Table II. Assessment of the credibility of the evidence across the 134 associations in the 12 eligible meta-analyses

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneitye	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Associations with convincing evidence[a]									
None of the meta-analyses was supported by strong evidence									
Associations with highly suggestive evidence[b]									
None of the meta-analyses was supported by highly suggestive evidence									
Associations with suggestive evidence[c]									
	Short-term outcomes (n = 5)								
Steffens, 2016 (8)	Episode of LBP	> 350 but < 500	LBP (prevention)	Exercise+education vs Control	> 10−6 but < 0.001	Including the null value	Not Large	Small-study effects	0.55 (0.41 to 0.74)
Kamper, 2014 (4)	Pain	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Excess significance bias	−0.55 (−0.83 to −0.27)
Kamper, 2014 (4)	Disability	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Excess significance bias	−0.41 (−0.62 to −0.19)
Van Middelkoop, 2011 (22)[f]	Pain intensity	> 350 but < 500	Chronic LBP	Multidis vs NT/WL	> 10^{-6} but < 0.001	NA	Not Large	No excess/Small-study effects NA	−0.45 (−0.67 to −0.22)
Guzman, 2002 (21)	Functional status	> 350 but < 500	Chronic LBP	Intensive (> 100h) daily Multidis with functional restoration vs Control	> 10^{-6} but < 0.001	Including the null value	Large	Neither	−0.66 (−1.02 to −0.31)
	Medium-term outcomes (n = 2)								
Kamper, 2014 (4)	Pain	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Both	−0.60 (−0.85 to −0.34)
Kamper, 2014 (4)	Disability	> 350 but < 1,000	Chronic LBP	Multidis vs TAU	> 10^{-6} but < 0.001	Including the null value	Large	Both	−0.43 (−0.66 to −0.19)
	Long-term outcomes (*n* = 1)								
Kamper, 2014 (4)	Work	> 1,000	Chronic LBP	Multidis vs Physical	> 10^{-6} but < 0.001	Excluding the null value	Not Large	Neither	1.87 (1.39 to 2.53)
	Associations with weak evidenced								
	Short-term outcomes (n = 19)								
Marin, 2017 (23)	Pain	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Not large	Neither	−0.40 (−0.74 to −0.06)
Marin, 2017 (23)	Disability	< 350	Subacute LBP	Multidis vs TAU	> 0.001 but < 0.05	Including the null value	Not large	Small-study effects	−0.38 (−0.63 to −0.14)
O'Keeffe, 2016 (6)	Disability	> 350 but 1,000	Chronic LBP + NP	Physical vs Physical+behavioural/ psychologically informed	> 0.001 but < 0.05	Including the null value	Large	Neither	0.27 (0.01 to 0.54)h
Kamper, 2014 (4)	Pain	> 1,000	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.30 (−0.54 to −0.06)
Kamper, 2014 (4)	Disability	> 1,000	Chronic LBP	Multidis vs Physical	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.39 (−0.68 to −0.10)
Kamper, 2014 (4)	Pain	< 350	Chronic LBP	Multidis vs WL	> 0.001 but < 0.05	Including the null value	Very large	Neither	−0.73 (−1.22 to −0.24)
Kamper, 2014 (4)	Disability	< 350	Chronic LBP	Multidis vs WL	> 10−6 but < 0.001	Including the null value	Not Large	Neither	−0.49 (−0.76 to −0.22)
Kamper, 2014 (4) [f]	QoL (MCS)	< 350	Chronic LBP	Multidis vs TAU	> 10−6 but < 0.001	NA	Not Large	No excess/Small-study effects NA	0.79 (0.45 to 1.14)
Kamper, 2014 (4)	Catastrophising	< 350	Chronic LBP	Multidis vs TAU	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	−0.43 (−0.83 to −0.03)
Kamper, 2014 (4)	Adverse events	> 350 but < 500	Chronic LBP	Multidis vs Surgery	> 0.001 but < 0.05	NA	Not Large	No excess/Small-study effects NA	28.25 (3.77 to 211.93)

Table II. *Cont.*

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneitye	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Van Middelkoop, 2011 (22)[f]	Pain intensity	>350 but <500	Chronic LBP	Multidis vs Active control	>0.001 but <0.05	NA	Large	No excess/Small-study effects NA	−0.56 (−0.98 to −0.15)
Van Middelkoop, 2011 (22)[f]	Disability	>350 but <500	Chronic LBP	Multidis vs NT/WL	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.34 (−0.54 to −0.15)
Norlund, 2009 (17)	Return to work	>1,000	Subacute and chronic LBP	Multidis vs Conservative	>0.001 but <0.05	Including the null value	Large	Excess significance bias	1.18 (1.06 to 1.31)
Häuser, 2009 (2)	Pain	<350	Fibromyalgia	Multidis vs Control	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.37 (−0.62 to −0.13)
Häuser, 2009 (2)	Fatigue	< 350	Fibromyalgia	Multidis vs Control	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.38 (−0.70 to −0.07)
Häuser, 2009 (2)	Depressive symptoms	<350	Fibromyalgia	Multidis vs Control	$>10^{-6}$ but <0.001	Including the null value	Large	Neither	−0.67 (−1.08 to −0.26)
Hoffman, 2009 (59)	Pain interference	>350 but <500	Chronic LBP	Multidis vs Active control	>0.001 but <0.05	NA	Not Large	NA	0.20 (0.02 to 0.37)
Guzman, 2002 (21)[f]	Pain rating	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	$>10^{-6}$ but <0.001	NA	Not Large	No excess/Small-study effects NA	−0.57 (−0.88 to −0.26)
Guzman, 2002 (21)	Employment status	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	$<10^{-6}$	NA	Not Large	No excess/Small-study effects NA	0.49 (0.31 to 0.68)
	Medium-term outcomes (n = 6)								
Kamper, 2014 (4)	Pain	>350 but <1,000	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	Including the null value	Large	Excess significance bias	−0.28 (−0.54 to −0.02)
Kamper, 2014 (4)	Work	<350	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	Including the null value	Not Large	Neither	2.14 (1.12 to 4.10)
Kamper, 2014 (4)[f]	QoL (PCS)	<350	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.42 (0.09 to 0.76)
Kamper, 2014 (4)[f]	QoL (MCS)	<350	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	NA	Not Large	No excess/mall-study effects NA	0.43 (0.09 to 0.76)
Kamper, 2014 (4)	Coping	<350	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	1.09 (0.31 to 1.87)
Hoffman, 2009 (59)	Disability: working	<350	Chronic LBP	Multidis vs Active control	>0.001 but <0.05	NA	Not Large	NA	0.36 (0.06 to 0.65)
	Long-term outcomes (n = 19)								
Marin, 2017 (23)	Pain	<350	Subacute LBP	Multidis vs TAU	$>10^{-6}$ but <0.001	Including the null value	Not large	Excess significance bias	−0.46 (−0.70 to −0.21)
Marin, 2017 (23)	Disability	<350	Subacute LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Large	Excess significance bias	−0.44 (−0.87 to −0.01)
Marin, 2017 (23)	Return-to-work	<350	Subacute LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Small-study effects	3.19 (1.46 to 6.98)
Marin, 2017 (23)	Sick leave periods	<350	Subacute LBP	Multidis vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.37 (−0.73 to −0.02)
Steffens, 2016 (8)	Episode of LBP	<350	LBP (prevention)	Exercise +education vs Control	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.73 (0.55 to 0.96)
O'Keeffe, 2016 (6)	Disability	>1,000	Chronic LBP+ NP	Physical vs Physical+behavioural/ psychologically informed	>0.001 but <0.05	Including the null value	Large	Both	0.25 (0.07 to 0.43)h
O'Keeffe, 2016[f] (6)	Pain	>1,000	Chronic LBP+ NP	Physical vs Physical+behavioural/ psychologically informed	>0.001 but <0.05	Including the null value	Not Large	Neither	0.18 (0.04 to 0.32)h

Table II. *Cont.*

Author, year	Outcome	Sample size (total N)	Pain condition	Intervention/control	Significance threshold reached (under the random-effects model)g	95% prediction interval rule	Estimate of heterogeneityе	Small-study effects or excess significance bias	Random-effects summary effect size (95% CI)
Kamper, 2014 (4)	Pain	>350 but <1,000	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.21 (−0.37 to −0.04)
Kamper, 2014 (4)	Disability	>350 but <1,000	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Excess significance bias	−0.23 (−0.40 to −0.06)
Kamper, 2014 (4)	Disability	>1,000	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	Including the null value	Very large	Small-study effects	−0.68 (−1.19 to −0.16)
Kamper, 2014 (4)	Catastrophizing	<350	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.40 (−0.76 to −0.05)
Kamper, 2014 (4)	Fear avoidance	>350 but <500	Chronic LBP	Multidis vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.29 (−0.49 to −0.08)
Kamper, 2014 (4)	Coping	<350	Chronic LBP	Multidis vs Physical	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.30 (0.06 to 0.54)
Schaafsma, 2013 (55)	Proportion off work	<350	Subacute LBP	Intense PCP vs Exercise	$>10^{-6}$ but <0.001	NA	Not Large	No excess/Small-study effects NA	0.57 (0.25 to 0.89)
Schaafsma, 2013 (55)	Time to return to work (>12 mo)	<350	Subacute LBP	Intense PCP + TAU vs TAU	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	−0.39 (−0.76 to −0.02)
Schaafsma, 2013 (55)	Time to return to work (12 mo)	>1,000	Chronic LBP	Intense PCP vs TAU	>0.001 but <0.05	Including the null value	Not Large	Neither	−0.23 (−0.42 to −0.03)
Hoffman, 2009 (59)	Disability: working	>350 but <1,000	Chronic LBP	Multidis vs Active control	>0.001 but <0.05	NA	Large	NA	0.53 (0.19 to 0.86)
Guzman, 2002 (21)	Functional status (60 mo)	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	>0.001 but <0.05	NA	Large	No excess/Small-study effects NA	−0.79 (−1.29 to −0.29)
Guzman, 2002 (21)	Employment status (12 mo)	<350	Chronic LBP	Intensive (>100 h) daily Multidis with functional restoration vs Control	>0.001 but <0.05	NA	Not Large	No excess/Small-study effects NA	0.34 (0.16 to 0.74)

[a]Convincing evidence criteria: >350 participants, significant summary associations ($p<10^{-6}$) per random-effects calculation, prediction intervals not including the null, heterogeneity not large ($I^2<50\%$), no evidence of small-study effects and no evidence of excess of significance bias.
[b]Highly suggestive evidence criteria: >350 participants, significant summary associations ($p<10^{-6}$) per random-effects calculation, and 95% prediction interval not including the null value.
[c]Suggestive evidence criteria: >350 participants and significant summary associations ($p>10^{-6}$ but <0.001) per random-effects calculation.
[d]Weak evidence criteria: all other treatment effects with $p\leq0.05$.
[e]Heterogeneity was categorized as not large ($I^2<50\%$), large ($I^2\geq50\%$ but $I^2<75\%$), and very large ($I^2\geq75\%$).
[f]On these comparisons MD is reported, instead of SMD.
[g]Random effects refer to summary effect (95% CI) using the random-effects model. The direction is arbitrary.
[h]Favour control, but in these 3 meta-analyses the control group was a multidisciplinary programme.
CBT: cognitive behavioural treatment; QoL: quality of life; PCS: physical component summary; MCS: mental component summary; LBP: low back pain; mo: months; NP: neck pain; Multidis: multidisciplinary programme; MBPSR: multidisciplinary bio-psychosocial rehabilitation programmes; PCP: physical conditioning programme; NT: no treatment; WL: waiting list; TAU: treatment as usual; CI: confidence interval; Control: not specified control group; NA: not applicable, because only 2 studies were available or information on included studies was not provided.