RR 1.2 (95% CI 0.5 to 2.6) Global patient assessment at least "somewhat better": 82% vs. 69% at 3 weeks, RR 1.2 (95% CI 1.0 to 1.4); 91% vs. 86% at 52 weeks, RR 1.1 (95% CI 1.0 to 1.2)

Author, Year **Duration of Followup** **LBP Duration** ***Quality***	**Intervention**	**Population**	**Pain Outcomes**	**Other Outcomes**
Hedeboe, 1982 (122) 3 months LBP duration not specified *Fair*	A: Dexamethasone: 4 mg/ml, 16 mg IM three times daily x 1 day, 8 mg three times daily x 1 day, 8 mg three times daily x 1 day, 4 mg three times daily x 1 day, 4 mg twice daily x 3 days (N=19) B: Placebo (n=20)	A. vs. B. Mean age: 44 vs. 40 years Female: 47% vs. 25% Baseline pain: Not reported Baseline function: Not reported		A. vs. B. Clear improvement (not otherwise defined): 68% (13/19) vs. 35% (7/20) at 9 days, RR 1.95, 95% CI 1.0 to 3.82; 32% (6/19) vs. 25% (5/20) at 3 months, RR 1.26, 95% CI 0.46 to 3.46
Holve, 2008 (123) 6 months Acute *Poor*	A: Prednisone: 60 mg by mouth once daily x 3 days, 40 mg by mouth once daily x 3 days, 20 mg by mouth once daily x 3 days (n=13) B: Placebo (n=14)	A. vs. B. Mean age: 39 vs. 46 years Female: 37% (overall) Baseline RDQ pain (mean, 0-5 VAS): 3.8 vs. 3.1 Baseline RDQ (mean, 0-24): 16 vs. 16	A. vs. B. RDQ Pain (mean, 0-5 RDQ pain, estimated from graph): 2.5 vs. 2.6 at 1 week, 1.8 vs. 2.1 at 2 weeks, 1.6 vs. 1.6 at 4 weeks, 1.5 vs. 1.0 at 3 months, 0.4 vs. 1.6 at 6 months (p>0.05)	A. vs. B. RDQ (mean, 0-24): 13 vs. 16 at 1 week, 8 vs. 13 at 2 weeks, 8 vs. 9 at 4 weeks, 3 vs. 2 at 3 months, 1 vs. 2 at 6 months (p>0.05) Return to baseline work hours: ~60% in each group by 2 months (p>0.05) NSAID and opioid use: No differences, data not provided Epidural injections: 15% (2/13) vs. 43% (6/14), RR 0.36 (95% CI 0.9 to 1.47)
Rodrigues, 2014 (124) 12 weeks LBP duration not specified *Fair*	A: Prednisone 1 mg/kg/day, reduced by 1/3 per week (n=31) B: Placebo (n=30)	A. vs. B. Mean age: 39 vs. 46 years Female: 37% (overall) Baseline RDQ pain (mean, 0-5 VAS): 3.8 vs. 3.1 Baseline RDQ (mean, 0-24): 16 vs. 16	A. vs. B. Pain (mean, 0-10 VAS): 7.68 vs. 7.07 at baseline, 5.68 vs. 5.50 at 3 weeks, 6.71 vs. 5.17 at 6 weeks, 6.61 vs. 5.97 at 12 weeks (p=0.02 at 6 weeks, otherwise p>0.05)	A. vs. B. RDQ (mean 0-24): 16.16 vs. 15.27 at baseline, 12.77 vs. 14.73 at 3 weeks, 14.71 vs. 13.80 at 6 weeks, 14.81 vs. 13.80 at 12 weeks (p>0.05 at all-time points) SF-36: No differences on any subscale Acetaminophen use: 19.42 vs. 19.6 (units unclear), p>0.05

CI=confidence interval, IM=intramuscular, LBP=low back pain, NSAIDS=nonsteroidal anti-inflammatory drug, ODI= Oswestry Disability Index, OR= odds ratio, RDQ= Roland-Morris Disability Questionnaire, RR=relative risk, VAS=visual analogue scale

Supplement Table 12. Pharmacological therapies versus active comparators for acute low back pain

Drug	Pain: Magnitude of Effect	Evidence	SOE
Acetaminophen vs. NSAID	No effect	1 SR (3 RCTs) + 1 RCT	Low
NSAID vs. NSAID	No effect in 15 of 21 RCTs	1 SR (21 RCTs)	Moderate
COX-2 selective NSAID vs. traditional NSAID	No effect	1 SR (3 RCTs)	Low
Skeletal muscle relaxant + NSAID vs. NSAID alone	RR 1.56 (0.92 to 2.70)	1 SR (2 RCTs) + 1 RCT	Low
Skeletal muscle relaxant vs. skeletal muscle relaxant	No effect	1 SR (2 RCTs)	Low

COX-2= cyclooxygenase-2, NSAID=nonsteroidal anti-inflammatory drug, RCT=randomized controlled trial, RR=relative risk, SOE=strength of evidence, SR=systematic review

Supplement Table 13. Pharmacological therapies versus active comparators for chronic low back pain

Drug	Pain: Magnitude of Effect	Evidence	SOE	Function: Magnitude of Effect	Evidence	SOE
Acetaminophen vs. NSAIDs	Unable to estimate	1 RCT	Insufficient	Unable to estimate	1 RCT	Insufficient
NSAID vs. NSAID	No difference	6 RCTs	Moderate	--	--	--
Opioids vs. NSAID	Unable to estimate (inconsistent)	3 RCTs	Insufficient	No difference	1 RCT	Insufficient
Long-acting opioids vs. long-acting opioids	No clear difference	4 RCTs	Moderate	No clear difference	4 RCTs	Moderate
Long-acting opioids vs. short-acting opioids	No clear difference*	6 RCTs	Low	--	--	--
Benzodiazepine (diazepam) vs. skeletal muscle relaxant	No difference	2 RCTs	Low	--	--	--
Skeletal muscle relaxant vs. skeletal muscle relaxant	No clear difference	1 SR (2 RCTs)	Low	--	--	--

NSAID=nonsteroidal anti-inflammatory drug, RCT=randomized controlled trial, SOE=strength of evidence, SR=systematic review

*Although some RCTs found long-acting opioids associated with greater pain relief, patients randomized to long-acting opioids also received higher doses of opioids

Supplement Table 14. Strength of evidence

Outcome	Study Design Number of Studies	Study Limitations	Consistency	Directness	Precision	Reporting Bias	Strength of Evidence
Acetaminophen							
Acetaminophen vs. Placebo, acute LBP : Pain and function	1 RCT	Low	Unable to determine	Direct	Precise	Undetected	Low
Acetaminophen vs. NSAID, acute LBP: Pain and global improvement	3 RCTs in systematic review and 1 RCT	High	Consistent	Direct	Precise	Undetected	Low
Acetaminophen vs. Placebo, chronic LBP	No studies	-	-	-	-	-	Insufficient
Acetaminophen vs. NSAID, chronic LBP	1 RCT	High	Unable to determine	Direct	Imprecise	Undetected	Insufficient
Acetaminophen vs. other interventions, acute LBP	4 RCTs	High	Consistent	Direct	Imprecise	Undetected	Insufficient
Acetaminophen vs. placebo: Adverse events (serious adverse events)	1 RCT	Low	Consistent	Direct	Imprecise	Undetected	Moderate
Acetaminophen vs. NSAIDs : Adverse events	3 RCTs in systematic reviews	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Acetaminophen vs. Placebo, NSAID or Other intervention, radicular LBP	No studies	-	-	-	-	-	Insufficient
NSAIDs							
NSAIDs vs. Placebo, acute LBP : Pain, function	4 RCTs in systematic review and 1 RCT for pain; 2 RCTs for function	Moderate	Consistent for pain Unable to determine for function	Direct	Precise for pain Imprecise for function	Undetected	Moderate for pain, low for function
NSAIDs vs. Placebo, chronic LBP : Pain, function	4 RCTs in systematic review and 2 RCTs for pain;4 RCTs for function	Moderate	Consistent	Direct	Precise for pain Imprecise for function	Undetected	Moderate for pain, low for function
NSAIDs vs. Placebo, radicular LBP : Pain	2 RCTs in systematic review	Moderate	Inconsistent	Direct	Imprecise	Undetected	Low
NSAID plus another intervention vs. Other intervention alone	2 RCTs	High	Inconsistent	Direct	Imprecise	Undetected	Insufficient
NSAIDs vs. Interventions other than acetaminophen and opioids	2 RCTs	High	Inconsistent	Direct	Imprecise	Undetected	Insufficient
NSAID vs. NSAID, acute or chronic LBP : Pain	27 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
NSAIDs vs. Placebo : Adverse events	10 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
COX-2-selective NSAIDs vs. nonselective NSAIDs : Adverse events	4 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate

Outcome	Study Design Number of Studies	Study Limitations	Consistency	Directness	Precision	Reporting Bias	Strength of Evidence
Opioids, Tramadol and Tapentadol							
Opioids vs. Placebo, chronic LBP : Pain and function	6 RCTs in systematic review and 4 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Tramadol vs. Placebo, chronic LBP : Pain and function	5 RCTs in systematic review and 2 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Buprenorphine patch vs. Placebo, subacute or chronic LBP : Pain and function	2 RCTs in systematic review	Moderate	Consistent for pain Inconsistent for function	Direct	Imprecise	Undetected	Low for pain Insufficient for function
Opioids vs. NSAIDs, chronic LBP : Pain relief, function	3 RCTs for pain 1RCT for function	Moderate	Inconsistent for pain Unable to determine for function	Direct	Imprecise	Undetected	Insufficient
Opioids vs. Acetaminophen, acute LBP : Days to return to work, pain	1 RCT for return to work No studies for pain	Moderate	Unable to determine	Direct	Imprecise	Undetected	Insufficient
Long acting opioids vs. Long acting opioids : Pain, function	4 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Long acting opioids vs. Short acting opioids : Pain	6 RCTs	Moderate	Inconsistent	Direct	Precise	Undetected	Low
Opioids vs. Placebo: Adverse events	16 RCTs in systematic review	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Skeletal Muscle Relaxants (SMR)							
SMRs vs. Placebo, acute LBP : Pain	4 RCTs in a systematic review and 1 RCT	Moderate	Consistent	Direct	Precise	Undetected	Moderate
SMR plus NSAID vs. NSAID alone, acute LBP : Pain	2 RCTs in systematic review and 1 RCT	Moderate	Consistent	Direct	Imprecise	Undetected	Low
SMR vs. Placebo, chronic LBP : Pain	3 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
SMR vs. SMR, acute or chronic LBP : Pain	3 RCTs in systematic review	Moderate	Consistent	Direct	Imprecise	Undetected	Low
SMR vs. Placebo, acute LBP : Adverse events	8 RCTs in systematic review and 1 RCT	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Benzodiazepines							

Outcome	Study Design Number of Studies	Study Limitations	Consistency	Directness	Precision	Reporting Bias	Strength of Evidence
Benzodiazepines vs. Placebo, acute LBP : Pain, function	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Tetrazepam vs. Placebo, chronic LBP: Pain, overall improvement	2 RCTs in systematic review	Moderate	Consistent	Direct	Imprecise	Undetected	Low
Diazepam vs. Placebo, acute or subacute radicular pain: Pain, function	1 RCT	Low	Unable to determine	Direct	Precise	Undetected	Low
Benzodiazepines vs. Skeletal muscle relaxants, chronic LBP: Pain, function	2 RCTs	Low	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Diazepam vs. Cyclobenzaprine, chronic LBP : Muscle spasms	1 RCT	Moderate	Unable to determine	Direct	Imprecise	Undetected	Low
Benzodiazepines vs. Placebo: Adverse events	8 RCTs in systematic review and 1 RCT	Moderate	Consistent	Direct	Imprecise	Undetected	Low
Antidepressants							
Tricyclic antidepressants or SSRI vs. Placebo, chronic LBP : Pain, function	4 RCTs of tricyclics and 3 RCTs of SSRIs in systematic review for pain; 2 RCTs evaluated function	Moderate	Consistent	Direct	Imprecise	Undetected	Moderate for pain, low for function
Duloxetine vs. Placebo, chronic LBP : Pain, Function	3 RCTs	Low	Consistent	Direct	Precise	Undetected	Moderate
Duloxetine vs. Tricyclic antidepressants	No studies	-	-	-	-	-	Insufficient
Antidepressants vs. Placebo : Adverse events, Serious adverse events	9 RCTs in systematic review and 3 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Antiseizure medications							
Antiseizure medications, acute non-radicular LBP	No studies	-	-	-	-	-	Insufficient
Gabapentin vs. Placebo, chronic non-radicular LBP	1 RCT (abstract only, excluded)	-	-	-	-	Suspected	Insufficient
Gabapentin vs. Placebo, chronic radicular LBP: Pain and function	3 RCTs	High	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Topiramate vs. Placebo, chronic radicular or mixed radicular and non-radicular LBP: Pain	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient

Outcome	Study Design Number of Studies	Study Limitations	Consistency	Directness	Precision	Reporting Bias	Strength of Evidence
Pregabalin vs. Placebo, chronic radicular LBP : pain, function	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Pregabalin plus transdermal buprenorphine vs. transdermal buprenorphine, chronic non-radicular LBP: Pain	1 RCT	Moderate	Unable to determine	Direct	Imprecise	Undetected	Insufficient
Pregabalin plus another analgesic vs. the other analgesic alone: Pain	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Gabapentin vs. Placebo : Adverse events	2 RCTs	Moderate	Consistent	Direct	Imprecise	Undetected	Low
Topiramate vs. Placebo : Withdrawal due to adverse events, sedation, diarrhea	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Pregabalin vs. Placebo : Withdrawal due to adverse events, somnolence, dizziness	2 RCTs	Moderate	Inconsistent	Direct	Imprecise	Undetected	Insufficient
Corticosteroids							
Systemic corticosteroids vs. Placebo, acute non- radicular LBP : Pain, function	2 RCTs	Moderate	Consistent	Direct	Imprecise	Undetected	Low
Systematic corticosteroids vs. Placebo, radicular LBP: Pain, function	6 RCTs	Moderate	Consistent	Direct	Precise	Undetected	Moderate
Systemic corticosteroids vs. Placebo, spinal stenosis: Pain, function	1 RCT	Moderate	Unable to determine	Direct	Precise	Undetected	Low
Systemic corticosteroids : Adverse events	12 RCTs	Moderate	Consistent	Direct	Imprecise	Undetected	Low

COX-2= cyclooxygenase-2, LBP=low back pain, NSAID=nonsteroidal anti-inflammatory drug, RCT=randomized controlled trial, SOE=strength of evidence, SMR=skeletal muscle relaxants, SSRI=selective serotonin reuptake inhibitor