Appendix G: GRADE tables and meta-analysis results

G.1 Recognition, referral and diagnosis

G.1.1 Signs, symptoms and risk factors of spondyloarthritis

Review questions 1 & 2

- What signs and symptoms should prompt a healthcare professional to think of spondyloarthritis?
- What risk factors should increase suspicion of spondyloarthritis?

G.1.1.1 Inflammatory back pain

IBP (ASAS criteria)

Table 1: IBP (ASAS criteria) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	4 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	1,776	1.61 (1.42, 1.83)	LOW
LR-			Serious[b]	Serious[d]	Serious[c]	Serious[e]		0.55 (0.42, 0.74)	VERY LOW
PERIPHERAL									

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	4 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	1,776	1.61 (1.42, 1.83)	LOW
LR-			Serious[b]	Serious[d]	Serious[c]	Serious[e]		0.55 (0.42, 0.74)	VERY LOW

[a] *van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *I2 ≥ 50%*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

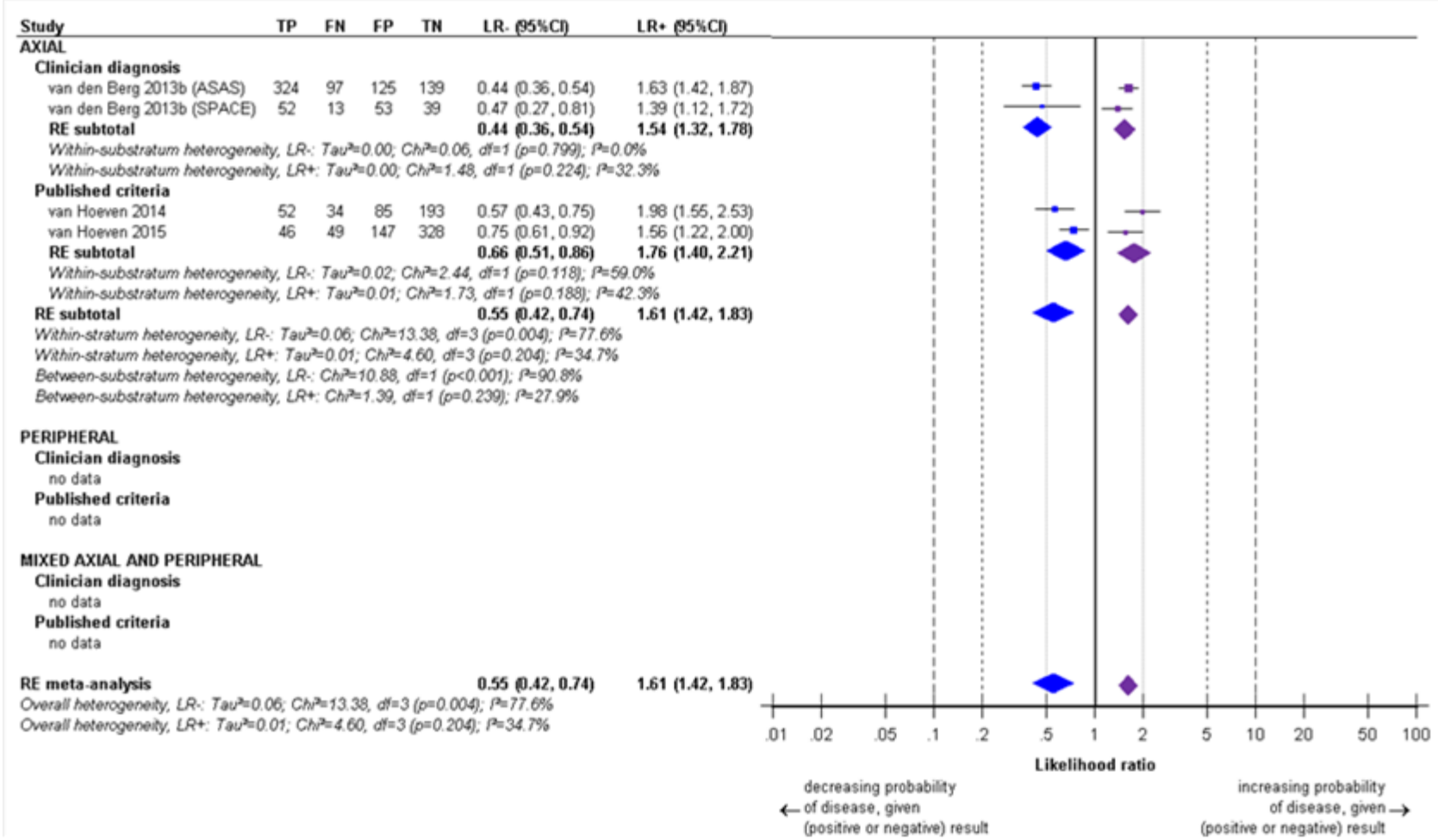


Figure 1 IBP (ASAS criteria) – forest plot: likelihood ratios

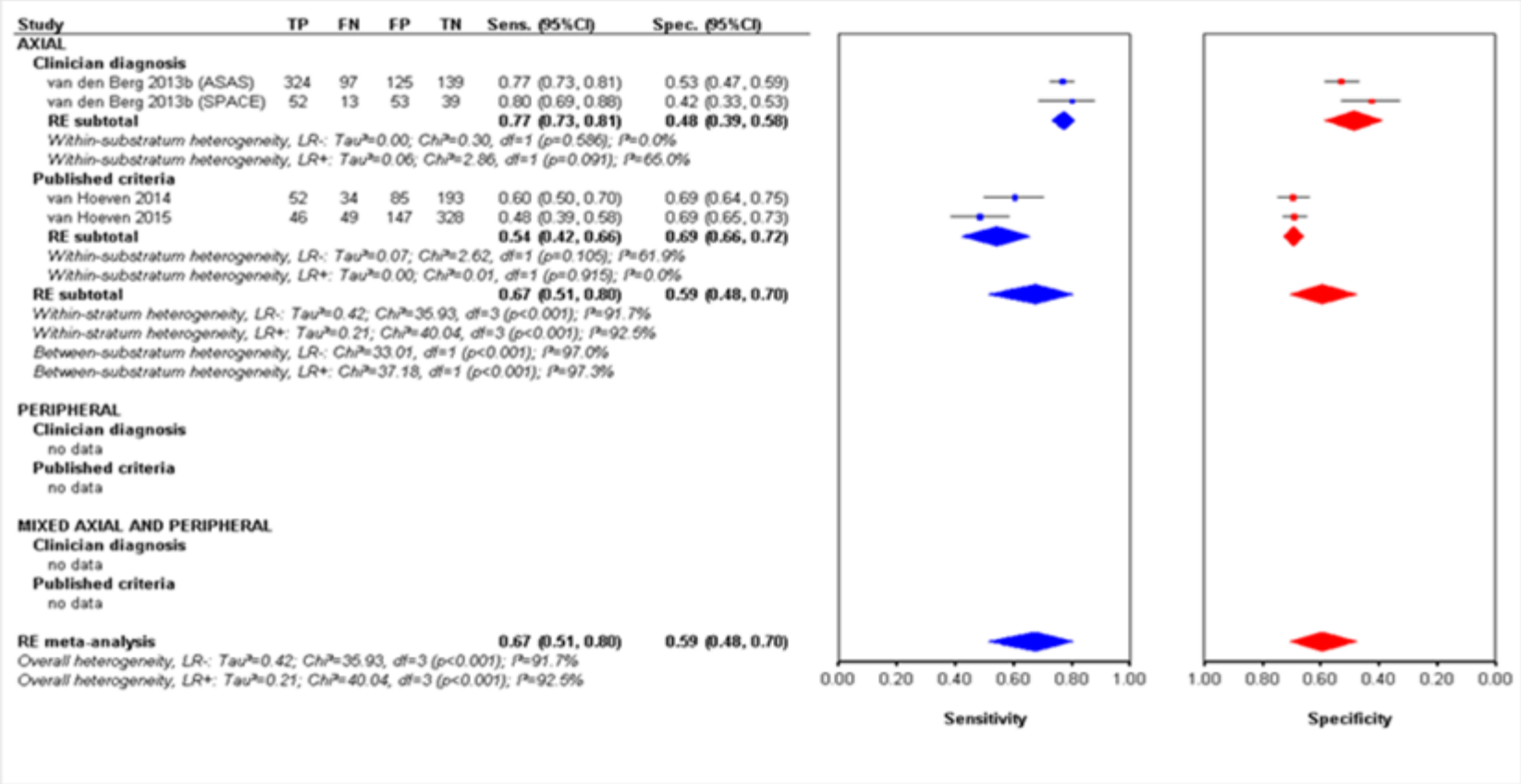


Figure 2: IBP (ASAS criteria) – forest plot: sensitivity and specificity

G.1.1.2 IBP (Berlin criteria)

Table 2: IBP (Berlin criteria) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	Serious[d]	Serious[e]	1,013	1.43 (0.98, 2.11)	VERY LOW
LR-			Serious[b]	No serious	No serious	No serious		0.58 (0.50, 0.68)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	Serious[d]	Serious[e]	1,013	1.43 (0.98, 2.11)	VERY LOW
LR-			Serious[b]	No serious	No serious	No serious		0.58 (0.50, 0.68)	MODERATE

[a] *Rudwaleit 2009 (ASAS); van Hoeven 2014*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	247	144	93	165	0.58 (0.49, 0.68)	1.75 (1.46, 2.10)
Published criteria						
van Hoeven 2014	67	19	183	95	0.65 (0.42, 0.99)	1.18 (1.03, 1.36)
RE subtotal					**0.58 (0.50, 0.68)**	**1.43 (0.98, 2.11)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.25, df=1 (p=0.620); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.07; Chi²=11.39, df=1 (p<0.001); I²=91.2%						
Between-substratum heterogeneity, LR-: Chi²=0.25, df=1 (p=0.620); I²=0.0%						
Between-substratum heterogeneity, LR+: Chi²=11.39, df=1 (p<0.001); I²=91.2%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.58 (0.50, 0.68)**	**1.43 (0.98, 2.11)**
Overall heterogeneity, LR-: Tau²=0.00; Chi²=0.25, df=1 (p=0.620); I²=0.0%						
Overall heterogeneity, LR+: Tau²=0.07; Chi²=11.39, df=1 (p<0.001); I²=91.2%						

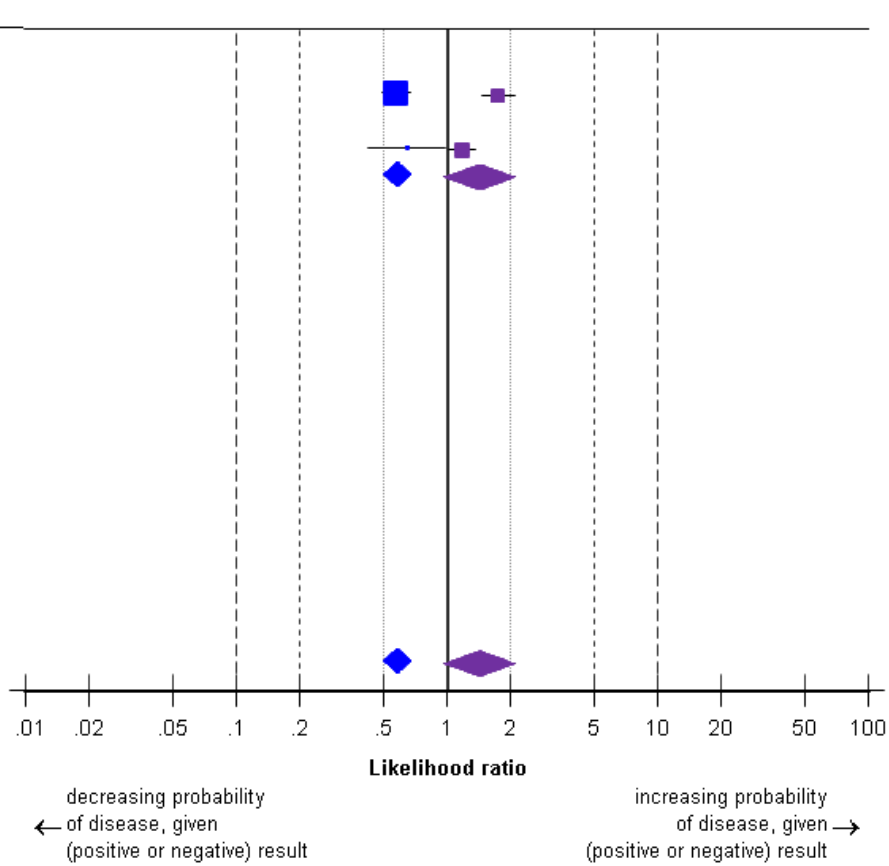


Figure 3: IBP (Berlin criteria) – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	247	144	93	165	0.63 (0.58, 0.68)	0.64 (0.58, 0.70)
Published criteria						
van Hoeven 2014	67	19	183	95	0.78 (0.68, 0.85)	0.34 (0.29, 0.40)
RE subtotal					**0.70 (0.54, 0.83)**	**0.49 (0.22, 0.76)**
Within-stratum heterogeneity, LR-: Tau²=0.22; Chi²=6.61, df=1 (p=0.010); I²=84.9%						
Within-stratum heterogeneity, LR+: Tau²=0.74; Chi²=46.04, df=1 (p<0.001); I²=97.8%						
Between-substratum heterogeneity, LR-: Chi²=6.61, df=1 (p=0.010); I²=84.9%						
Between-substratum heterogeneity, LR+: Chi²=46.04, df=1 (p<0.001); I²=97.8%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.70 (0.54, 0.83)**	**0.49 (0.22, 0.76)**
Overall heterogeneity, LR-: Tau²=0.22; Chi²=6.61, df=1 (p=0.010); I²=84.9%						
Overall heterogeneity, LR+: Tau²=0.74; Chi²=46.04, df=1 (p<0.001); I²=97.8%						

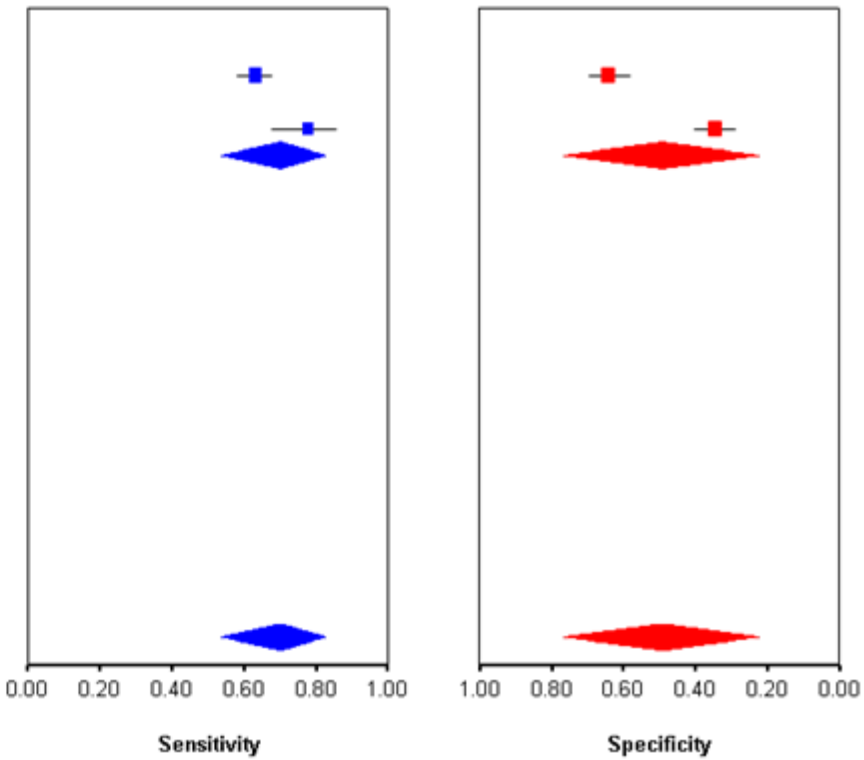


Figure 4: IBP (Berlin criteria) – forest plot: sensitivity and specificity

G.1.1.3 IBP (Calin criteria)

Table 3: IBP (Calin criteria) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	3 studies[a]	Cross-sectional	Serious[b]	Serious[c]	Serious[d]	No serious	1,105	1.34 (1.18, 1.53)	VERY LOW
LR-			Serious[b]	No serious	Serious[d]	No serious		0.36 (0.28, 0.47)	LOW
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	Serious	n/a	No serious	Serious[f]	81	11.19 (1.62, 77.17)	LOW
LR-			Serious	n/a	No serious	Serious[g]		0.51 (0.39, 0.68)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[h]	Cross-sectional	No serious	n/a	No serious	No serious	99	0.97 (0.76, 1.24)	HIGH
LR-			No serious	n/a	No serious	Serious[g]		1.09 (0.58, 2.04)	MODERATE
ALL EVIDENCE POOLED									
LR+	5 studies[i]	Cross-sectional	No serious	Serious[c]	Serious[d]	No serious	1,285	1.29 (1.08, 1.53)	LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[g]		0.49 (0.34, 0.70)	VERY LOW

[a] *Hermann 2009; Rudwaleit 2009 (ASAS); van Hoeven 2014*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[e] *Sadek 2007*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[g] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[h] *D'Agostino 2011*
[i] *D'Agostino 2011; Hermann 2009; Rudwaleit 2009 (ASAS); Sadek 2007; van Hoeven 2014*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	336	55	154	104	0.35 (0.26, 0.46)	1.44 (1.29, 1.60)
Published criteria						
Hermann 2009	27	3	39	23	0.27 (0.09, 0.83)	1.43 (1.14, 1.79)
van Hoeven 2014	75	11	200	78	0.46 (0.25, 0.82)	1.21 (1.09, 1.35)
RE subtotal					**0.41 (0.24, 0.68)**	**1.28 (1.10, 1.49)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.66, df=1 (p=0.415); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.01; Chi²=1.68, df=1 (p=0.194); I²=40.6%						
RE subtotal					**0.36 (0.28, 0.47)**	**1.34 (1.18, 1.53)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.93, df=2 (p=0.628); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.01; Chi²=5.23, df=2 (p=0.073); I²=61.8%						
Between-substratum heterogeneity, LR-: Chi²=0.27, df=1 (p=0.606); I²=0.0%						
Between-substratum heterogeneity, LR+: Chi²=3.55, df=1 (p=0.060); I²=71.8%						
PERIPHERAL						
Clinician diagnosis						
Sadek 2007	30	29	1	21	0.51 (0.39, 0.68)	11.19 (1.62, 77.17)
Published criteria						
no data						
RE subtotal					**0.51 (0.39, 0.68)**	**11.19 (1.62, 77.17)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	36	15	35	13	1.09 (0.58, 2.04)	0.97 (0.76, 1.24)
Published criteria						
no data						
RE subtotal					**1.09 (0.58, 2.04)**	**0.97 (0.76, 1.24)**
RE meta-analysis					**0.49 (0.34, 0.70)**	**1.29 (1.08, 1.53)**
Overall heterogeneity, LR-: Tau²=0.10; Chi²=12.24, df=4 (p=0.016); I²=67.3%						
Overall heterogeneity, LR+: Tau²=0.02; Chi²=15.93, df=4 (p=0.003); I²=74.9%						
Between-stratum heterogeneity, LR-: Chi²=11.31, df=2 (p=0.003); I²=82.3%						
Between-stratum heterogeneity, LR+: Chi²=10.69, df=2 (p=0.005); I²=81.3%						

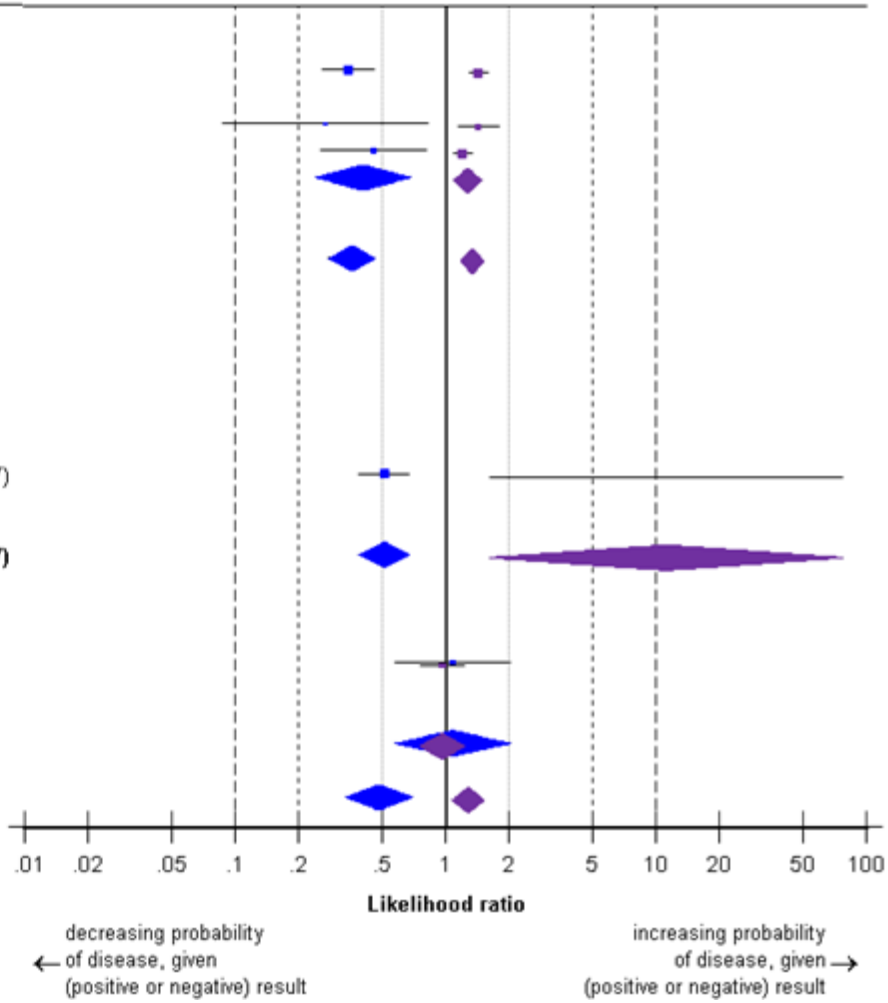


Figure 5: IBP (Calin criteria) – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	336	55	154	104	0.86 (0.82, 0.89)	0.40 (0.34, 0.46)
Published criteria						
Hermann 2009	27	3	39	23	0.90 (0.73, 0.97)	0.37 (0.26, 0.50)
van Hoeven 2014	75	11	200	78	0.87 (0.78, 0.93)	0.28 (0.23, 0.34)
RE subtotal					**0.88 (0.81, 0.93)**	**0.31 (0.23, 0.40)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.16, df=1 (p=0.687); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.04; Chi²=1.97, df=1 (p=0.161); I²=49.2%						
RE subtotal					**0.86 (0.83, 0.89)**	**0.35 (0.26, 0.44)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.45, df=2 (p=0.800); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.09; Chi²=9.09, df=2 (p=0.011); I²=78.0%						
Between-substratum heterogeneity, LR-: Chi²=0.28, df=1 (p=0.594); I²=0.0%						
Between-substratum heterogeneity, LR+: Chi²=7.12, df=1 (p=0.008); I²=86.0%						
PERIPHERAL						
Clinician diagnosis						
Sadek 2007	30	29	1	21	0.51 (0.38, 0.63)	0.95 (0.74, 0.99)
Published criteria						
no data						
RE subtotal					**0.51 (0.38, 0.63)**	**0.95 (0.74, 0.99)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	36	15	35	13	0.71 (0.57, 0.81)	0.27 (0.16, 0.41)
Published criteria						
no data						
RE subtotal					**0.71 (0.57, 0.81)**	**0.27 (0.16, 0.41)**
RE meta-analysis					**0.79 (0.63, 0.89)**	**0.37 (0.27, 0.49)**
Overall heterogeneity, LR-: Tau²=0.72; Chi²=42.67, df=4 (p<0.001); I²=90.6%						
Overall heterogeneity, LR+: Tau²=0.21; Chi²=23.21, df=4 (p<0.001); I²=82.8%						
Between-stratum heterogeneity, LR-: Chi²=42.22, df=2 (p<0.001); I²=95.3%						
Between-stratum heterogeneity, LR+: Chi²=14.12, df=2 (p<0.001); I²=85.8%						

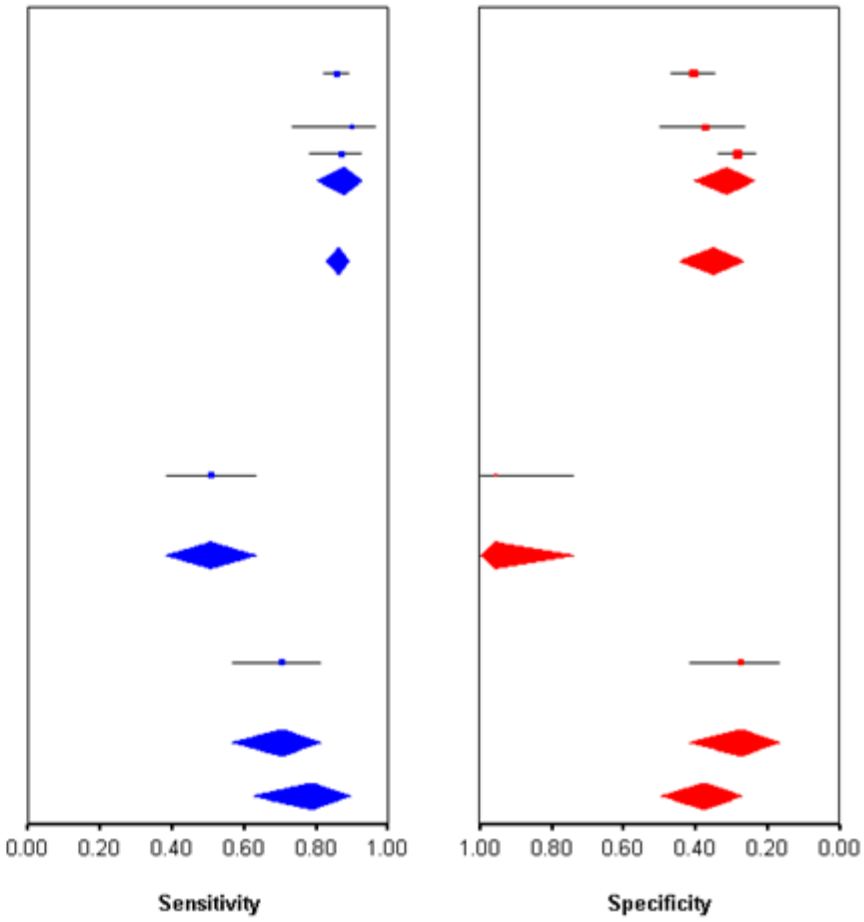


Figure 6: IBP (Calin criteria) – forest plot: sensitivity and specificity

G.1.1.4 IBP (ad hoc or unspecified definitions)

Table 4: IBP (ad hoc or unspecified definitions) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	3 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	No serious	2,107	1.25 (0.97, 1.60)	LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[d]		0.51 (0.23, 1.13)	VERY LOW
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	No serious	n/a	No serious	Serious[f]	266	1.42 (0.69, 2.91)	MODERATE
LR-			No serious	n/a	No serious	No serious		0.95 (0.87, 1.04)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[g]	Cross-sectional	No serious	Serious[c]	No serious	Serious[f]	880	1.47 (1.03, 2.08)	LOW
LR-			Serious[b]	Serious[c]	Serious[h]	Serious[d]		0.60 (0.44, 0.83)	VERY LOW
ALL EVIDENCE POOLED									
LR+	6 studies[i]	Cross-sectional	Serious[b]	Serious[c]	No serious	No serious	3,253	1.31 (1.10, 1.57)	LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[d]		0.60 (0.42, 0.87)	VERY LOW

[a] *Poddubnyy 2011; Rudwaleit 2009 (ASAS); Sieper 2013*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[e] *Rudwaleit 2011*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[g] *Althoff 2009; Tomero 2014*
[h] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[i] *Althoff 2009; Poddubnyy 2011; Rudwaleit 2009 (ASAS); Rudwaleit 2011; Sieper 2013; Tomero 2014*

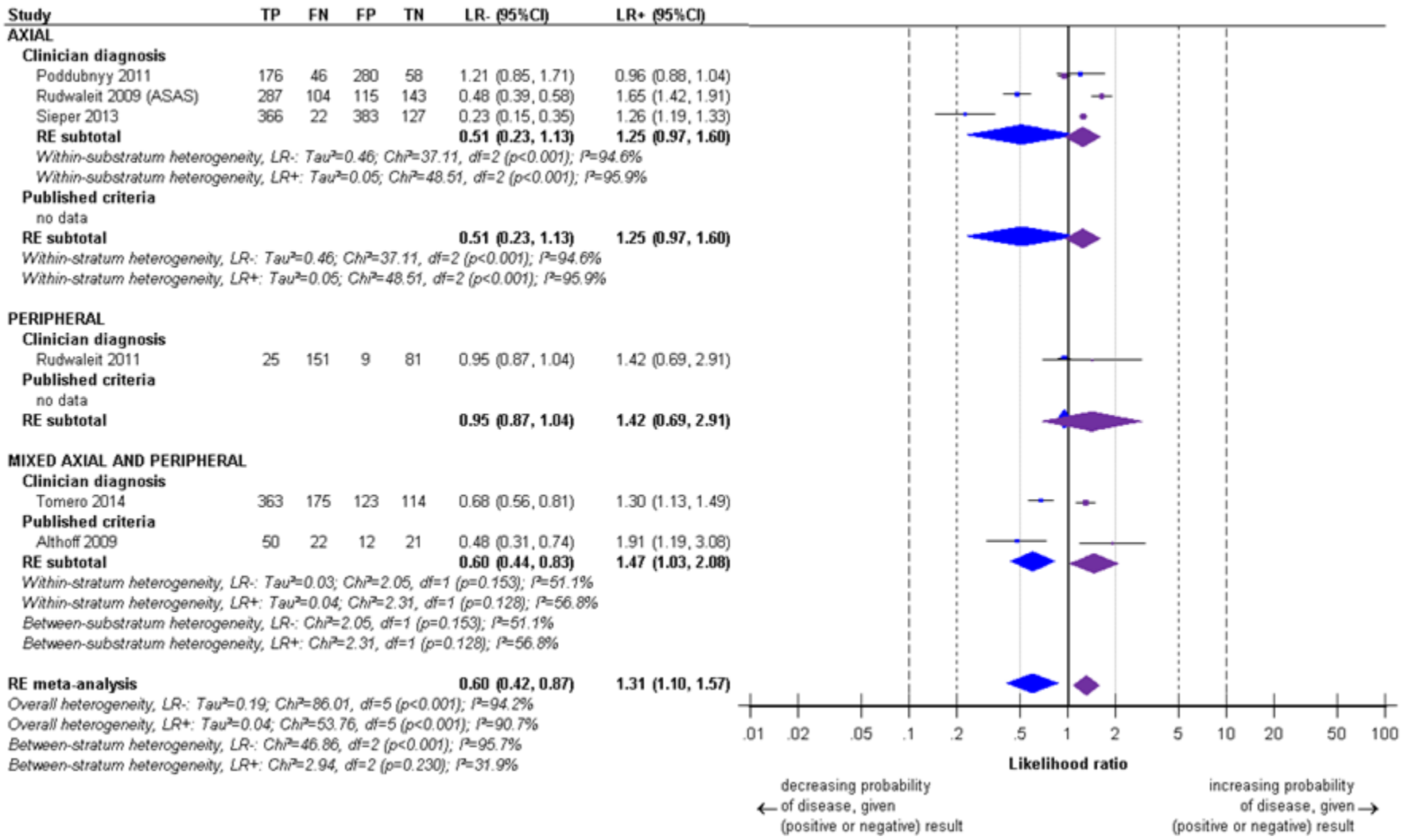


Figure 7: IBP (ad hoc or unspecified definitions) – forest plot: likelihood ratios

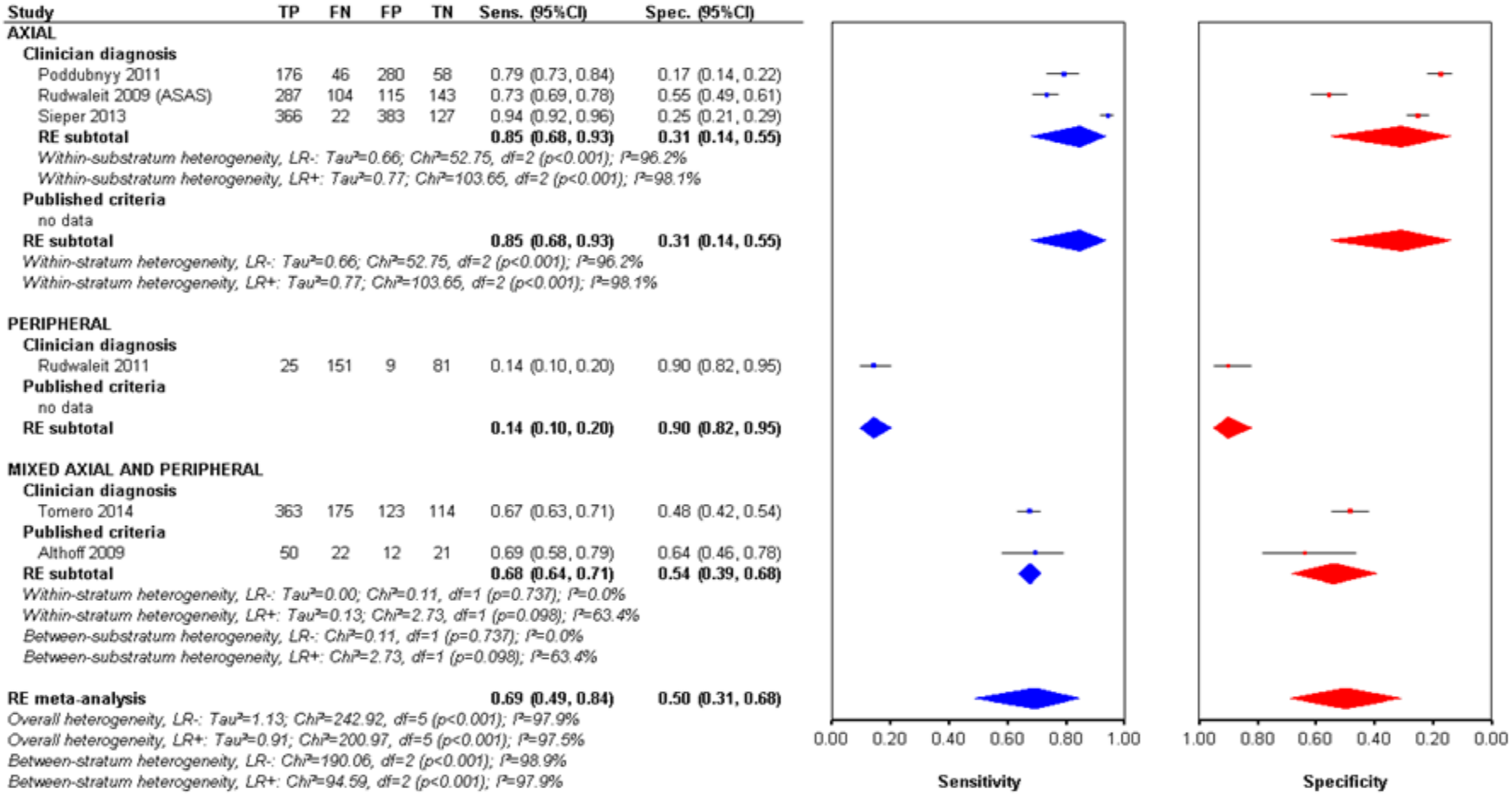


Figure 8: IBP (ad hoc or unspecified definitions) – forest plot: sensitivity and specificity

G.1.1.5 Back pain (in people with other presenting complaints)

Table 5: Back pain (in people with other presenting complaints) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	372	1.42 (0.88, 2.29)	LOW
LR-			No serious	n/a	Serious[b]	No serious		0.89 (0.74, 1.07)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[d]	Cross-sectional	No serious	Serious[e]	Serious[f]	No serious	876	0.98 (0.89, 1.09)	LOW
LR-			No serious	No serious	No serious	No serious		1.27 (0.95, 1.71)	HIGH
ALL EVIDENCE POOLED									
LR+	3 studies[g]	Cross-sectional	No serious	Serious[e]	Serious[f]	No serious	1,248	1.00 (0.89, 1.12)	LOW
LR-			No serious	Serious[e]	Serious[f]	No serious		1.03 (0.73, 1.46)	LOW

[a] *Kvien 1994*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[d] *Haroon 2015; Tomero 2014*
[e] *I2 ≥ 50%*
[f] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[g] *Haroon 2015; Kvien 1994; Tomero 2014*

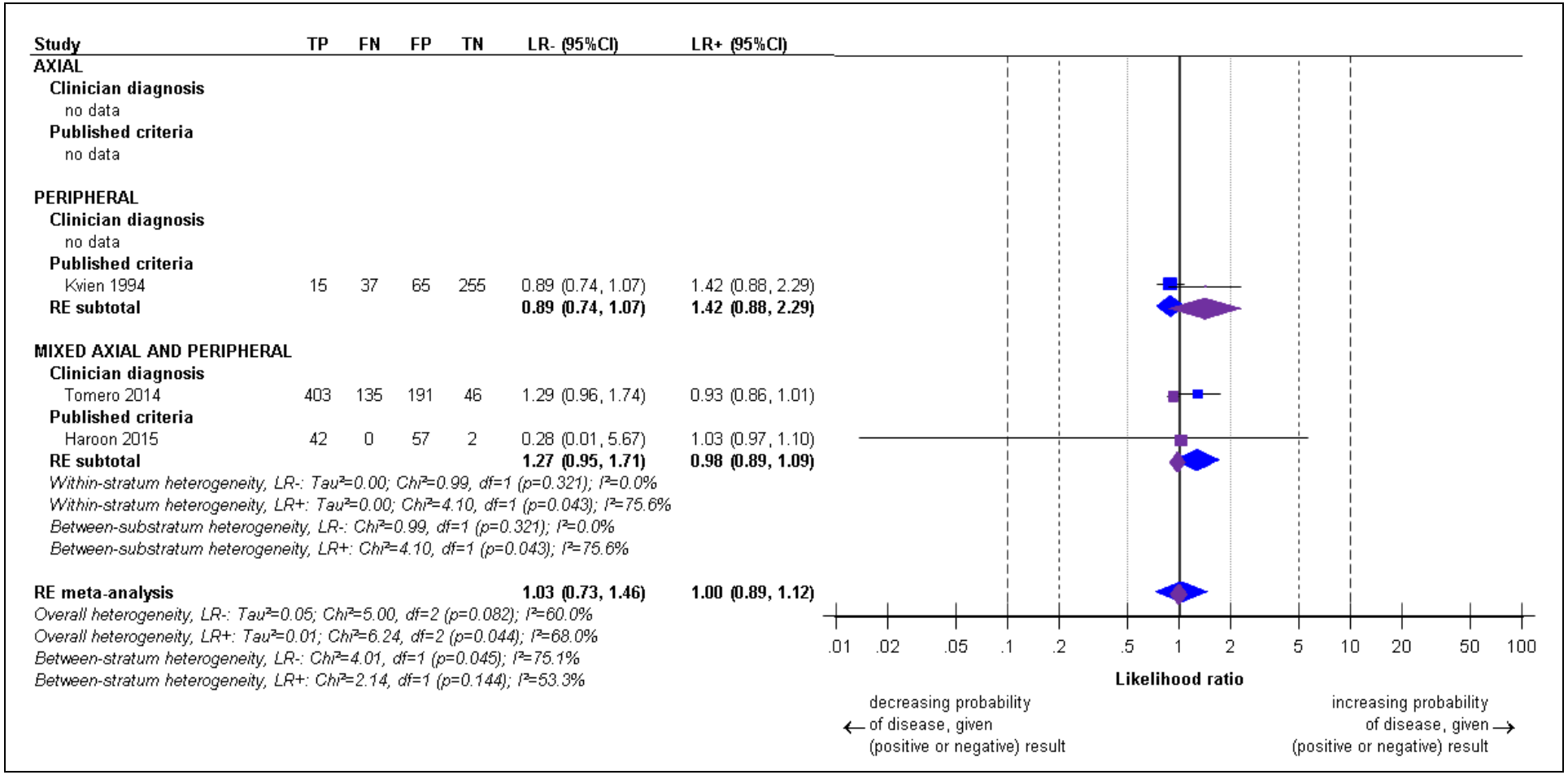


Figure 9 Back pain (in people with other presenting complaints) – forest plot: likelihood ratios

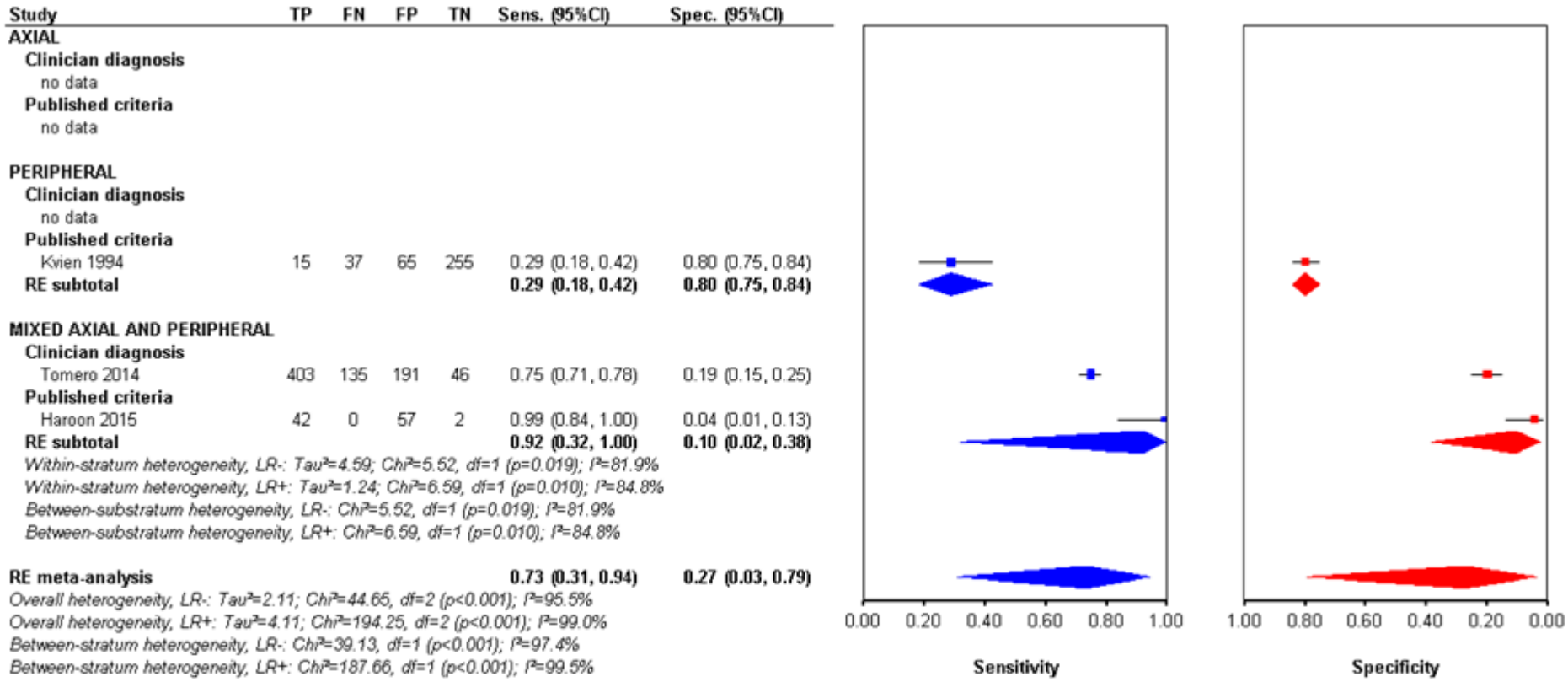


Figure 10: Back pain (in people with other presenting complaints) – forest plot: likelihood ratios

G.1.1.6 Age

Age <45 at onset of back pain

Table 6: Age <45 at onset of back pain – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	Serious[b]	No serious	787	3.29 (2.74, 3.96)	LOW
LR-			Serious	n/a	Serious[b]	No serious		0.34 (0.24, 0.48)	LOW
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	Serious[b]	No serious	787	3.29 (2.74, 3.96)	LOW
LR-			Serious	n/a	Serious[b]	No serious		0.34 (0.24, 0.48)	LOW

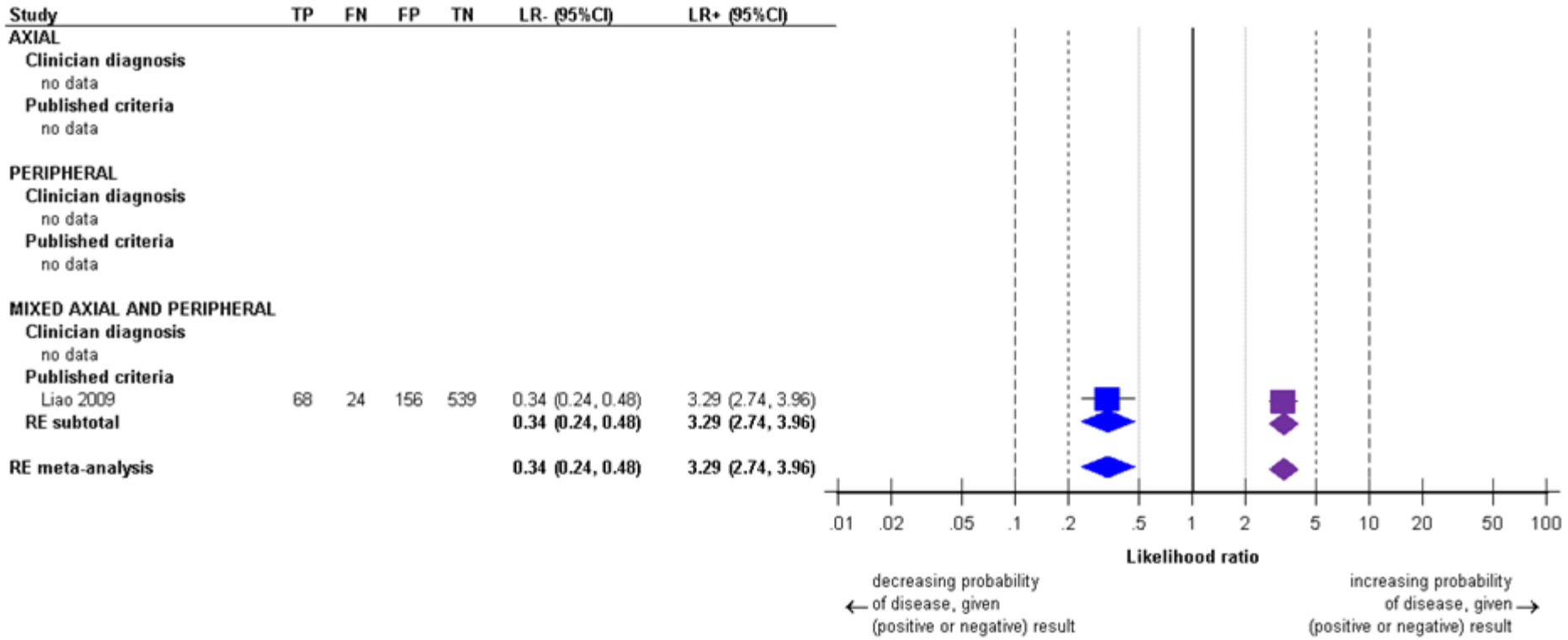


Figure 11: Age <45 at onset of back pain – forest plot: likelihood ratios

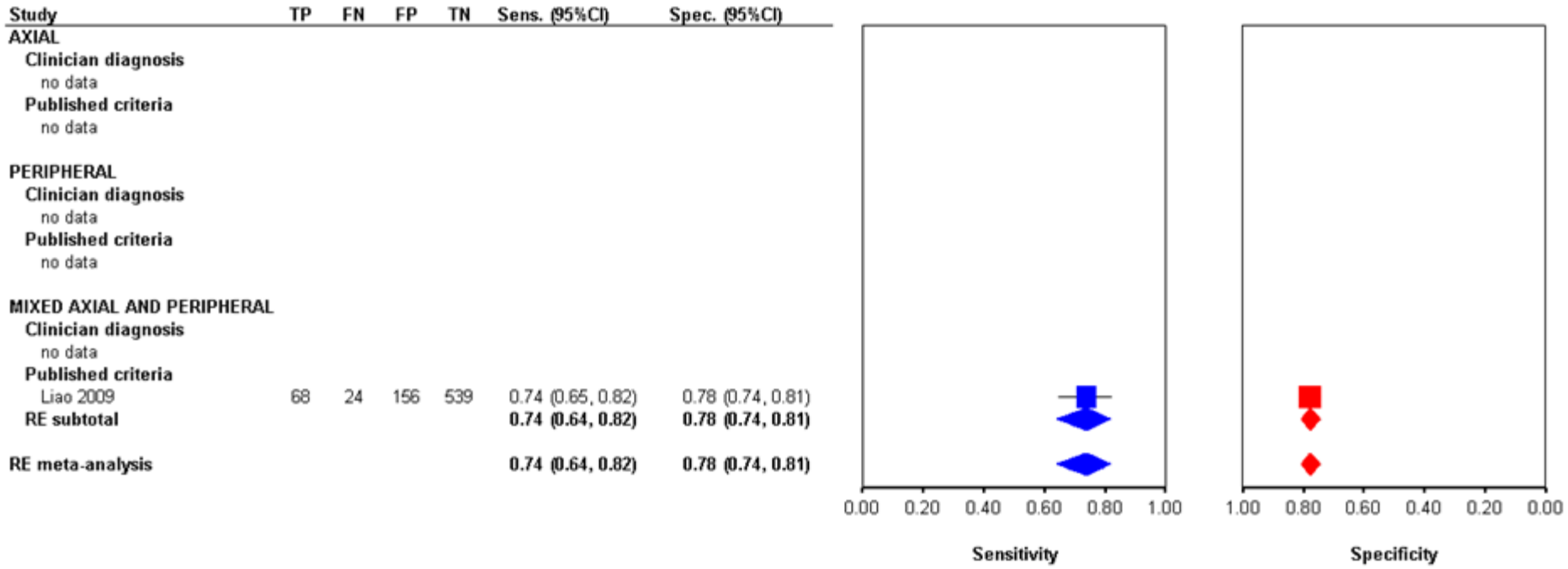


Figure 12: Age <45 at onset of back pain – forest plot: sensitivity and specificity

G.1.1.7 Age <35 at onset of back pain (in people aged <45 at onset of back pain)

Table 7: Age <35 at onset of back pain (in people aged <45 at onset of back pain) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	322	1.36 (1.17, 1.59)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.53 (0.36, 0.77)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	322	1.36 (1.17, 1.59)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.53 (0.36, 0.77)	MODERATE

[a] *Braun 2011*

[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Braun 2011	87	26	118	91	0.53 (0.36, 0.77)	1.36 (1.17, 1.59)
Published criteria						
no data						
RE subtotal					**0.53 (0.36, 0.77)**	**1.36 (1.17, 1.59)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.53 (0.36, 0.77)**	**1.36 (1.17, 1.59)**

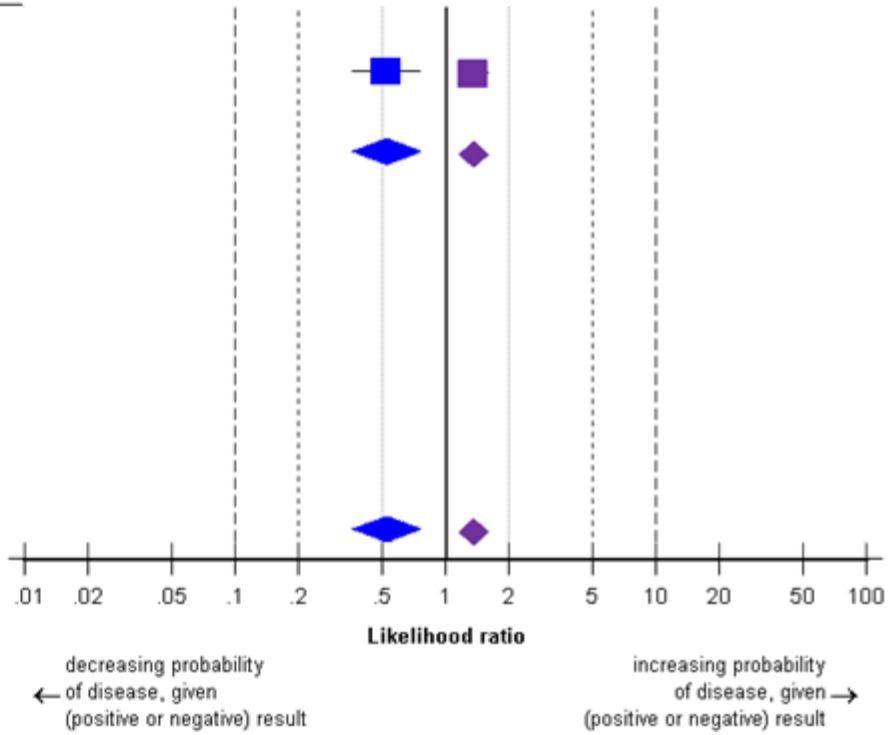


Figure 13: Age <35 at onset of back pain (in people aged <45 at onset of back pain) – forest plot: likelihood ratios

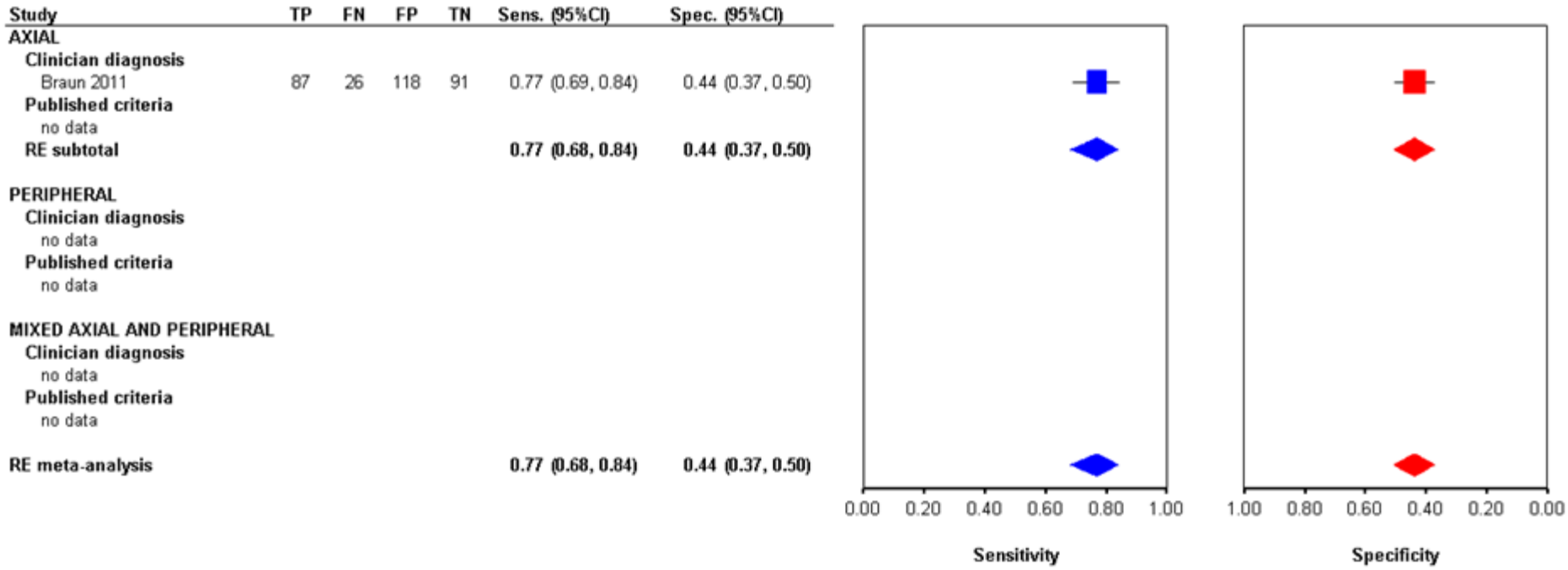


Figure 14: Age <35 at onset of back pain (in people aged <45 at onset of back pain) – forest plot: sensitivity and specificity

G.1.1.8 Age <40 at onset of back pain (in people aged <45 at onset of back pain)

Table 8: Age <40 at onset of back pain (in people aged <45 at onset of back pain) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	649	1.07 (1.01, 1.13)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.54 (0.33, 0.88)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	649	1.07 (1.01, 1.13)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.54 (0.33, 0.88)	MODERATE

[a] *Rudwaleit 2009 (ASAS)*

[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

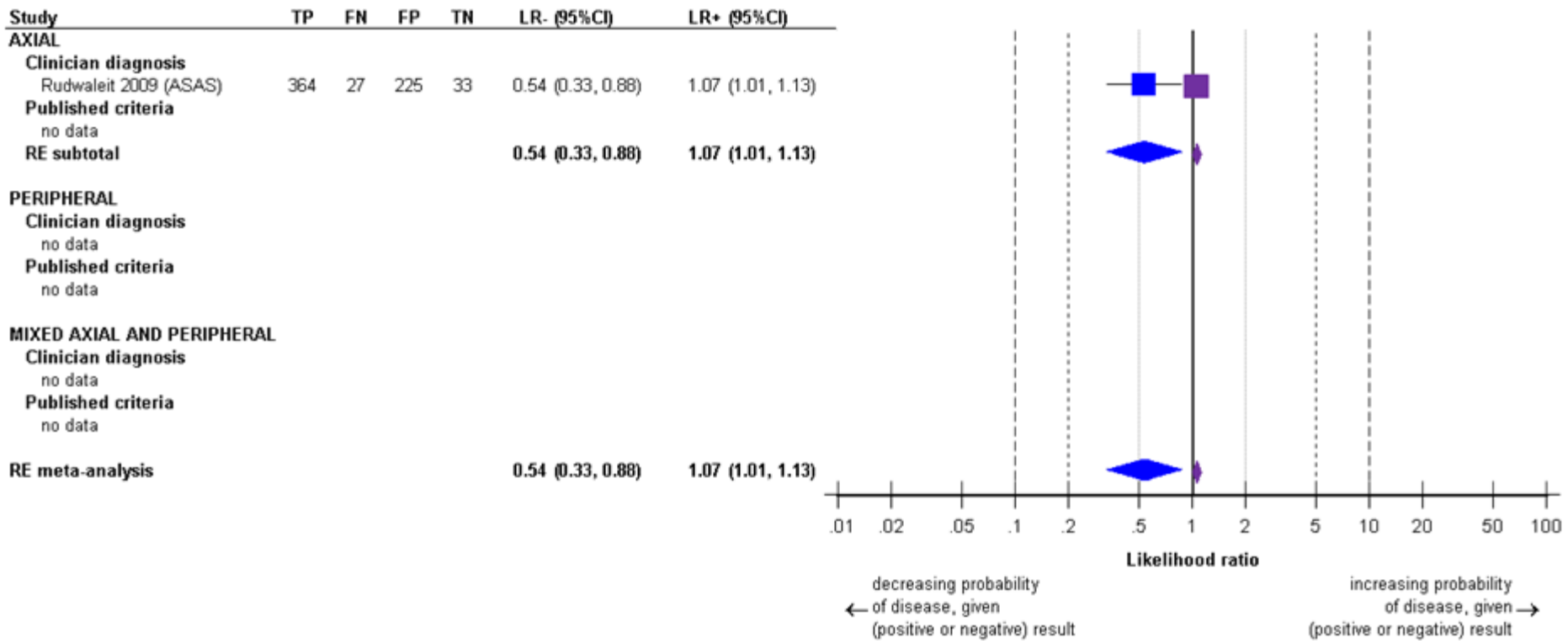


Figure 15: Age <40 at onset of back pain (in people aged <45 at onset of back pain) – forest plot: likelihood ratios

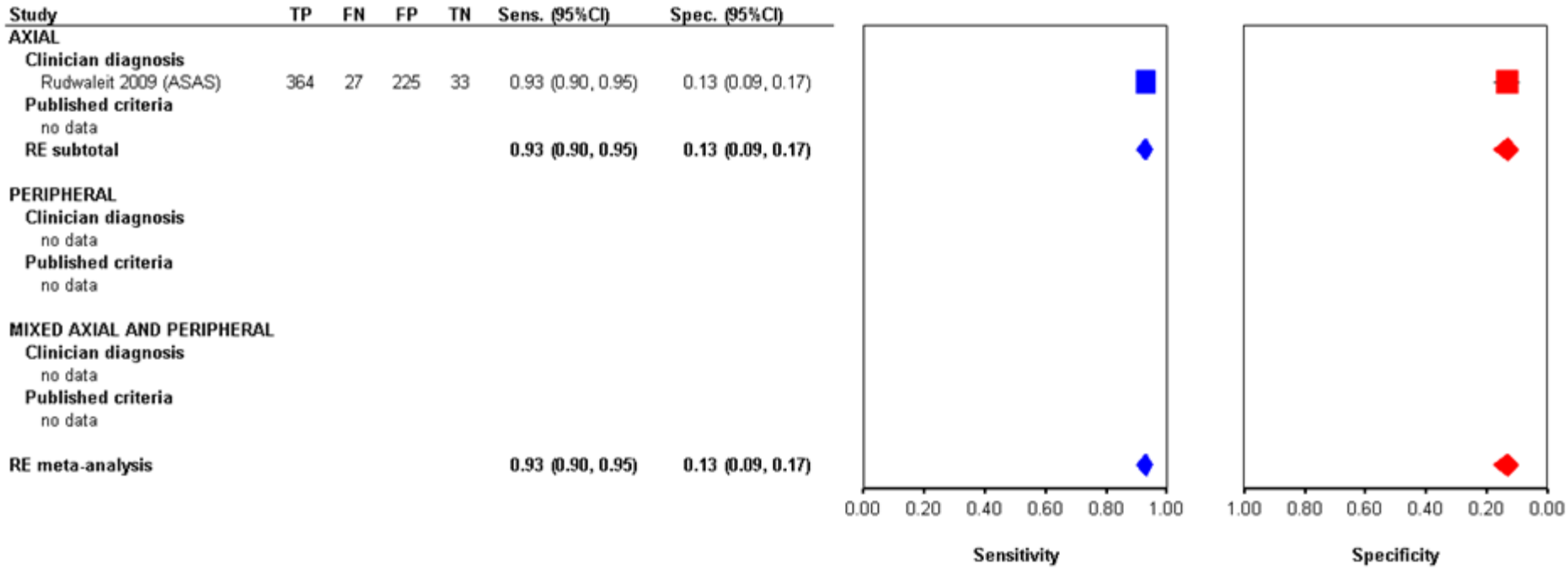


Figure 16: Age <40 at onset of back pain (in people aged <45 at onset of back pain) – forest plot: sensitivity and specificity

G.1.1.9 Back pain with age of onset <45 (in people with acute anterior uveitis)

Table 9: Back pain with age of onset <45 (in people with acute anterior uveitis) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	No serious	101	1.50 (1.25, 1.81)	MODERATE
LR-			No serious	n/a	Serious[b]	Serious[c]		0.03 (0.00, 0.55)	LOW
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	No serious	101	1.50 (1.25, 1.81)	MODERATE
LR-			No serious	n/a	Serious[b]	Serious[c]		0.03 (0.00, 0.55)	LOW

[a] *Haroon 2015*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

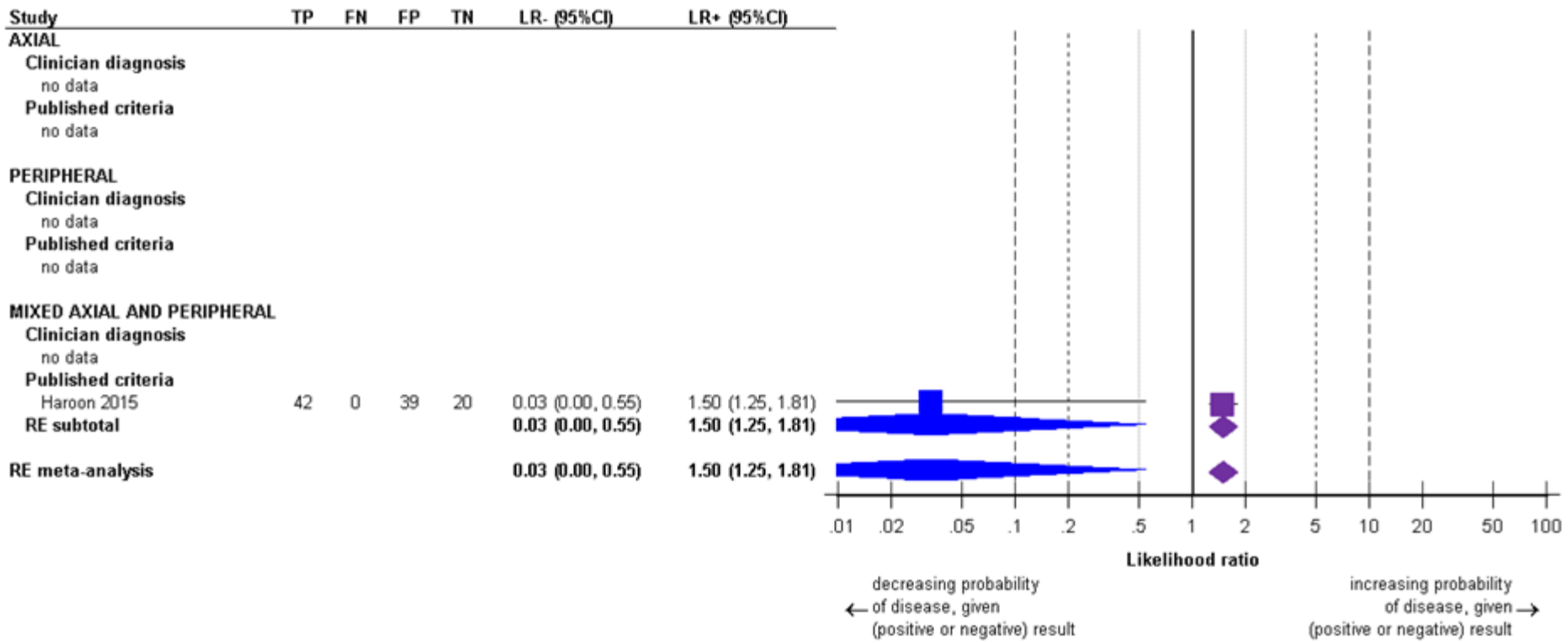


Figure 17: Back pain with age of onset <45 (in people with acute anterior uveitis) – forest plot: likelihood ratios

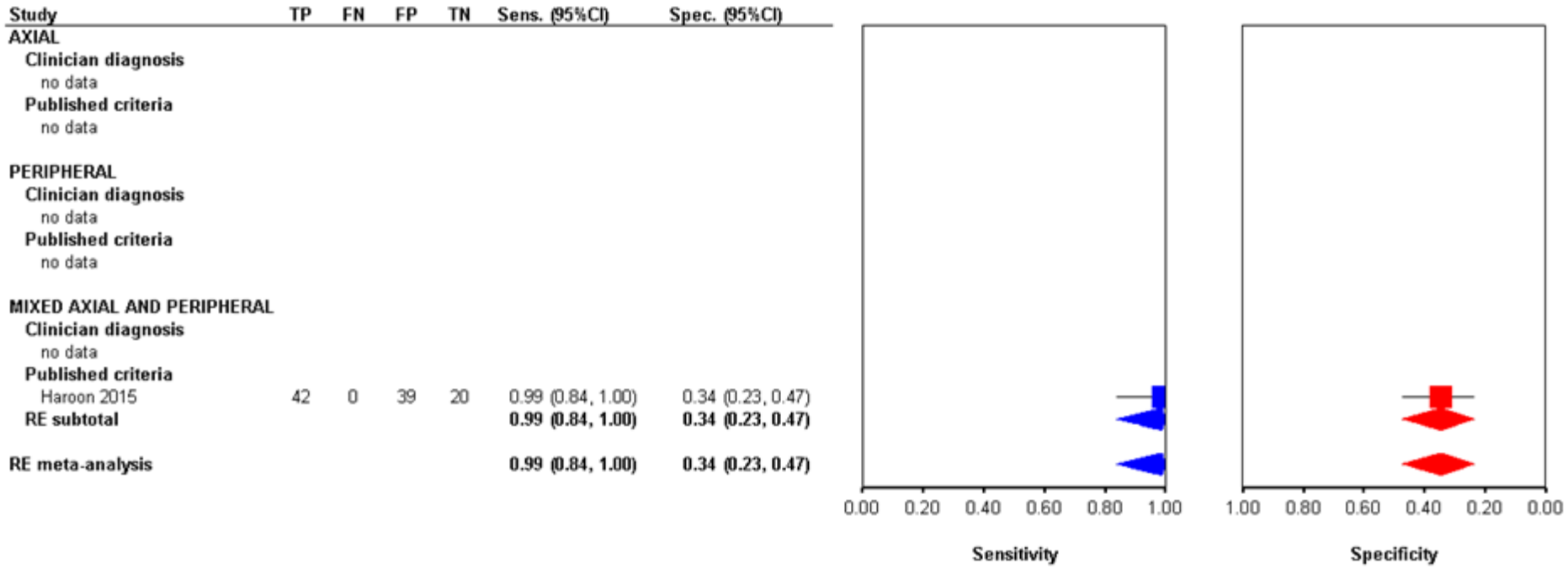


Figure 18: Back pain with age of onset <45 (in people with acute anterior uveitis) – forest plot: sensitivity and specificity

G.1.1.10 Morning stiffness

Table 10: Morning stiffness – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	322	1.06 (0.77, 1.45)	HIGH
LR-			No serious	n/a	No serious	No serious		0.97 (0.82, 1.15)	HIGH
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[b]	Cross-sectional	Serious	n/a	Serious[c]	No serious	787	5.25 (4.18, 6.58)	LOW
LR-			Serious	n/a	Serious[c]	No serious		0.33 (0.24, 0.45)	LOW
ALL EVIDENCE POOLED									
LR+	2 studies[d]	Cross-sectional	Serious[e]	Serious[f]	Serious[g]	V. serious[h]	1,109	2.36 (0.49, 11.37)	VERY LOW
LR-			Serious[e]	Serious[f]	Serious[g]	Serious[i]		0.57 (0.20, 1.65)	VERY LOW

[a] *Braun 2011*
[b] *Liao 2009*
[c] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *Braun 2011; Liao 2009*
[e] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[f] *I2 ≥ 50%*
[g] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[h] *At a 95% confidence level, data are consistent with meaningful predictive value in either direction and no predictive value at all (i.e. 95% CI for LR+ spans both 0.5 and 2).*
[i] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

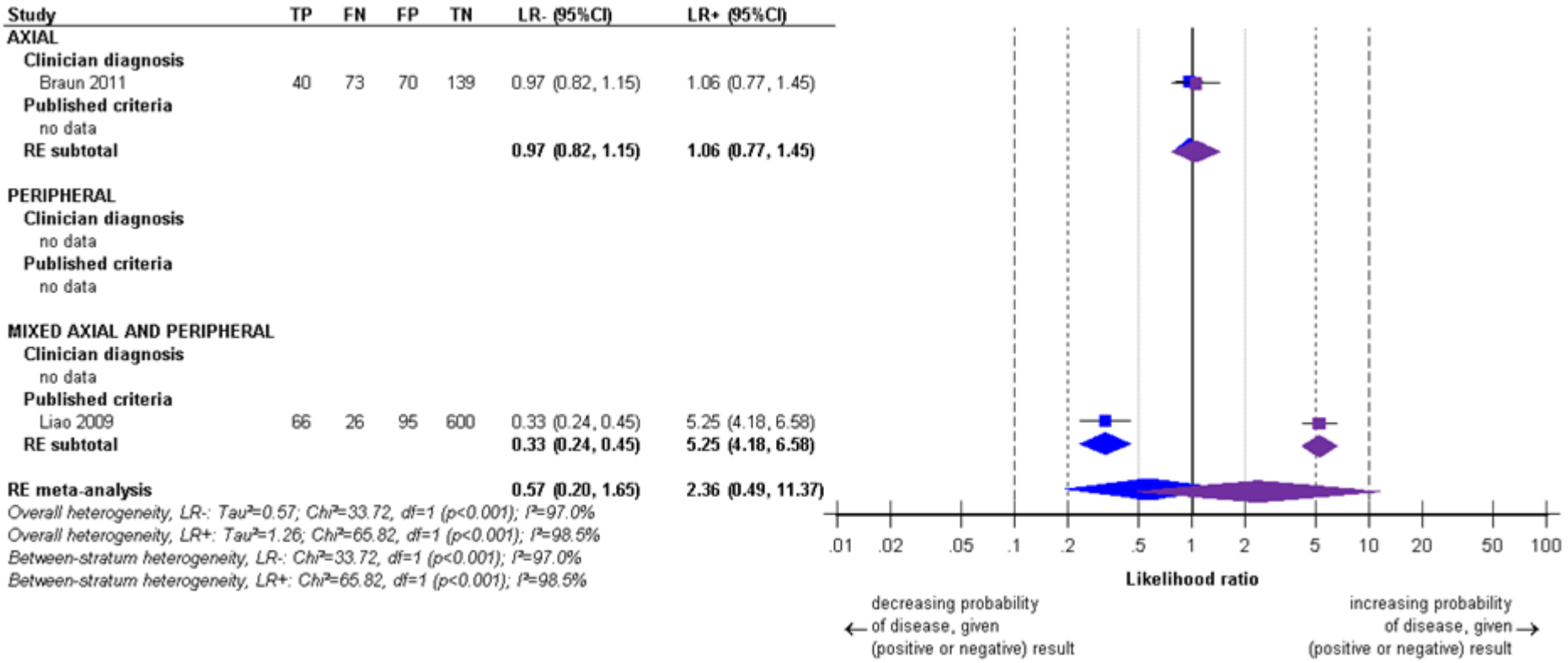


Figure 19: Morning stiffness – forest plot: likelihood ratios

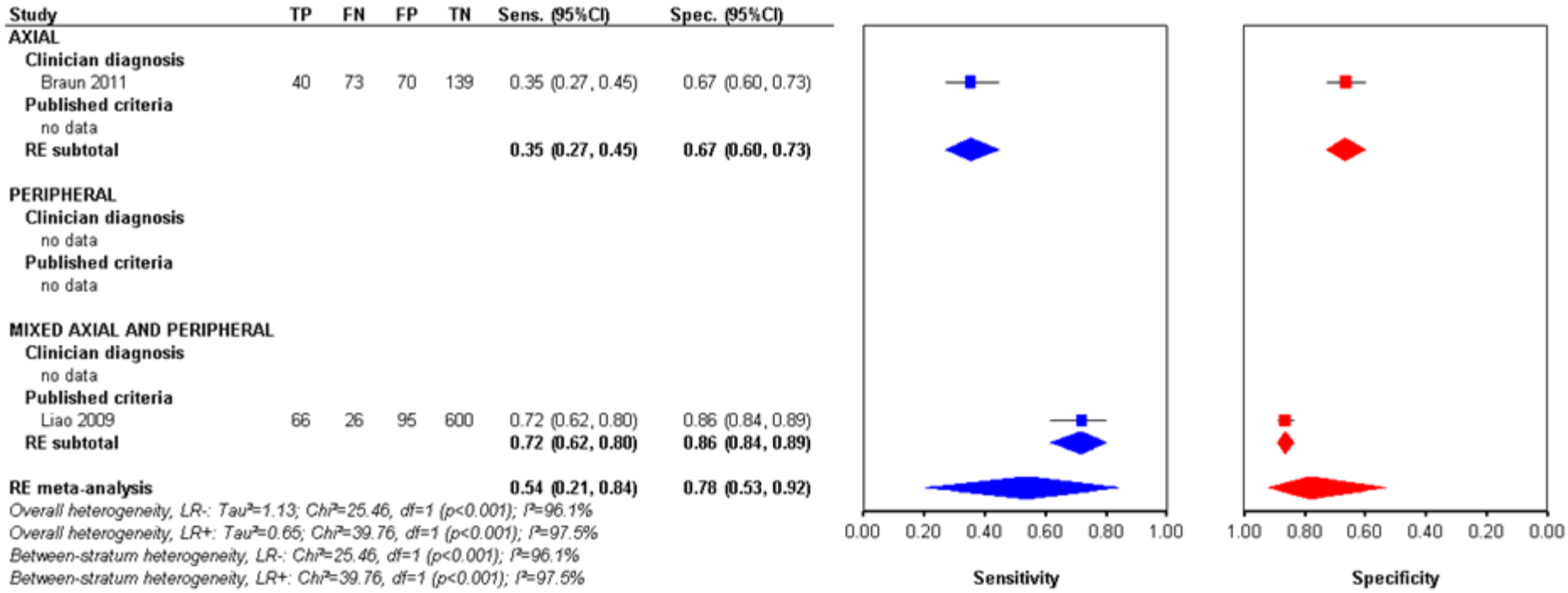


Figure 20: Morning stiffness – forest plot: sensitivity and specificity

G.1.1.11 Neck pain

Table 11: Neck pain – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	Serious[b]	Serious[c]	92	0.14 (0.04, 0.56)	VERY LOW
LR-			Serious	n/a	Serious[b]	Serious[d]		1.75 (1.36, 2.26)	VERY LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	Serious[b]	Serious[c]	92	0.14 (0.04, 0.56)	VERY LOW
LR-			Serious	n/a	Serious[b]	Serious[d]		1.75 (1.36, 2.26)	VERY LOW

[a] *Hermann 2009*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
Hermann 2009	2	28	29	33	1.75 (1.36, 2.26)	0.14 (0.04, 0.56)
RE subtotal					**1.75 (1.36, 2.26)**	**0.14 (0.04, 0.56)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**1.75 (1.36, 2.26)**	**0.14 (0.04, 0.56)**

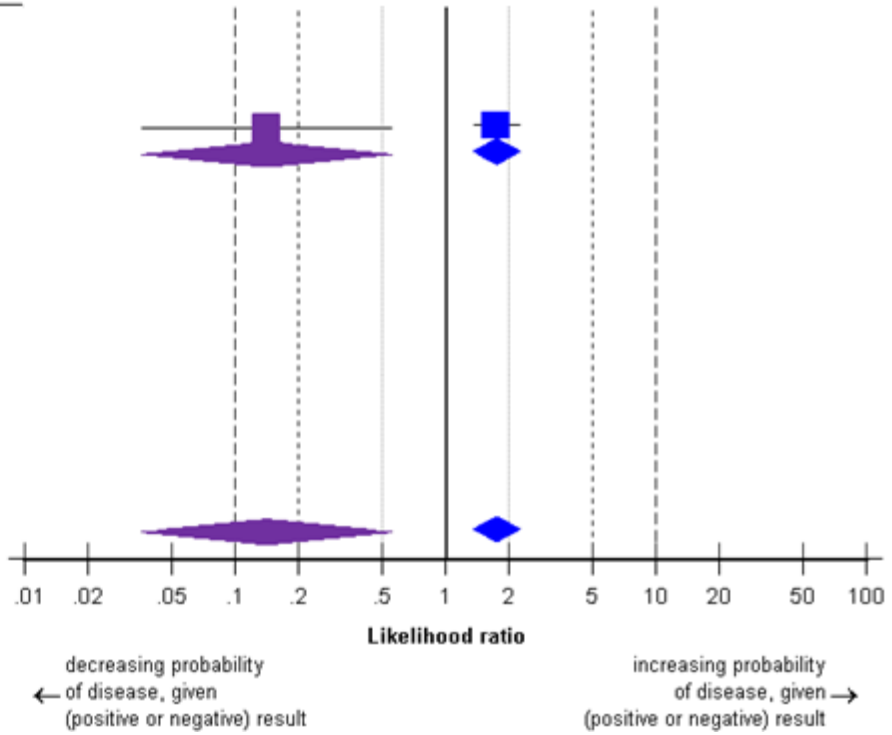


Figure 21: Neck pain – forest plot: likelihood ratios

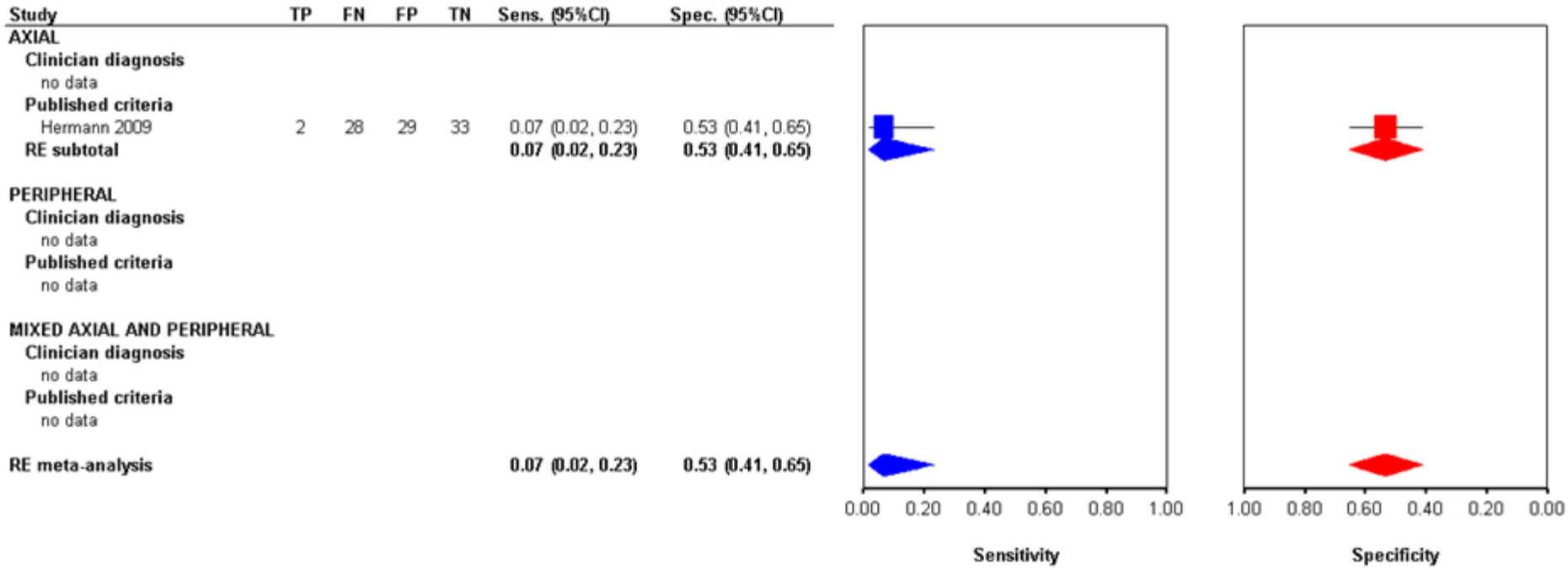


Figure 22: Neck pain – forest plot: sensitivity and specificity

G.1.1.12 Response to NSAIDs

Table 12 Response to NSAIDs – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	7 studies[a]	Cross-sectional	No serious	Serious[b]	No serious	No serious	3,145	1.52 (1.25, 1.85)	MODERATE
LR-			No serious	Serious[b]	No serious	Serious[c]		0.61 (0.48, 0.79)	LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[d]	Cross-sectional	No serious	No serious	No serious	No serious	874	1.45 (1.26, 1.67)	HIGH
LR-			No serious	No serious	No serious	No serious		0.64 (0.55, 0.75)	HIGH
ALL EVIDENCE POOLED									
LR+	9 studies[e]	Cross-sectional	No serious	Serious[b]	No serious	No serious	4,019	1.51 (1.30, 1.76)	MODERATE
LR-			No serious	Serious[b]	No serious	No serious		0.62 (0.51, 0.75)	MODERATE

[a] *Braun 2011; Poddubnyy 2011; Sieper 2013; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *I2 ≥ 50%*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[d] *D'Agostino 2011; Tomero 2014*
[e] *Braun 2011; D'Agostino 2011; Poddubnyy 2011; Sieper 2013; Tomero 2014; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

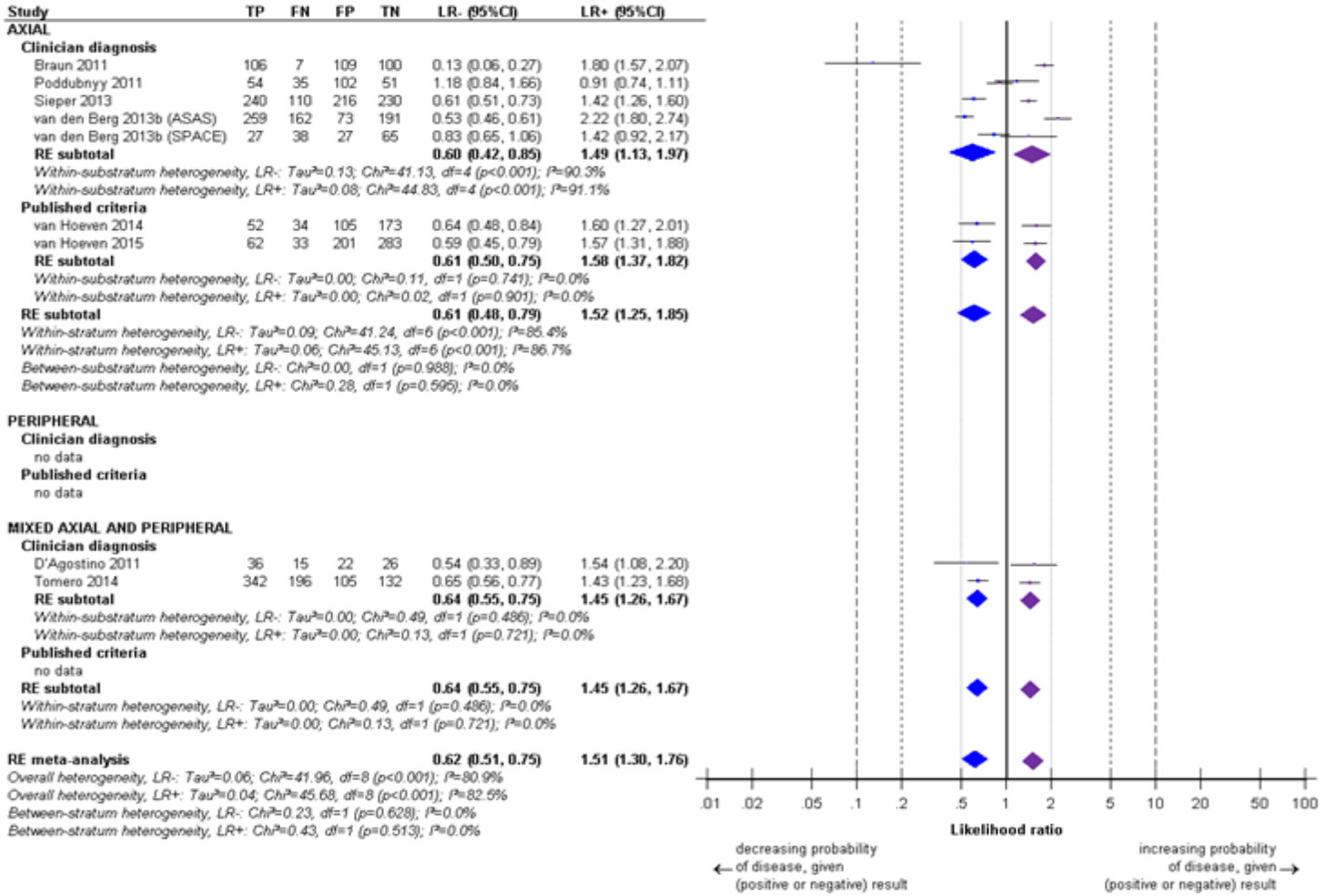


Figure 23: Response to NSAIDs – forest plot: likelihood ratios

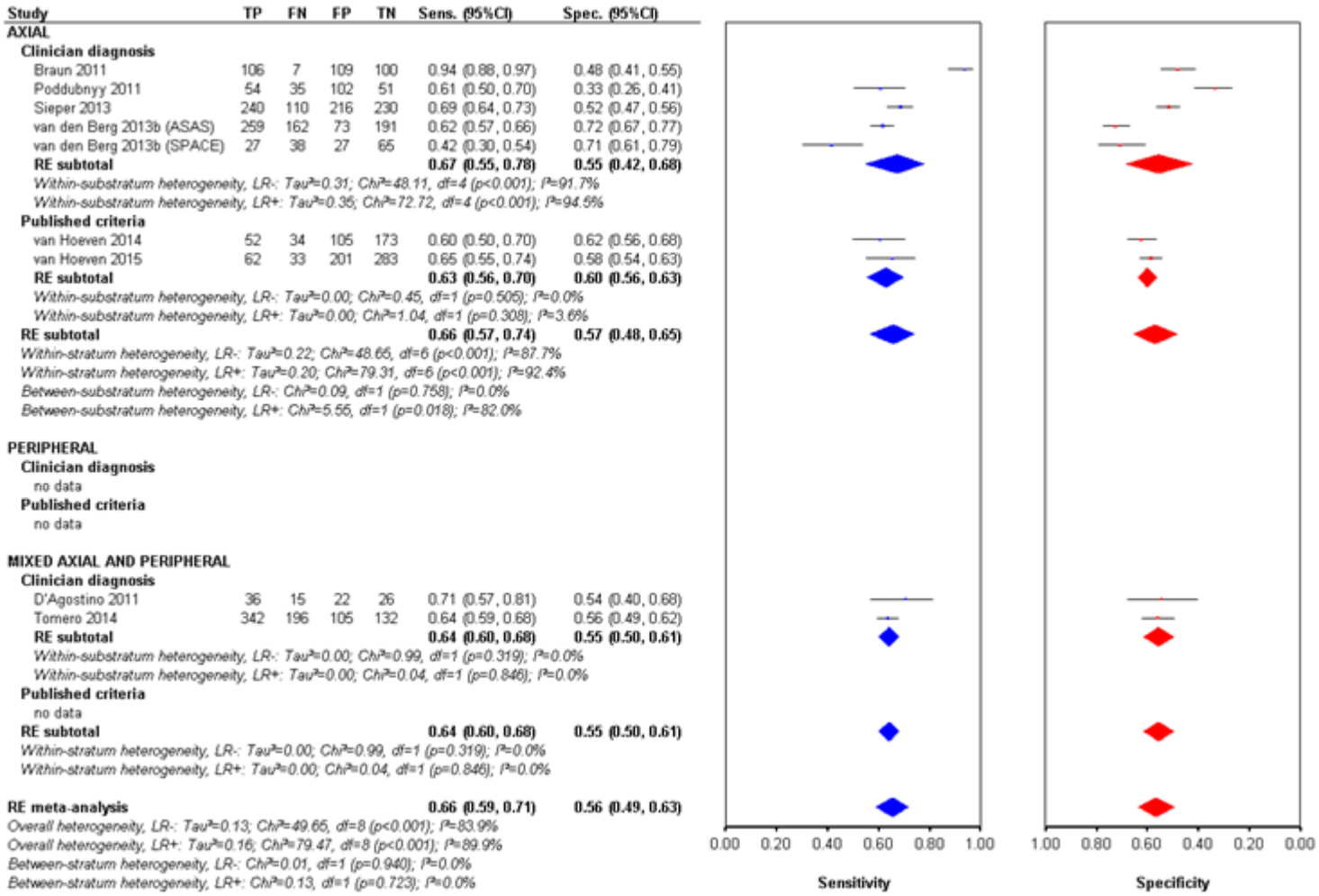


Figure 24 Response to NSAIDs – forest plot: sensitivity and specificity

G.1.1.13 Enthesitis

Table 13: Enthesitis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	7 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	3,023	1.05 (0.81, 1.37)	MODERATE
LR-			No serious	Serious[c]	Serious[b]	No serious		1.00 (0.95, 1.05)	LOW
PERIPHERAL									
LR+	4 studies[d]	Cross-sectional	No serious	Serious[c]	No serious	Serious[e]	867	3.42 (0.54, 21.57)	LOW
LR-			No serious	Serious[c]	Serious[b]	Serious[f]		0.70 (0.47, 1.03)	VERY LOW
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[g]	Cross-sectional	No serious	Serious[c]	No serious	Serious[e]	907	1.86 (1.16, 3.00)	LOW
LR-			No serious	No serious	No serious	No serious		0.79 (0.74, 0.85)	HIGH
ALL EVIDENCE POOLED									
LR+	14 studies[h]	Cross-sectional	No serious	Serious[c]	Serious[b]	No serious	4,797	1.37 (0.99, 1.89)	LOW
LR-			No serious	Serious[c]	Serious[b]	No serious		0.90 (0.82, 0.98)	LOW

[a] *Braun 2011; Dougados 2011 (DESIR); Hulsemann 1995; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *I2 ≥ 50%*
[d] *Kvien 1994; Rudwaleit 2011; Sadek 2007; You 2015*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[g] *D'Agostino 2011; Godfrin 2004 ; Tomero 2014*
[h] *Braun 2011; Dougados 2011 (DESIR); D'Agostino 2011; Godfrin 2004 ; Hulsemann 1995; Kvien 1994; Rudwaleit 2011; Sadek 2007; Tomero 2014; You 2015; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

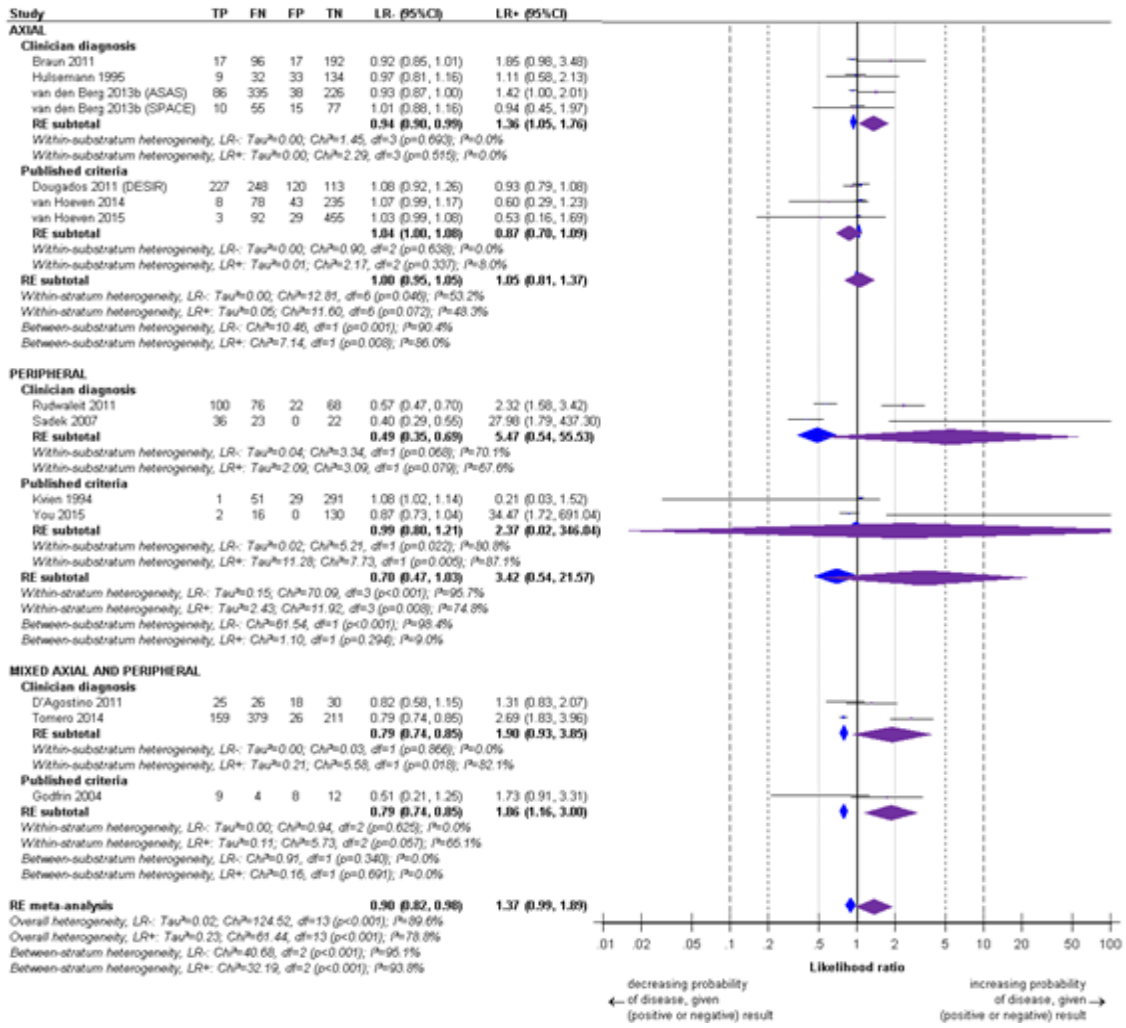


Figure 25: Enthesitis – forest plot: likelihood ratios

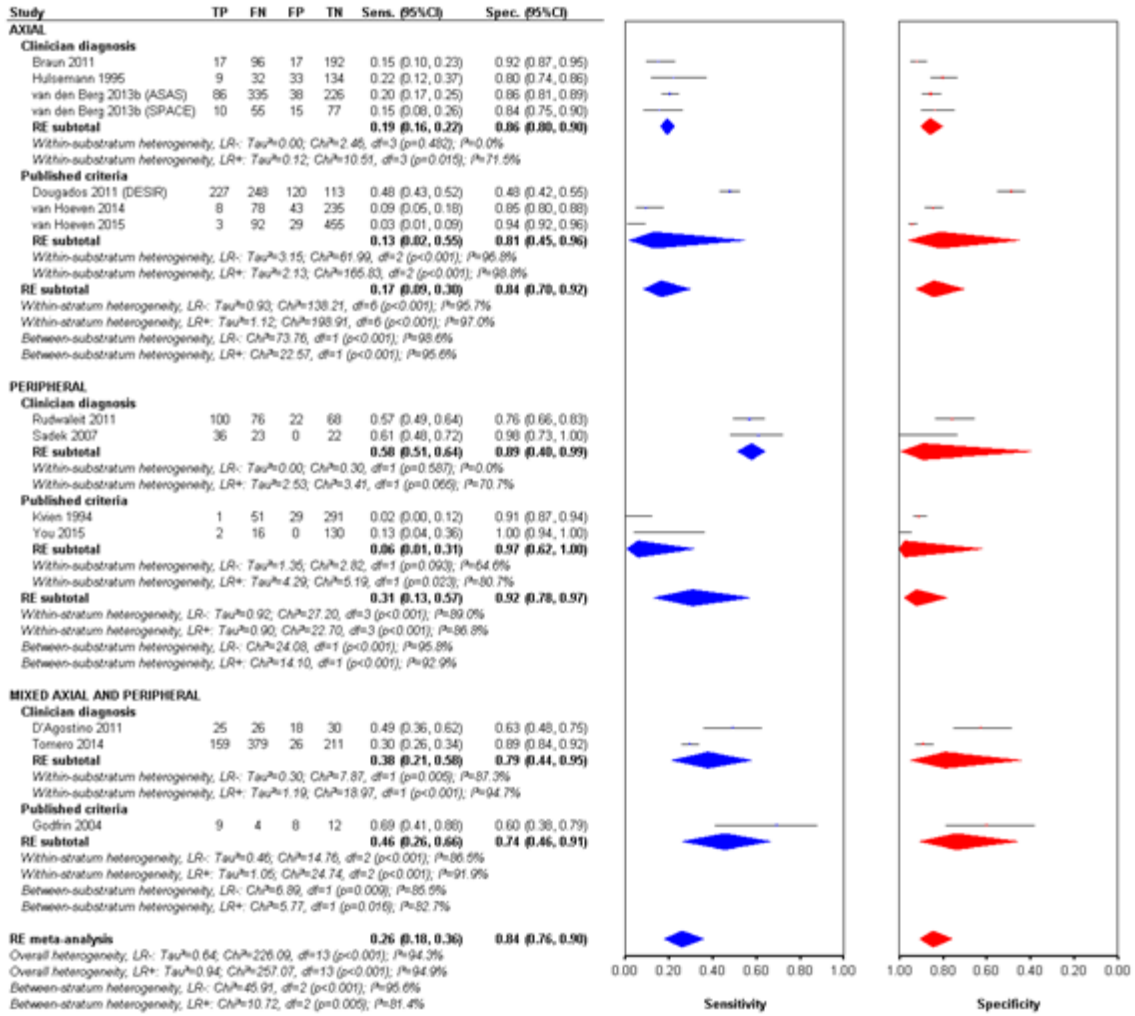


Figure 26: Enthesitis – forest plot: sensitivity and specificity

G.1.1.14 Enthesitis (heel)

Table 14: Enthesitis (heel) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	1,357	0.84 (0.71, 0.98)	MODERATE
LR-			No serious	Serious[c]	Serious[b]	No serious		1.08 (0.93, 1.24)	LOW
PERIPHERAL									
LR+	1 study[d]	Cross-sectional	No serious	n/a	No serious	Serious[e]	266	2.34 (1.32, 4.15)	MODERATE
LR-			No serious	n/a	No serious	No serious		0.79 (0.70, 0.90)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[f]	Cross-sectional	Serious[g]	No serious	Serious[b]	Serious[e]	1,562	3.45 (1.63, 7.29)	VERY LOW
LR-			Serious[g]	Serious[c]	Serious[b]	No serious		0.90 (0.79, 1.01)	VERY LOW
ALL EVIDENCE POOLED									
LR+	5 studies[h]	Cross-sectional	No serious	Serious[c]	Serious[b]	Serious[e]	3,185	1.73 (0.96, 3.15)	VERY LOW
LR-			No serious	Serious[c]	Serious[b]	No serious		0.94 (0.85, 1.04)	LOW

[a] *Dougados 2011 (DESIR); Rudwaleit 2009 (ASAS)*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *I2 ≥ 50%*
[d] *Rudwaleit 2011*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[f] *Liao 2009; Tomero 2014*
[g] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[h] *Dougados 2011 (DESIR); Liao 2009; Rudwaleit 2009 (ASAS); Rudwaleit 2011; Tomero 2014*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	66	325	46	212	1.01 (0.94, 1.09)	0.95 (0.67, 1.33)
Published criteria						
Dougados 2011 (DESIR)	183	292	111	122	1.17 (1.02, 1.35)	0.81 (0.68, 0.96)
RE subtotal					**1.08 (0.93, 1.24)**	**0.84 (0.71, 0.98)**
Within-stratum heterogeneity, LR-: Tau²=0.01; Chi²=3.37, df=1 (p=0.066); I²=70.3%						
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=0.64, df=1 (p=0.422); I²=0.0%						
Between-substratum heterogeneity, LR-: Chi²=3.37, df=1 (p=0.066); I²=70.3%						
Between-substratum heterogeneity, LR+: Chi²=0.64, df=1 (p=0.422); I²=0.0%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	55	121	12	78	0.79 (0.70, 0.90)	2.34 (1.32, 4.15)
Published criteria						
no data						
RE subtotal					**0.79 (0.70, 0.90)**	**2.34 (1.32, 4.15)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Tomero 2014	123	415	20	217	0.84 (0.79, 0.89)	2.71 (1.73, 4.24)
Published criteria						
Liao 2009	5	87	6	689	0.95 (0.91, 1.00)	6.30 (1.96, 20.22)
RE subtotal					**0.90 (0.79, 1.01)**	**3.45 (1.63, 7.29)**
Within-stratum heterogeneity, LR-: Tau²=0.01; Chi²=9.78, df=1 (p=0.002); I²=89.8%						
Within-stratum heterogeneity, LR+: Tau²=0.15; Chi²=1.75, df=1 (p=0.186); I²=42.8%						
Between-substratum heterogeneity, LR-: Chi²=9.78, df=1 (p=0.002); I²=89.8%						
Between-substratum heterogeneity, LR+: Chi²=1.75, df=1 (p=0.186); I²=42.8%						
RE meta-analysis					**0.94 (0.85, 1.04)**	**1.73 (0.96, 3.15)**
Overall heterogeneity, LR-: Tau²=0.01; Chi²=32.88, df=4 (p<0.001); I²=87.8%						
Overall heterogeneity, LR+: Tau²=0.38; Chi²=42.67, df=4 (p<0.001); I²=90.6%						
Between-stratum heterogeneity, LR-: Chi²=19.73, df=2 (p<0.001); I²=89.9%						
Between-stratum heterogeneity, LR+: Chi²=40.28, df=2 (p<0.001); I²=95.0%						

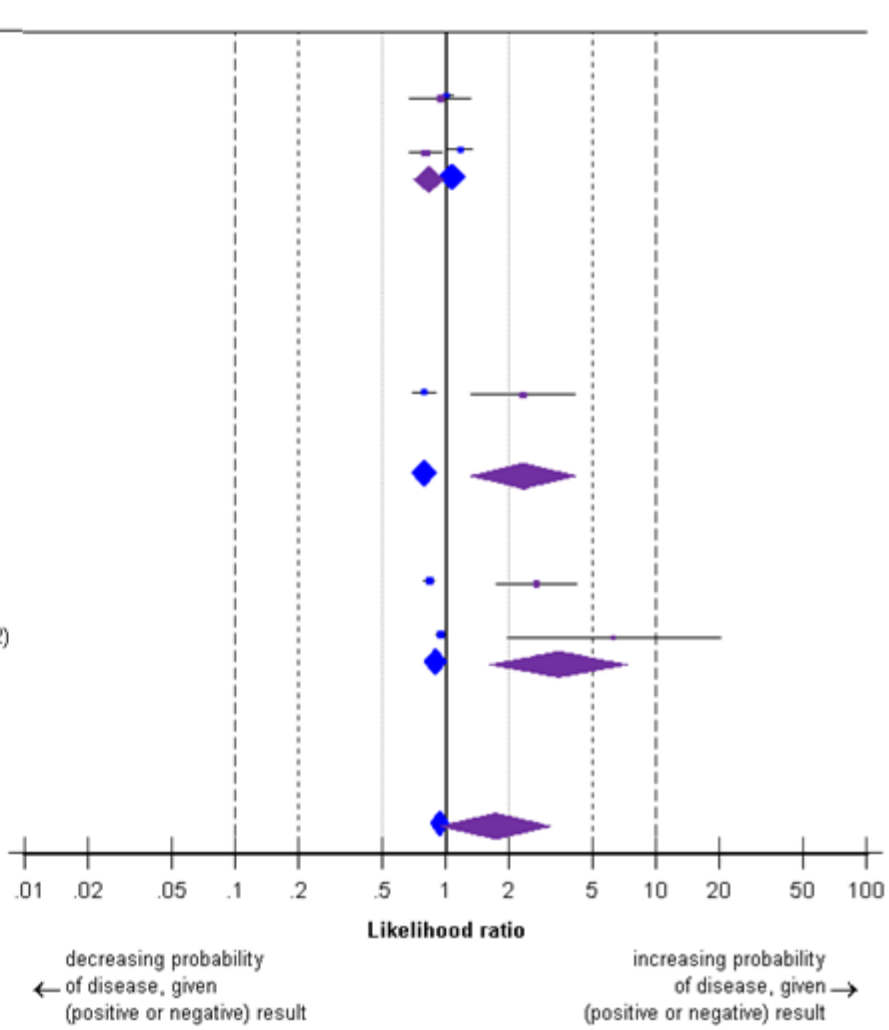


Figure 27: Enthesitis (heel) – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	66	325	46	212	0.17 (0.13, 0.21)	0.82 (0.77, 0.86)
Published criteria						
Dougados 2011 (DESIR)	183	292	111	122	0.39 (0.34, 0.43)	0.52 (0.46, 0.59)
RE subtotal					**0.26 (0.11, 0.52)**	**0.69 (0.36, 0.90)**
Within-stratum heterogeneity, LR-: Tau²=0.62; Chi²=46.83, df=1 (p<0.001); I²=97.9%						
Within-stratum heterogeneity, LR+: Tau²=1.01; Chi²=47.06, df=1 (p<0.001); I²=97.9%						
Between-substratum heterogeneity, LR-: Chi²=46.83, df=1 (p<0.001); I²=97.9%						
Between-substratum heterogeneity, LR+: Chi²=47.06, df=1 (p<0.001); I²=97.9%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	55	121	12	78	0.31 (0.25, 0.38)	0.87 (0.78, 0.92)
Published criteria						
no data						
RE subtotal					**0.31 (0.25, 0.38)**	**0.87 (0.78, 0.92)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Tomero 2014	123	415	20	217	0.23 (0.20, 0.27)	0.92 (0.87, 0.94)
Published criteria						
Liao 2009	5	87	6	689	0.05 (0.02, 0.12)	0.99 (0.98, 1.00)
RE subtotal					**0.12 (0.03, 0.41)**	**0.97 (0.77, 1.00)**
Within-stratum heterogeneity, LR-: Tau²=1.23; Chi²=12.12, df=1 (p<0.001); I²=91.7%						
Within-stratum heterogeneity, LR+: Tau²=2.67; Chi²=24.99, df=1 (p<0.001); I²=96.0%						
Between-substratum heterogeneity, LR-: Chi²=12.12, df=1 (p<0.001); I²=91.7%						
Between-substratum heterogeneity, LR+: Chi²=24.99, df=1 (p<0.001); I²=96.0%						
RE meta-analysis					**0.22 (0.14, 0.32)**	**0.89 (0.70, 0.97)**
Overall heterogeneity, LR-: Tau²=0.32; Chi²=73.59, df=4 (p<0.001); I²=94.6%						
Overall heterogeneity, LR+: Tau²=1.95; Chi²=180.12, df=4 (p<0.001); I²=97.8%						
Between-stratum heterogeneity, LR-: Chi²=14.64, df=2 (p<0.001); I²=86.3%						
Between-stratum heterogeneity, LR+: Chi²=108.07, df=2 (p<0.001); I²=98.1%						

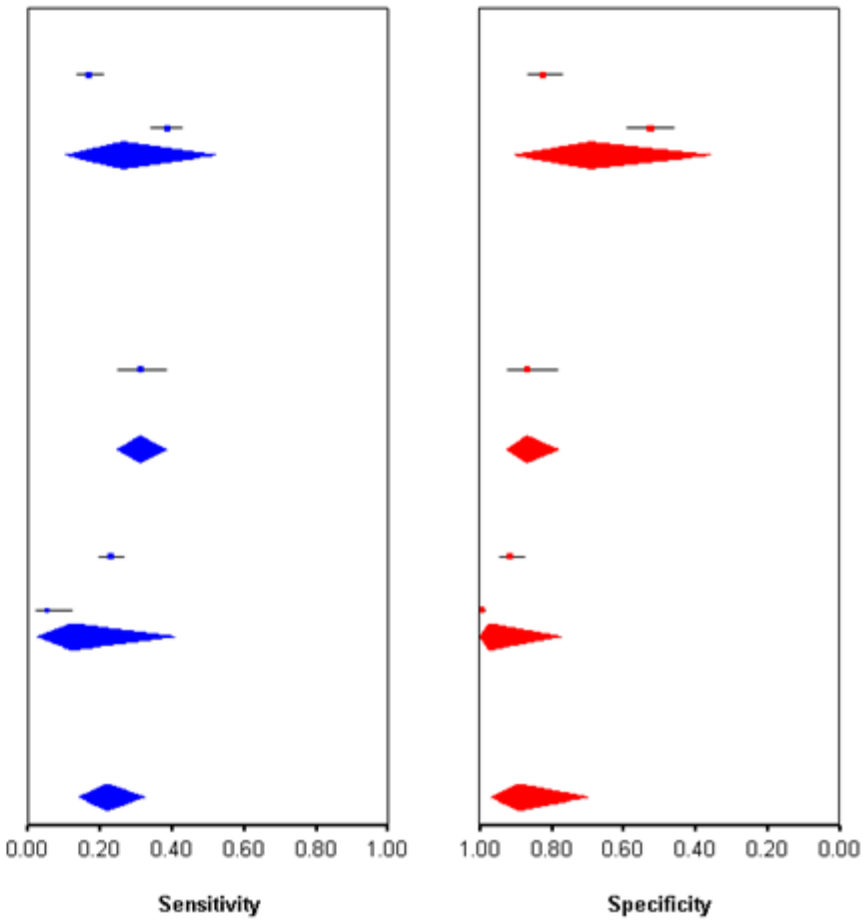


Figure 28: Enthesitis (heel) – forest plot: sensitivity and specificity

G.1.1.15 Psoriasis

Table 15: Psoriasis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	5 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	2,493	1.25 (0.88, 1.79)	MODERATE
LR-			No serious	No serious	Serious[b]	No serious		0.98 (0.96, 1.01)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	4 studies[c]	Cross-sectional	No serious	No serious	No serious	Serious[d]	1,694	2.65 (1.50, 4.68)	MODERATE
LR-			Serious[e]	Serious[f]	Serious[b]	No serious		0.92 (0.86, 0.99)	VERY LOW
ALL EVIDENCE POOLED									
LR+	9 studies[g]	Cross-sectional	No serious	Serious[f]	Serious[b]	Serious[d]	4,187	1.74 (1.16, 2.60)	VERY LOW
LR-			No serious	Serious[f]	Serious[b]	No serious		0.96 (0.93, 1.00)	LOW

[a] *Dougados 2011 (DESIR); van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *D'Agostino 2011; Godfrin 2004 ; Liao 2009; Tomero 2014*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[f] *I2 ≥ 50%*
[g] *Dougados 2011 (DESIR); D'Agostino 2011; Godfrin 2004 ; Liao 2009; Tomero 2014; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	36	385	13	251	0.96 (0.92, 1.00)	1.74 (0.94, 3.21)
van den Berg 2013b (SPACE)	10	55	6	86	0.91 (0.81, 1.02)	2.36 (0.90, 6.17)
RE subtotal					**0.96 (0.92, 0.99)**	**1.90 (1.13, 3.19)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.93, df=1 (p=0.335); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.28, df=1 (p=0.599); I^2=0.0%						
Published criteria						
Dougados 2011 (DESIR)	76	399	37	196	1.00 (0.93, 1.07)	1.01 (0.70, 1.44)
van Hoeven 2014	5	81	13	265	0.99 (0.93, 1.05)	1.24 (0.46, 3.39)
van Hoeven 2015	3	92	23	461	1.02 (0.98, 1.06)	0.66 (0.20, 2.17)
RE subtotal					**1.01 (0.98, 1.04)**	**1.00 (0.72, 1.38)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.66, df=2 (p=0.720); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.64, df=2 (p=0.726); I^2=0.0%						
RE subtotal					**0.98 (0.96, 1.01)**	**1.25 (0.88, 1.79)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=5.77, df=4 (p=0.217); I^2=30.7%						
Within-stratum heterogeneity, LR+: Tau^2=0.04; Chi^2=5.15, df=4 (p=0.272); I^2=22.4%						
Between-substratum heterogeneity, LR-: Chi^2=4.19, df=1 (p=0.041); I^2=76.1%						
Between-substratum heterogeneity, LR+: Chi^2=4.24, df=1 (p=0.040); I^2=76.4%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	17	34	8	40	0.80 (0.63, 1.01)	2.00 (0.95, 4.20)
Tomero 2014	75	463	11	226	0.90 (0.86, 0.94)	3.00 (1.63, 5.55)
RE subtotal					**0.90 (0.86, 0.94)**	**2.55 (1.59, 4.09)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=1.00, df=1 (p=0.316); I^2=0.4%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.68, df=1 (p=0.408); I^2=0.0%						
Published criteria						
Godfrin 2004	6	7	5	15	0.72 (0.41, 1.26)	1.85 (0.71, 4.82)
Liao 2009	3	89	1	694	0.97 (0.93, 1.01)	22.66 (2.38, 215.60)
RE subtotal					**0.96 (0.85, 1.08)**	**5.21 (0.46, 58.68)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=1.08, df=1 (p=0.298); I^2=7.6%						
Within-substratum heterogeneity, LR+: Tau^2=2.36; Chi^2=4.03, df=1 (p=0.045); I^2=75.2%						
RE subtotal					**0.92 (0.86, 0.99)**	**2.65 (1.50, 4.68)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=8.45, df=3 (p=0.038); I^2=64.5%						
Within-stratum heterogeneity, LR+: Tau^2=0.12; Chi^2=4.73, df=3 (p=0.193); I^2=36.6%						
Between-substratum heterogeneity, LR-: Chi^2=6.36, df=1 (p=0.012); I^2=84.3%						
Between-substratum heterogeneity, LR+: Chi^2=0.02, df=1 (p=0.901); I^2=0.0%						
RE meta-analysis					**0.96 (0.93, 1.00)**	**1.74 (1.16, 2.60)**
Overall heterogeneity, LR-: Tau^2=0.00; Chi^2=21.35, df=8 (p=0.006); I^2=62.5%						
Overall heterogeneity, LR+: Tau^2=0.20; Chi^2=18.95, df=8 (p=0.015); I^2=57.8%						
Between-stratum heterogeneity, LR-: Chi^2=7.13, df=1 (p=0.008); I^2=86.0%						
Between-stratum heterogeneity, LR+: Chi^2=9.07, df=1 (p=0.003); I^2=89.0%						

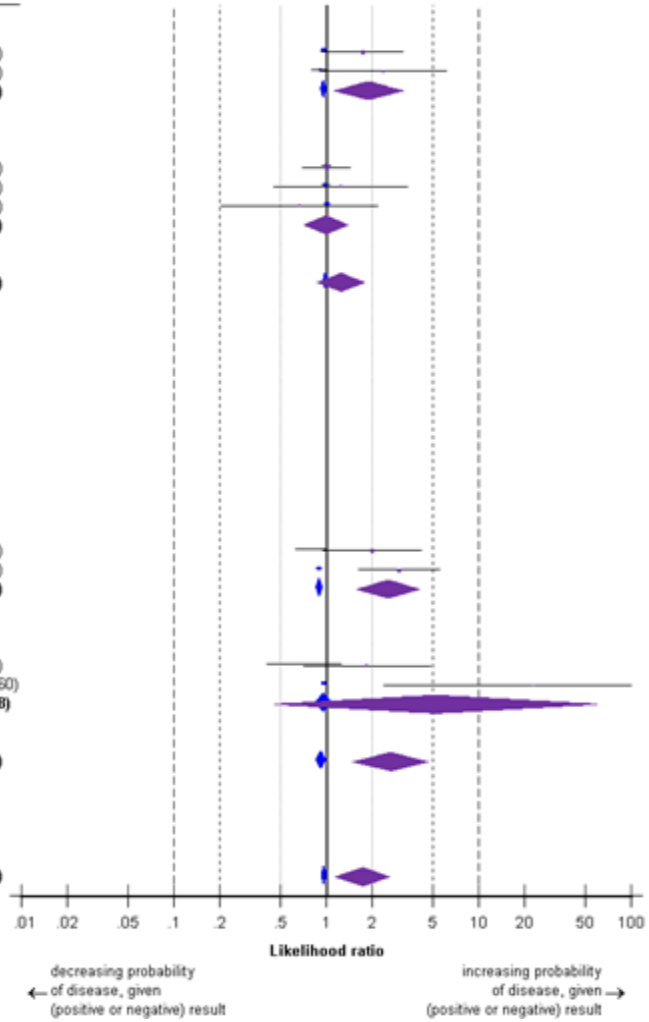


Figure 29: Psoriasis – forest plot: likelihood ratios

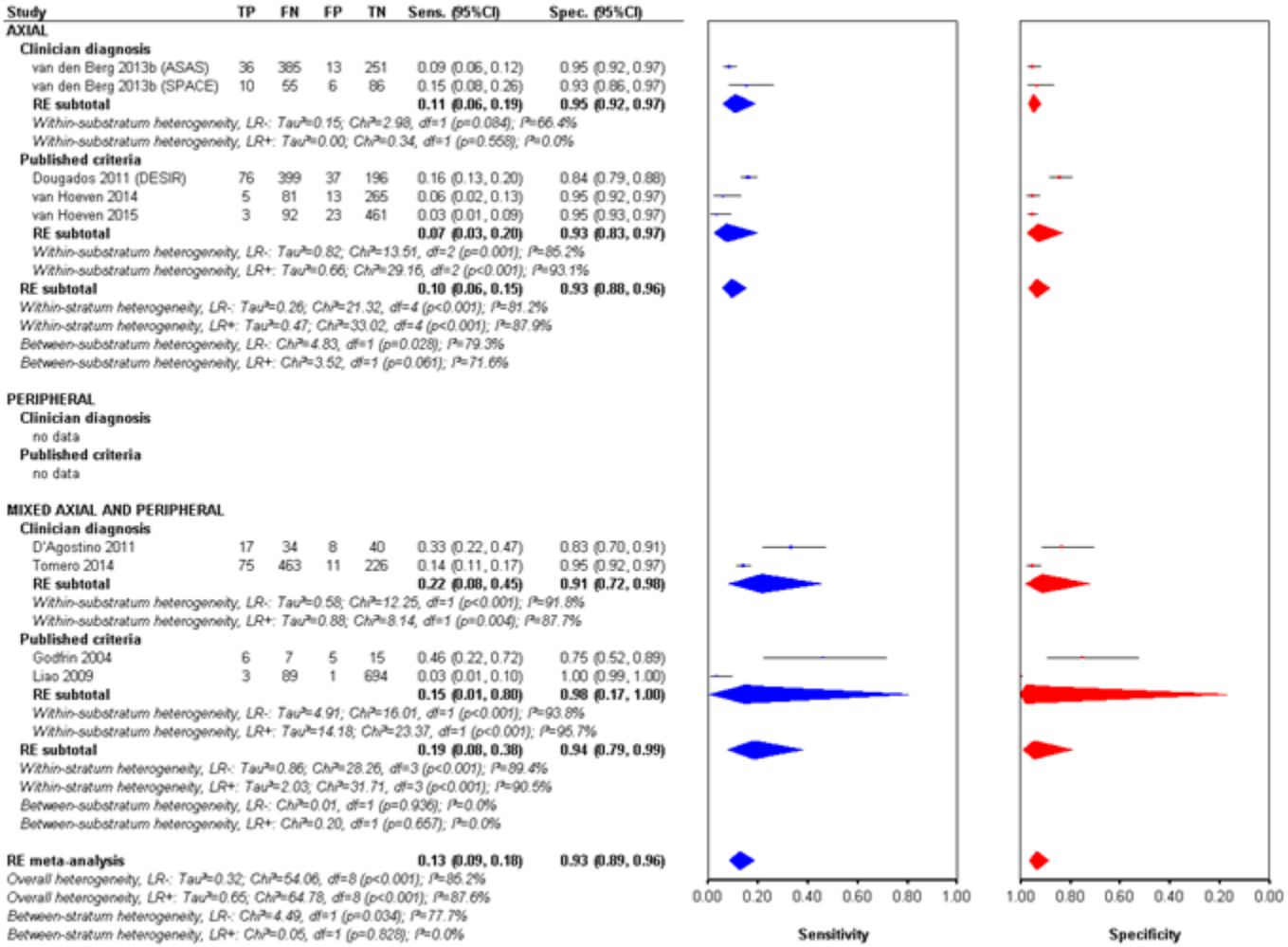


Figure 30 Psoriasis – forest plot: sensitivity and specificity

G.1.1.16 Uveitis

Table 16: Uveitis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	4 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	Serious[c]	1,914	1.58 (1.12, 2.22)	LOW
LR-			No serious	No serious	Serious[b]	No serious		0.97 (0.94, 0.99)	MODERATE
PERIPHERAL									
LR+	5 studies[d]	Cross-sectional	No serious	Serious[e]	Serious[b]	Serious[c]	1,038	3.66 (0.97, 13.80)	VERY LOW
LR-			Serious[f]	Serious[e]	Serious[b]	No serious		0.93 (0.85, 1.02)	VERY LOW
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[g]	Cross-sectional	No serious	No serious	No serious	Serious[c]	935	3.93 (1.16, 13.30)	MODERATE
LR-			No serious	No serious	No serious	No serious		0.95 (0.87, 1.03)	HIGH
ALL EVIDENCE POOLED									
LR+	11 studies[h]	Cross-sectional	No serious	No serious	Serious[b]	Serious[c]	3,887	2.34 (1.51, 3.63)	LOW
LR-			No serious	No serious	Serious[b]	No serious		0.96 (0.94, 0.99)	MODERATE

[a] *Dougados 2011 (DESIR); van Hoeven 2014; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[d] *Kvien 1994; Mattila 1998; Munch 1985; Mäki-Ikola 1991; Rigby 1993*
[e] *I2 ≥ 50%*
[f] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[g] *Salvarini 2001; Tomero 2014*
[h] *Dougados 2011 (DESIR); Kvien 1994; Mattila 1998; Munch 1985; Mäki-Ikola 1991; Rigby 1993; Salvarini 2001; Tomero 2014; van Hoeven 2014; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

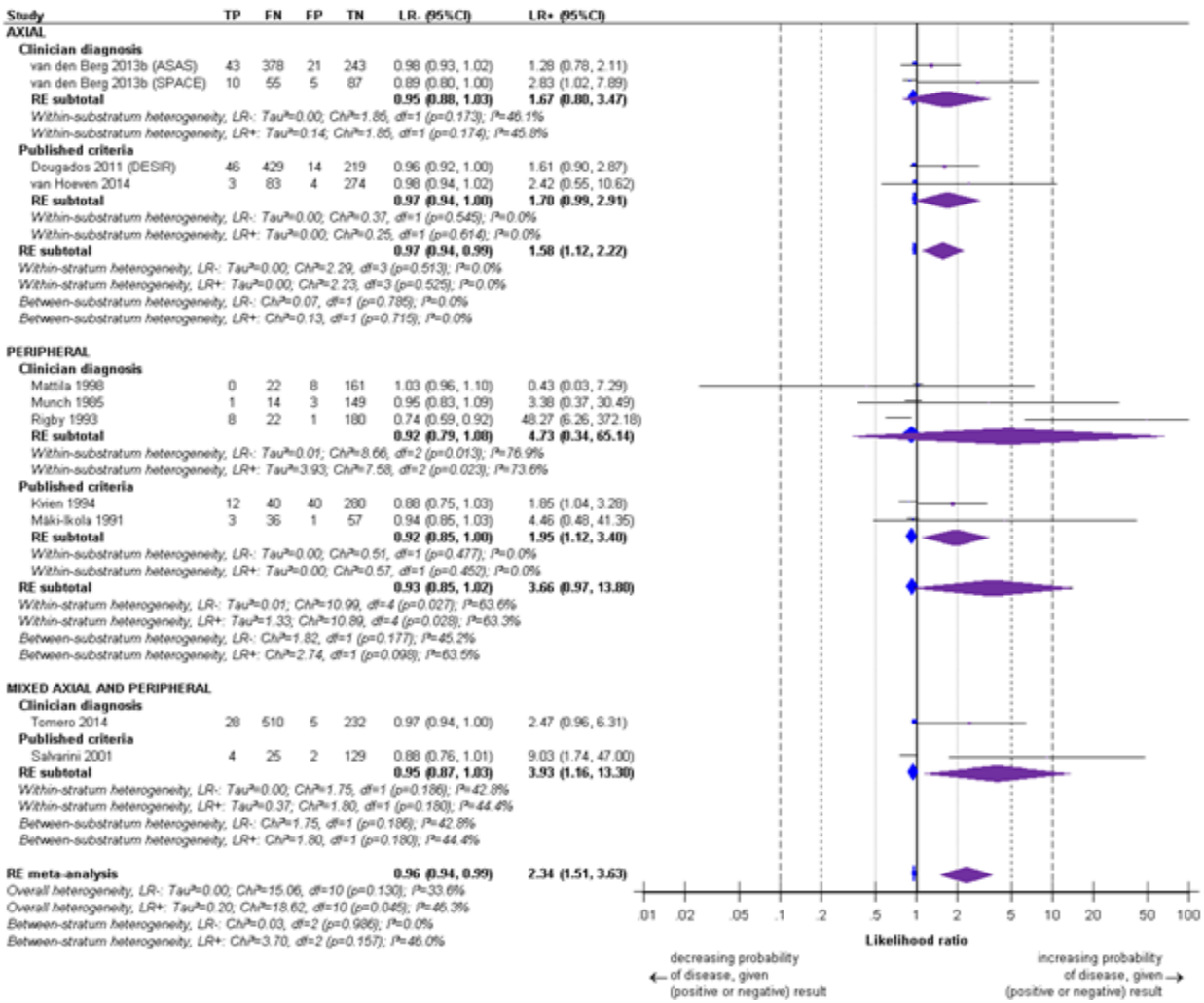

Study TP FN FP TN LR- (95%CI) LR+ (95%CI)
AXIAL
Clinician diagnosis
van den Berg 2013b (ASAS) 43 378 21 243 0.98 (0.93, 1.02) 1.28 (0.78, 2.11)
van den Berg 2013b (SPACE) 10 55 5 87 0.89 (0.80, 1.00) 2.83 (1.02, 7.89)
RE subtotal 0.95 (0.88, 1.03) 1.67 (0.80, 3.47)
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=1.86, df=1 (p=0.173); I²=46.1%
Within-substratum heterogeneity, LR+: Tau²=0.14; Chi²=1.86, df=1 (p=0.174); I²=45.8%
Published criteria
Dougados 2011 (DESIR) 46 429 14 219 0.96 (0.92, 1.00) 1.61 (0.90, 2.87)
van Hoeven 2014 3 83 4 274 0.98 (0.94, 1.02) 2.42 (0.55, 10.62)
RE subtotal 0.97 (0.94, 1.00) 1.70 (0.99, 2.91)
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.37, df=1 (p=0.545); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.25, df=1 (p=0.614); I²=0.0%
RE subtotal 0.97 (0.94, 0.99) 1.58 (1.12, 2.22)
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=2.29, df=3 (p=0.513); I²=0.0%
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=2.23, df=3 (p=0.525); I²=0.0%
Between-substratum heterogeneity, LR-: Chi²=0.07, df=1 (p=0.785); I²=0.0%
Between-substratum heterogeneity, LR+: Chi²=0.13, df=1 (p=0.715); I²=0.0%
PERIPHERAL
Clinician diagnosis
Mattila 1998 0 22 8 161 1.03 (0.96, 1.10) 0.43 (0.03, 7.29)
Munch 1995 1 14 3 149 0.95 (0.83, 1.09) 3.38 (0.37, 30.49)
Rigby 1993 8 22 1 180 0.74 (0.59, 0.92) 48.27 (6.26, 372.18)
RE subtotal 0.92 (0.79, 1.08) 4.73 (0.34, 65.14)
Within-substratum heterogeneity, LR-: Tau²=0.01; Chi²=8.66, df=2 (p=0.013); I²=76.9%
Within-substratum heterogeneity, LR+: Tau²=3.93; Chi²=7.58, df=2 (p=0.023); I²=73.6%
Published criteria
Kvien 1994 12 40 40 280 0.88 (0.75, 1.03) 1.85 (1.04, 3.28)
Mäki-Ikola 1991 3 36 1 57 0.94 (0.85, 1.03) 4.46 (0.48, 41.35)
RE subtotal 0.92 (0.85, 1.00) 1.95 (1.12, 3.40)
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.51, df=1 (p=0.477); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.57, df=1 (p=0.452); I²=0.0%
RE subtotal 0.93 (0.85, 1.02) 3.66 (0.97, 13.80)
Within-stratum heterogeneity, LR-: Tau²=0.01; Chi²=10.99, df=4 (p=0.027); I²=63.6%
Within-stratum heterogeneity, LR+: Tau²=1.33; Chi²=10.89, df=4 (p=0.028); I²=63.3%
Between-substratum heterogeneity, LR-: Chi²=1.82, df=1 (p=0.177); I²=45.2%
Between-substratum heterogeneity, LR+: Chi²=2.74, df=1 (p=0.098); I²=63.5%
MIXED AXIAL AND PERIPHERAL
Clinician diagnosis
Tornero 2014 28 510 5 232 0.97 (0.94, 1.00) 2.47 (0.96, 6.31)
Published criteria
Salvarini 2001 4 25 2 129 0.88 (0.76, 1.01) 9.03 (1.74, 47.00)
RE subtotal 0.95 (0.87, 1.03) 3.93 (1.16, 13.30)
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=1.75, df=1 (p=0.186); I²=42.8%
Within-stratum heterogeneity, LR+: Tau²=0.37; Chi²=1.80, df=1 (p=0.180); I²=44.4%
Between-substratum heterogeneity, LR-: Chi²=1.75, df=1 (p=0.186); I²=42.8%
Between-substratum heterogeneity, LR+: Chi²=1.80, df=1 (p=0.180); I²=44.4%
RE meta-analysis 0.96 (0.94, 0.99) 2.34 (1.51, 3.63)
Overall heterogeneity, LR-: Tau²=0.00; Chi²=15.06, df=10 (p=0.130); I²=33.6%
Overall heterogeneity, LR+: Tau²=0.20; Chi²=18.62, df=10 (p=0.045); I²=46.3%
Between-stratum heterogeneity, LR-: Chi²=0.03, df=2 (p=0.986); I²=0.0%
Between-stratum heterogeneity, LR+: Chi²=3.70, df=2 (p=0.157); I²=46.0%
.01 .02 .05 .1 .2 .5 1 2 5 10 20 50 100
Likelihood ratio
decreasing probability of disease, given (positive or negative) result
increasing probability of disease, given (positive or negative) result

Figure 31: Uveitis – forest plot: likelihood ratios

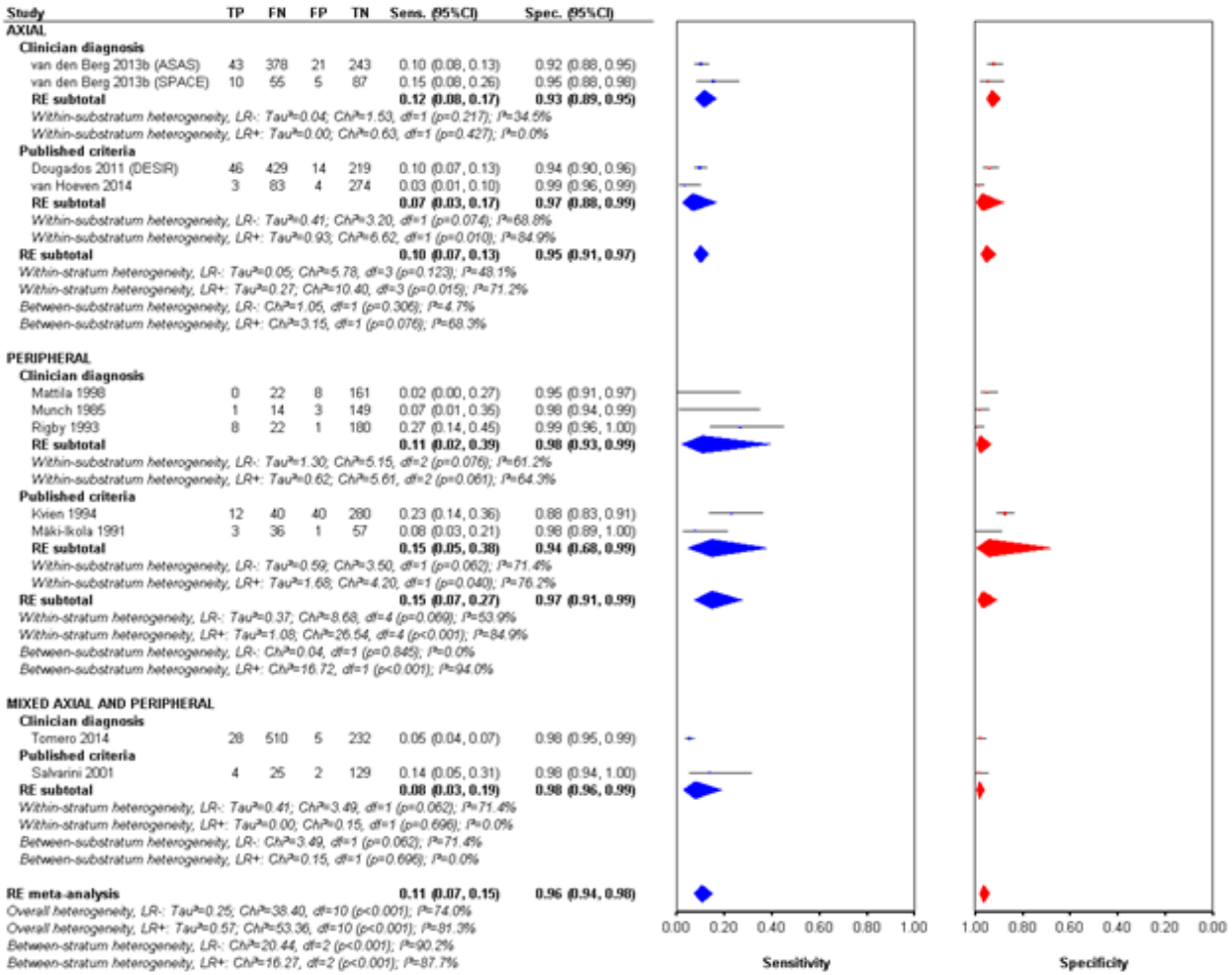


Figure 32: History of uveitis

Table 17: History of uveitis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	579	1.42 (0.54, 3.72)	LOW
LR-			No serious	n/a	Serious[b]	No serious		0.98 (0.94, 1.03)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	579	1.42 (0.54, 3.72)	LOW
LR-			No serious	n/a	Serious[b]	No serious		0.98 (0.94, 1.03)	MODERATE

[a] *van Hoeven 2015*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
van Hoeven 2015	5	90	18	466	0.98 (0.94, 1.03)	1.42 (0.54, 3.72)
RE subtotal					**0.98 (0.94, 1.03)**	**1.42 (0.54, 3.72)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.98 (0.94, 1.03)**	**1.42 (0.54, 3.72)**

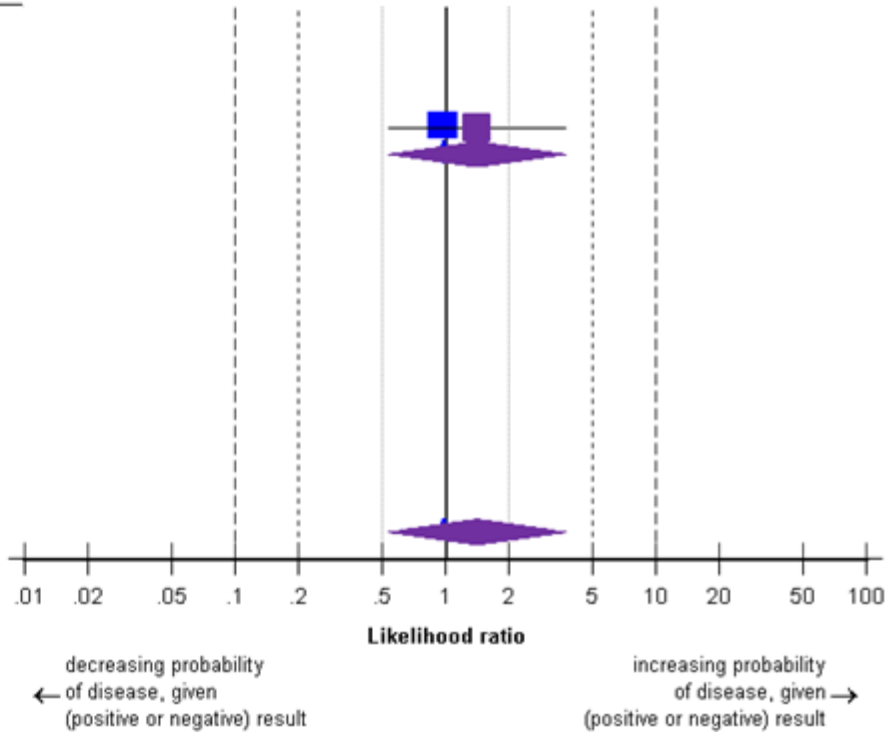


Figure 33: History of uveitis – forest plot: likelihood ratios

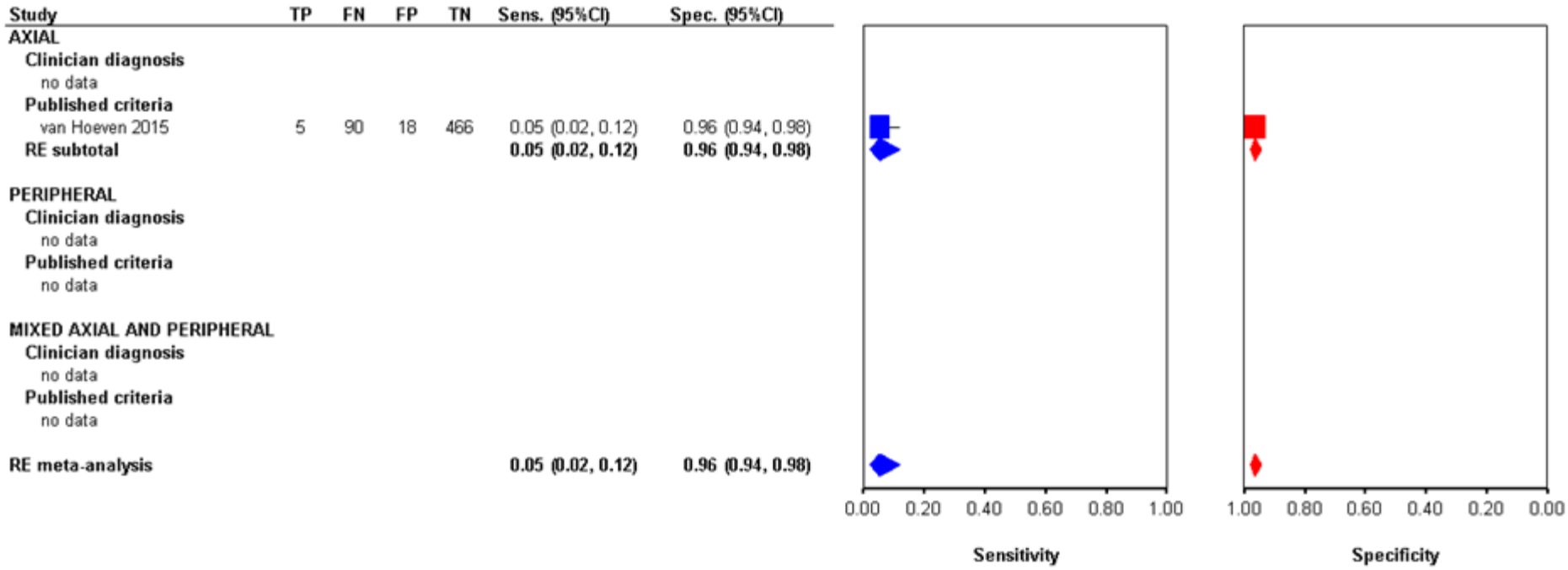


Figure 34: History of uveitis – forest plot: sensitivity and specificity

G.1.1.17 Inflammatory bowel disease

Table 18 Inflammatory bowel disease – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	4 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	2,129	1.16 (0.68, 1.97)	MODERATE
LR-			No serious	No serious	Serious[b]	No serious		1.00 (0.98, 1.01)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[c]	Cross-sectional	No serious	No serious	No serious	Serious[d]	1,661	1.69 (0.83, 3.43)	MODERATE
LR-			Serious[e]	No serious	Serious[b]	No serious		0.99 (0.98, 1.01)	LOW
ALL EVIDENCE POOLED									
LR+	7 studies[f]	Cross-sectional	No serious	No serious	Serious[b]	Serious[d]	3,790	1.33 (0.86, 2.03)	LOW
LR-			Serious[e]	No serious	Serious[b]	No serious		0.99 (0.98, 1.00)	LOW

[a] *Dougados 2011 (DESIR); van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *D'Agostino 2011; Liao 2009; Tomero 2014*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[f] *Dougados 2011 (DESIR); D'Agostino 2011; Liao 2009; Tomero 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	14	407	4	260	0.98 (0.96, 1.00)	2.19 (0.73, 6.60)
van den Berg 2013b (SPACE)	4	61	5	87	0.99 (0.92, 1.07)	1.13 (0.32, 4.06)
RE subtotal					**0.98 (0.96, 1.00)**	**1.65 (0.72, 3.81)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.07, df=1 (p=0.795); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.59, df=1 (p=0.441); I²=0.0%						
Published criteria						
Dougados 2011 (DESIR)	20	455	10	223	1.00 (0.97, 1.03)	0.98 (0.47, 2.06)
van Hoeven 2015	1	94	11	473	1.01 (0.99, 1.04)	0.46 (0.06, 3.55)
RE subtotal					**1.01 (0.99, 1.03)**	**0.90 (0.45, 1.80)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.30, df=1 (p=0.584); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.46, df=1 (p=0.497); I²=0.0%						
RE subtotal					**1.00 (0.98, 1.01)**	**1.16 (0.68, 1.97)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=3.28, df=3 (p=0.351); I²=8.4%						
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=2.27, df=3 (p=0.519); I²=0.0%						
Between-substratum heterogeneity, LR-: Chi²=2.91, df=1 (p=0.088); I²=65.6%						
Between-substratum heterogeneity, LR+: Chi²=1.21, df=1 (p=0.270); I²=17.7%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	4	47	1	47	0.94 (0.86, 1.03)	3.76 (0.44, 32.50)
Tomero 2014	27	511	8	229	0.98 (0.95, 1.01)	1.49 (0.69, 3.22)
RE subtotal					**0.98 (0.95, 1.01)**	**1.65 (0.80, 3.43)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.80, df=1 (p=0.371); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.63, df=1 (p=0.427); I²=0.0%						
Published criteria						
Liao 2009	0	92	1	694	1.00 (0.98, 1.01)	2.49 (0.10, 60.79)
RE subtotal					**0.99 (0.98, 1.01)**	**1.69 (0.83, 3.43)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=2.01, df=2 (p=0.366); I²=0.6%						
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=0.69, df=2 (p=0.707); I²=0.0%						
Between-substratum heterogeneity, LR-: Chi²=1.21, df=1 (p=0.271); I²=17.3%						
Between-substratum heterogeneity, LR+: Chi²=0.06, df=1 (p=0.806); I²=0.0%						
RE meta-analysis					**0.99 (0.98, 1.00)**	**1.33 (0.86, 2.03)**
Overall heterogeneity, LR-: Tau²=0.00; Chi²=5.44, df=6 (p=0.489); I²=0.0%						
Overall heterogeneity, LR+: Tau²=0.00; Chi²=3.66, df=6 (p=0.723); I²=0.0%						
Between-stratum heterogeneity, LR-: Chi²=0.15, df=1 (p=0.695); I²=0.0%						
Between-stratum heterogeneity, LR+: Chi²=0.70, df=1 (p=0.404); I²=0.0%						

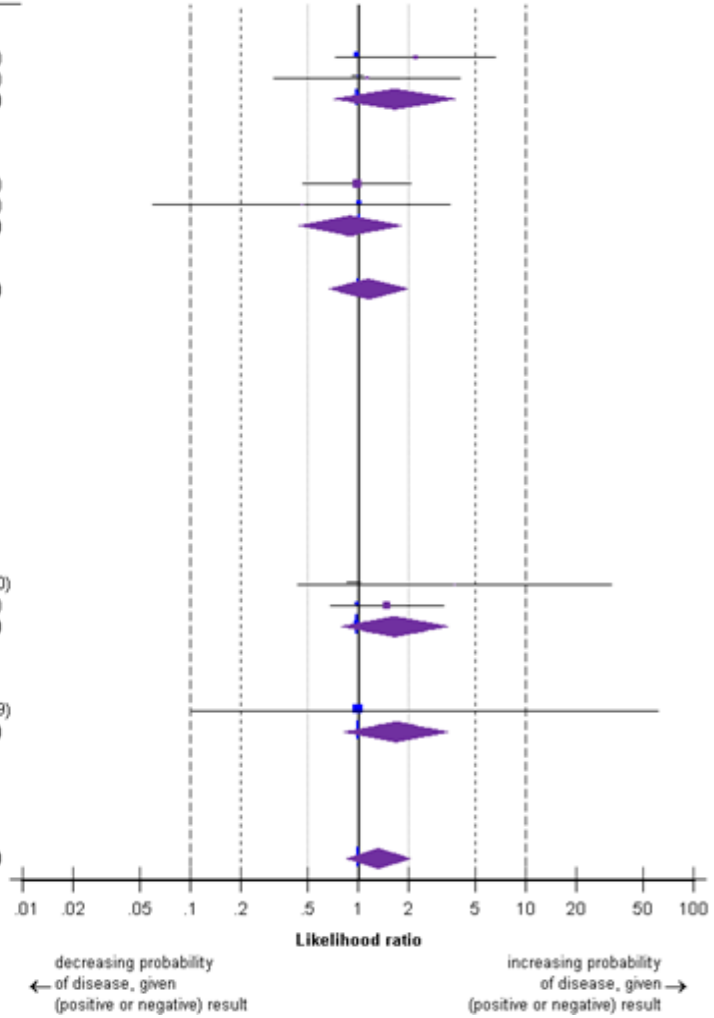


Figure 35: Inflammatory bowel disease – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	14	407	4	260	0.03 (0.02, 0.06)	0.98 (0.96, 0.99)
van den Berg 2013b (SPACE)	4	61	5	87	0.06 (0.02, 0.15)	0.95 (0.88, 0.98)
RE subtotal					**0.04 (0.02, 0.07)**	**0.97 (0.90, 0.99)**
Within-substratum heterogeneity, LR-: Tau²=0.04; Chi²=1.22, df=1 (p=0.269); I²=18.3%						
Within-substratum heterogeneity, LR+: Tau²=0.64; Chi²=3.73, df=1 (p=0.053); I²=73.2%						
Published criteria						
Dougados 2011 (DESIR)	20	455	10	223	0.04 (0.03, 0.06)	0.96 (0.92, 0.98)
van Hoeven 2015	1	94	11	473	0.01 (0.00, 0.07)	0.98 (0.96, 0.99)
RE subtotal					**0.03 (0.01, 0.09)**	**0.97 (0.94, 0.98)**
Within-substratum heterogeneity, LR-: Tau²=0.47; Chi²=1.89, df=1 (p=0.169); I²=47.2%						
Within-substratum heterogeneity, LR+: Tau²=0.12; Chi²=2.18, df=1 (p=0.140); I²=54.2%						
RE subtotal					**0.04 (0.03, 0.05)**	**0.97 (0.95, 0.98)**
Within-stratum heterogeneity, LR-: Tau²=0.01; Chi²=3.13, df=3 (p=0.372); I²=4.1%						
Within-stratum heterogeneity, LR+: Tau²=0.14; Chi²=5.92, df=3 (p=0.116); I²=49.3%						
Between-substratum heterogeneity, LR-: Chi²=0.01, df=1 (p=0.914); I²=0.0%						
Between-substratum heterogeneity, LR+: Chi²=0.00, df=1 (p=0.993); I²=0.0%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	4	47	1	47	0.08 (0.03, 0.19)	0.98 (0.87, 1.00)
Tomero 2014	27	511	8	229	0.05 (0.03, 0.07)	0.97 (0.93, 0.98)
RE subtotal					**0.05 (0.04, 0.07)**	**0.97 (0.94, 0.98)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.73, df=1 (p=0.392); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.21, df=1 (p=0.644); I²=0.0%						
Published criteria						
Liao 2009	0	92	1	694	0.01 (0.00, 0.08)	1.00 (0.99, 1.00)
RE subtotal					**0.05 (0.02, 0.10)**	**0.99 (0.93, 1.00)**
Within-stratum heterogeneity, LR-: Tau²=0.19; Chi²=3.41, df=2 (p=0.182); I²=41.3%						
Within-stratum heterogeneity, LR+: Tau²=1.91; Chi²=9.72, df=2 (p=0.008); I²=79.4%						
Between-substratum heterogeneity, LR-: Chi²=2.68, df=1 (p=0.102); I²=62.6%						
Between-substratum heterogeneity, LR+: Chi²=9.50, df=1 (p=0.002); I²=89.5%						
RE meta-analysis					**0.04 (0.03, 0.06)**	**0.98 (0.96, 0.99)**
Overall heterogeneity, LR-: Tau²=0.04; Chi²=7.94, df=6 (p=0.242); I²=24.5%						
Overall heterogeneity, LR+: Tau²=0.34; Chi²=16.59, df=6 (p=0.011); I²=63.8%						
Between-stratum heterogeneity, LR-: Chi²=1.41, df=1 (p=0.236); I²=28.8%						
Between-stratum heterogeneity, LR+: Chi²=0.96, df=1 (p=0.327); I²=0.0%						

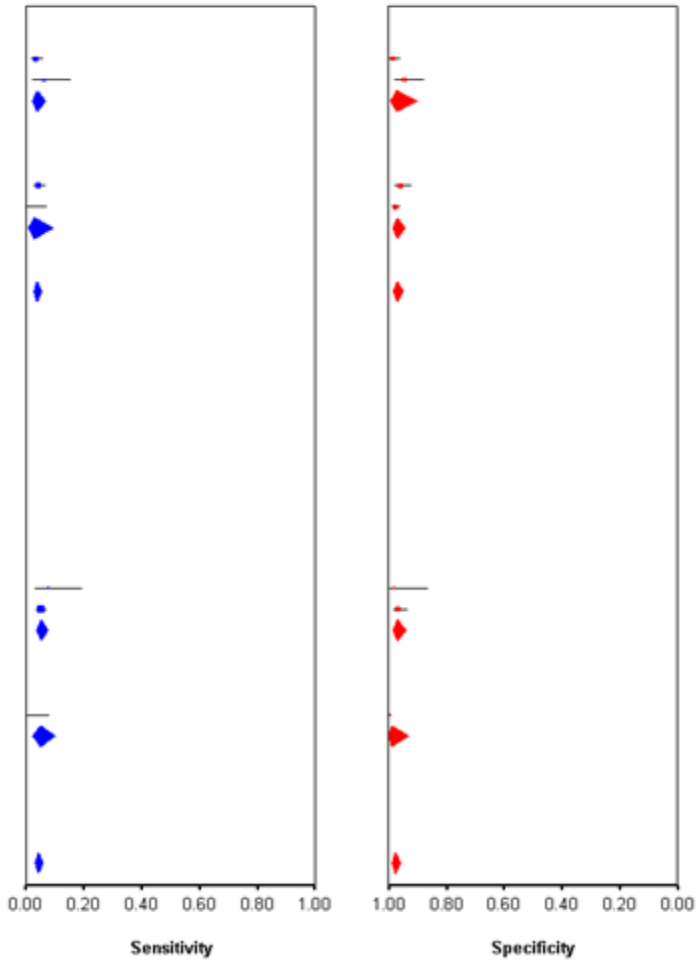


Figure 36: Inflammatory bowel disease – forest plot: sensitivity and specificity

G.1.1.18 Dactylitis

Table 19: Dactylitis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	4 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	Serious[c]	1,785	2.28 (1.31, 3.96)	LOW
LR-			No serious	No serious	Serious[b]	No serious		0.96 (0.94, 0.98)	MODERATE
PERIPHERAL									
LR+	2 studies[d]	Cross-sectional	Serious[e]	Serious[f]	Serious[b]	Serious[c]	229	9.59 (1.15, 80.06)	VERY LOW
LR-			No serious	Serious[f]	No serious	Serious[g]		0.66 (0.28, 1.57)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[h]	Cross-sectional	No serious	No serious	No serious	No serious	874	14.67 (2.87, 75.08)	HIGH
LR-			No serious	No serious	No serious	No serious		0.92 (0.90, 0.95)	HIGH
ALL EVIDENCE POOLED									
LR+	8 studies[i]	Cross-sectional	No serious	Serious[f]	Serious[b]	Serious[c]	2,888	4.26 (1.90, 9.56)	VERY LOW
LR-			No serious	Serious[f]	No serious	No serious		0.95 (0.92, 0.98)	MODERATE

a *van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
b *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
c *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
d *Sadek 2007; You 2015*
e *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
f *I2 ≥ 50%*
g *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
h *D'Agostino 2011; Tomero 2014*
i *D'Agostino 2011; Sadek 2007; Tomero 2014; You 2015; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	28	393	5	259	0.95 (0.92, 0.98)	3.51 (1.37, 8.98)
van den Berg 2013b (SPACE)	4	61	2	90	0.96 (0.90, 1.03)	2.83 (0.53, 15.00)
RE subtotal					**0.95 (0.93, 0.98)**	**3.33 (1.47, 7.56)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.04, df=1 (p=0.833); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.05, df=1 (p=0.825); I^2=0.0%						
Published criteria						
van Hoeven 2014	4	82	9	269	0.99 (0.94, 1.04)	1.44 (0.45, 4.55)
van Hoeven 2015	5	90	14	470	0.98 (0.93, 1.03)	1.82 (0.67, 4.93)
RE subtotal					**0.98 (0.95, 1.02)**	**1.64 (0.77, 3.50)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.07, df=1 (p=0.784); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.09, df=1 (p=0.761); I^2=0.0%						
RE subtotal					**0.96 (0.94, 0.98)**	**2.28 (1.31, 3.96)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=1.64, df=3 (p=0.651); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=1.69, df=3 (p=0.639); I^2=0.0%						
Between-substratum heterogeneity, LR-: Chi^2=1.52, df=1 (p=0.218); I^2=34.0%						
Between-substratum heterogeneity, LR+: Chi^2=1.55, df=1 (p=0.213); I^2=35.5%						
PERIPHERAL						
Clinician diagnosis						
Sadek 2007	2	57	0	22	0.98 (0.90, 1.06)	1.92 (0.10, 38.42)
Published criteria						
You 2015	11	7	4	126	0.40 (0.22, 0.72)	19.86 (7.07, 55.79)
RE subtotal					**0.66 (0.28, 1.57)**	**9.59 (1.15, 80.06)**
Within-stratum heterogeneity, LR-: Tau^2=0.35; Chi^2=8.93, df=1 (p=0.003); I^2=88.8%						
Within-stratum heterogeneity, LR+: Tau^2=1.42; Chi^2=2.09, df=1 (p=0.148); I^2=52.1%						
Between-substratum heterogeneity, LR-: Chi^2=8.93, df=1 (p=0.003); I^2=88.8%						
Between-substratum heterogeneity, LR+: Chi^2=2.09, df=1 (p=0.148); I^2=52.1%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	4	47	0	48	0.92 (0.84, 1.01)	8.48 (0.47, 153.45)
Tomero 2014	43	495	1	236	0.92 (0.90, 0.95)	18.94 (2.62, 136.75)
RE subtotal					**0.92 (0.90, 0.95)**	**14.67 (2.87, 75.08)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.00, df=1 (p=0.980); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.20, df=1 (p=0.653); I^2=0.0%						
Published criteria						
no data						
RE subtotal					**0.92 (0.90, 0.95)**	**14.67 (2.87, 75.08)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.00, df=1 (p=0.980); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.20, df=1 (p=0.653); I^2=0.0%						
RE meta-analysis					**0.95 (0.92, 0.98)**	**4.26 (1.90, 9.56)**
Overall heterogeneity, LR-: Tau^2=0.00; Chi^2=16.70, df=7 (p=0.019); I^2=58.1%						
Overall heterogeneity, LR+: Tau^2=0.73; Chi^2=17.78, df=7 (p=0.013); I^2=60.6%						
Between-stratum heterogeneity, LR-: Chi^2=6.13, df=2 (p=0.047); I^2=67.4%						
Between-stratum heterogeneity, LR+: Chi^2=13.80, df=2 (p=0.001); I^2=85.5%						

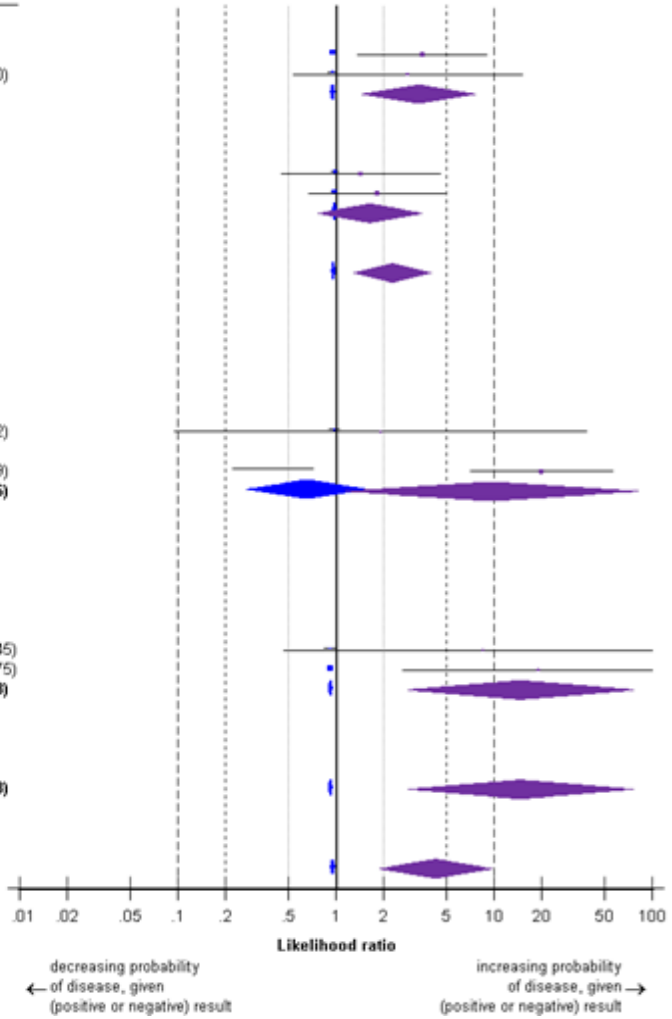


Figure 37: Dactylitis – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	28	393	5	259	0.07 (0.05, 0.09)	0.98 (0.96, 0.99)
van den Berg 2013b (SPACE)	4	61	2	90	0.06 (0.02, 0.15)	0.98 (0.92, 0.99)
RE subtotal					**0.07 (0.05, 0.09)**	**0.98 (0.96, 0.99)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.02, df=1 (p=0.881); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.03, df=1 (p=0.868); I^2=0.0%						
Published criteria						
van Hoeven 2014	4	82	9	269	0.05 (0.02, 0.12)	0.97 (0.94, 0.98)
van Hoeven 2015	5	90	14	470	0.05 (0.02, 0.12)	0.97 (0.95, 0.98)
RE subtotal					**0.05 (0.03, 0.09)**	**0.97 (0.95, 0.98)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.04, df=1 (p=0.850); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.07, df=1 (p=0.789); I^2=0.0%						
RE subtotal					**0.06 (0.05, 0.08)**	**0.97 (0.96, 0.98)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.64, df=3 (p=0.887); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=1.11, df=3 (p=0.775); I^2=0.0%						
Between-substratum heterogeneity, LR-: Chi^2=0.58, df=1 (p=0.445); I^2=0.0%						
Between-substratum heterogeneity, LR+: Chi^2=1.01, df=1 (p=0.315); I^2=1.1%						
PERIPHERAL						
Clinician diagnosis						
Sadek 2007	2	57	0	22	0.04 (0.01, 0.13)	0.98 (0.73, 1.00)
Published criteria						
You 2015	11	7	4	126	0.61 (0.38, 0.80)	0.97 (0.92, 0.99)
RE subtotal					**0.21 (0.01, 0.90)**	**0.97 (0.93, 0.99)**
Within-stratum heterogeneity, LR-: Tau^2=6.11; Chi^2=19.76, df=1 (p<0.001); I^2=94.9%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.06, df=1 (p=0.814); I^2=0.0%						
Between-substratum heterogeneity, LR-: Chi^2=19.76, df=1 (p<0.001); I^2=94.9%						
Between-substratum heterogeneity, LR+: Chi^2=0.06, df=1 (p=0.814); I^2=0.0%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	4	47	0	48	0.09 (0.03, 0.20)	0.99 (0.86, 1.00)
Tornero 2014	43	495	1	236	0.08 (0.06, 0.11)	1.00 (0.97, 1.00)
RE subtotal					**0.08 (0.06, 0.11)**	**0.99 (0.97, 1.00)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.03, df=1 (p=0.867); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.26, df=1 (p=0.609); I^2=0.0%						
Published criteria						
no data						
RE subtotal					**0.08 (0.06, 0.11)**	**0.99 (0.97, 1.00)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.03, df=1 (p=0.867); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.26, df=1 (p=0.609); I^2=0.0%						
RE meta-analysis					**0.09 (0.05, 0.15)**	**0.97 (0.96, 0.98)**
Overall heterogeneity, LR-: Tau^2=0.54; Chi^2=40.58, df=7 (p<0.001); I^2=82.7%						
Overall heterogeneity, LR+: Tau^2=0.00; Chi^2=5.14, df=7 (p=0.643); I^2=0.0%						
Between-stratum heterogeneity, LR-: Chi^2=20.14, df=2 (p<0.001); I^2=90.1%						
Between-stratum heterogeneity, LR+: Chi^2=3.71, df=2 (p=0.156); I^2=46.2%						

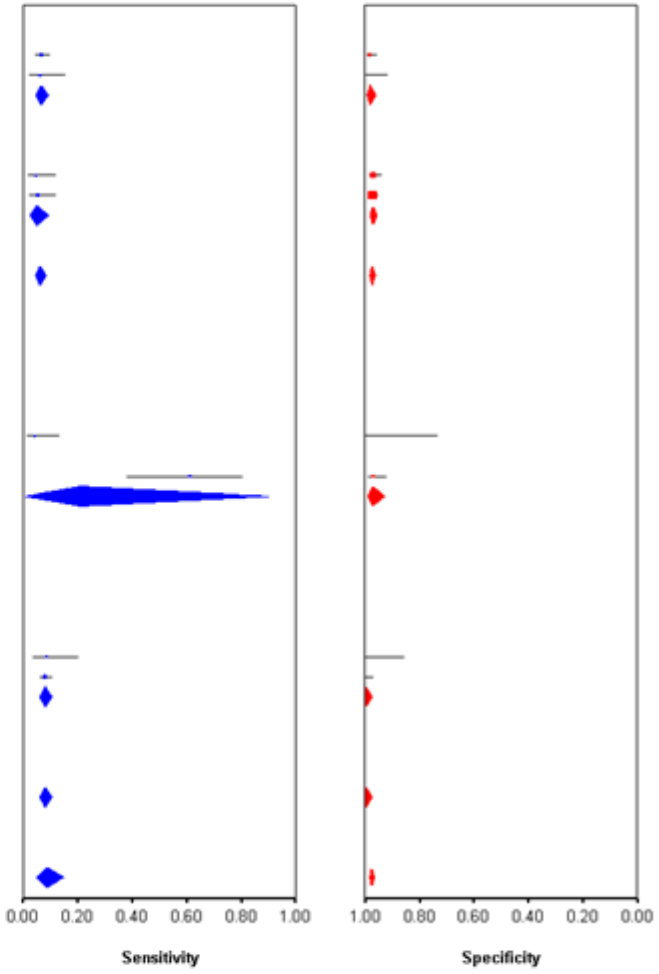


Figure 38: Dactylitis – forest plot: sensitivity and specificity

G.1.1.19 Arthritis

Arthritis / peripheral arthritis

Table 20: Arthritis / peripheral arthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	6 studies[a]	Cross-sectional	No serious	Serious[b]	Serious[c]	No serious	2,670	1.08 (0.84, 1.38)	LOW
LR-			No serious	Serious[b]	Serious[c]	No serious		1.00 (0.93, 1.07)	LOW
PERIPHERAL									
LR+	1 study[d]	Cross-sectional	Serious	n/a	No serious	No serious	191	3.74 (2.88, 4.85)	MODERATE
LR-			Serious	n/a	No serious	No serious		0.03 (0.00, 0.46)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[e]	Cross-sectional	No serious	Serious[b]	No serious	Serious[f]	874	2.32 (0.70, 7.70)	LOW
LR-			No serious	No serious	No serious	No serious		0.86 (0.82, 0.90)	HIGH
ALL EVIDENCE POOLED									
LR+	9 studies[g]	Cross-sectional	No serious	Serious[b]	Serious[c]	Serious[f]	3,735	1.57 (0.98, 2.53)	VERY LOW
LR-			No serious	Serious[b]	Serious[c]	No serious		0.96 (0.88, 1.05)	LOW

[a] *Dougados 2011 (DESIR); Hulsemann 1995; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *I2 ≥ 50%*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *Mattila 1998*
[e] *D'Agostino 2011; Tomero 2014*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[g] *Dougados 2011 (DESIR); D'Agostino 2011; Hulsemann 1995; Mattila 1998; Tomero 2014; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

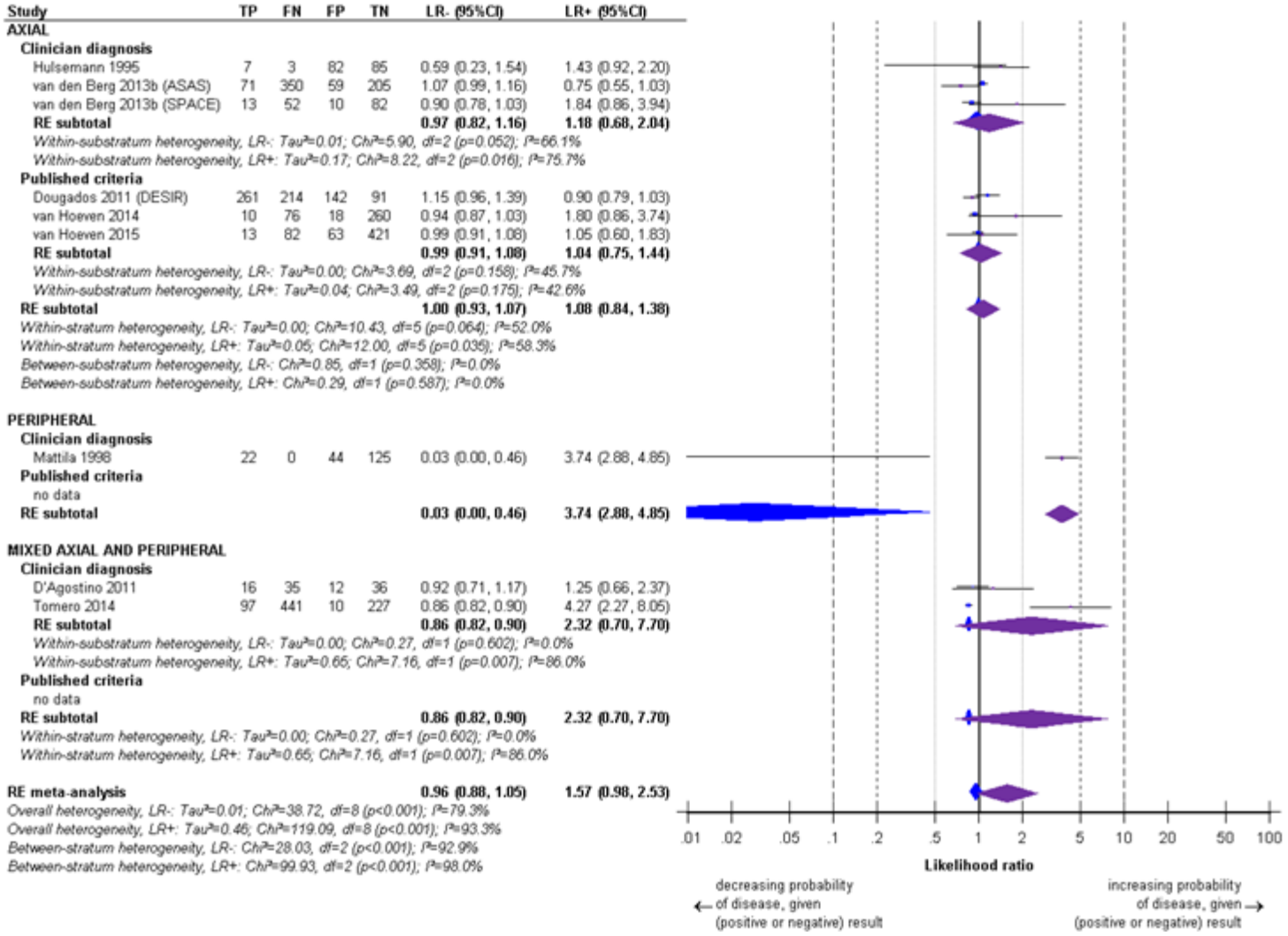


Figure 39: Arthritis / peripheral arthritis – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Hulsemann 1995	7	3	82	85	0.70 (0.38, 0.90)	0.51 (0.43, 0.58)
van den Berg 2013b (ASAS)	71	350	59	205	0.17 (0.14, 0.21)	0.78 (0.72, 0.82)
van den Berg 2013b (SPACE)	13	52	10	82	0.20 (0.12, 0.31)	0.89 (0.81, 0.94)
RE subtotal					**0.27 (0.13, 0.48)**	**0.75 (0.51, 0.90)**
Within-substratum heterogeneity, LR-: Tau^2=0.48; Chi^2=12.24, df=2 (p=0.002); I^2=83.7%						
Within-substratum heterogeneity, LR+: Tau^2=0.84; Chi^2=48.62, df=2 (p<0.001); I^2=95.9%						
Published criteria						
Dougados 2011 (DESIR)	261	214	142	91	0.55 (0.50, 0.59)	0.39 (0.33, 0.45)
van Hoeven 2014	10	76	18	260	0.12 (0.06, 0.20)	0.94 (0.90, 0.96)
van Hoeven 2015	13	82	63	421	0.14 (0.08, 0.22)	0.87 (0.84, 0.90)
RE subtotal					**0.23 (0.05, 0.61)**	**0.80 (0.38, 0.96)**
Within-substratum heterogeneity, LR-: Tau^2=2.11; Chi^2=77.36, df=2 (p<0.001); I^2=97.4%						
Within-substratum heterogeneity, LR+: Tau^2=2.66; Chi^2=207.04, df=2 (p<0.001); I^2=99.0%						
RE subtotal					**0.26 (0.12, 0.49)**	**0.78 (0.57, 0.90)**
Within-stratum heterogeneity, LR-: Tau^2=1.40; Chi^2=179.55, df=5 (p<0.001); I^2=97.2%						
Within-stratum heterogeneity, LR+: Tau^2=1.42; Chi^2=257.38, df=5 (p<0.001); I^2=98.1%						
Between-substratum heterogeneity, LR-: Chi^2=89.95, df=1 (p<0.001); I^2=98.9%						
Between-substratum heterogeneity, LR+: Chi^2=1.72, df=1 (p=0.190); I^2=41.8%						
PERIPHERAL						
Clinician diagnosis						
Mattila 1998	22	0	44	125	0.98 (0.73, 1.00)	0.74 (0.67, 0.80)
Published criteria						
no data						
RE subtotal					**0.98 (0.73, 1.00)**	**0.74 (0.67, 0.80)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	16	35	12	36	0.31 (0.20, 0.45)	0.75 (0.61, 0.85)
Tomero 2014	97	441	10	227	0.18 (0.15, 0.22)	0.96 (0.92, 0.98)
RE subtotal					**0.23 (0.13, 0.38)**	**0.89 (0.53, 0.98)**
Within-substratum heterogeneity, LR-: Tau^2=0.22; Chi^2=5.16, df=1 (p=0.023); I^2=80.6%						
Within-substratum heterogeneity, LR+: Tau^2=1.94; Chi^2=19.00, df=1 (p<0.001); I^2=94.7%						
Published criteria						
no data						
RE subtotal					**0.23 (0.13, 0.38)**	**0.89 (0.53, 0.98)**
Within-stratum heterogeneity, LR-: Tau^2=0.22; Chi^2=5.16, df=1 (p=0.023); I^2=80.6%						
Within-stratum heterogeneity, LR+: Tau^2=1.94; Chi^2=19.00, df=1 (p<0.001); I^2=94.7%						
RE meta-analysis					**0.30 (0.17, 0.47)**	**0.80 (0.66, 0.90)**
Overall heterogeneity, LR-: Tau^2=1.05; Chi^2=240.04, df=8 (p<0.001); I^2=96.7%						
Overall heterogeneity, LR+: Tau^2=1.26; Chi^2=302.79, df=8 (p<0.001); I^2=97.4%						
Between-stratum heterogeneity, LR-: Chi^2=55.33, df=2 (p<0.001); I^2=96.4%						
Between-stratum heterogeneity, LR+: Chi^2=26.41, df=2 (p<0.001); I^2=92.4%						

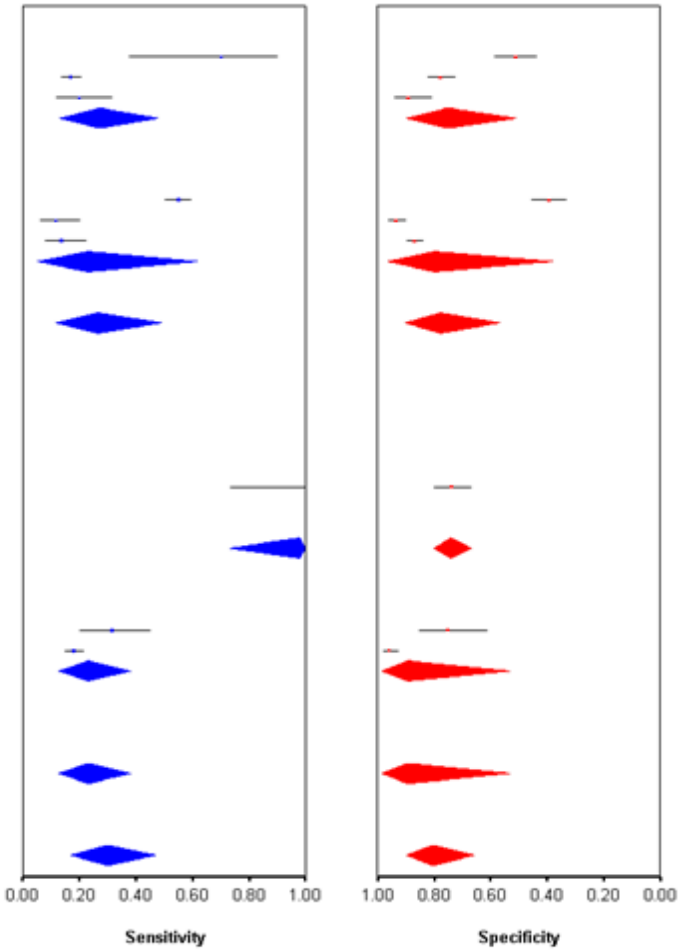


Figure 40: Arthritis / peripheral arthritis – forest plot: sensitivity and specificity

G.1.1.20 Oligoarthritis (in people with symptoms of peripheral arthritis)

Table 21 Oligoarthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	299	28.58 (2.85, 286.02)	LOW
LR-			Serious[b]	Serious[d]	Serious[c]	No serious		0.76 (0.64, 0.90)	VERY LOW
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	299	28.58 (2.85, 286.02)	LOW
LR-			Serious[b]	Serious[d]	Serious[c]	No serious		0.76 (0.64, 0.90)	VERY LOW

[a] *Sadek 2007; Tinazzi 2012*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *I2 ≥ 50%*

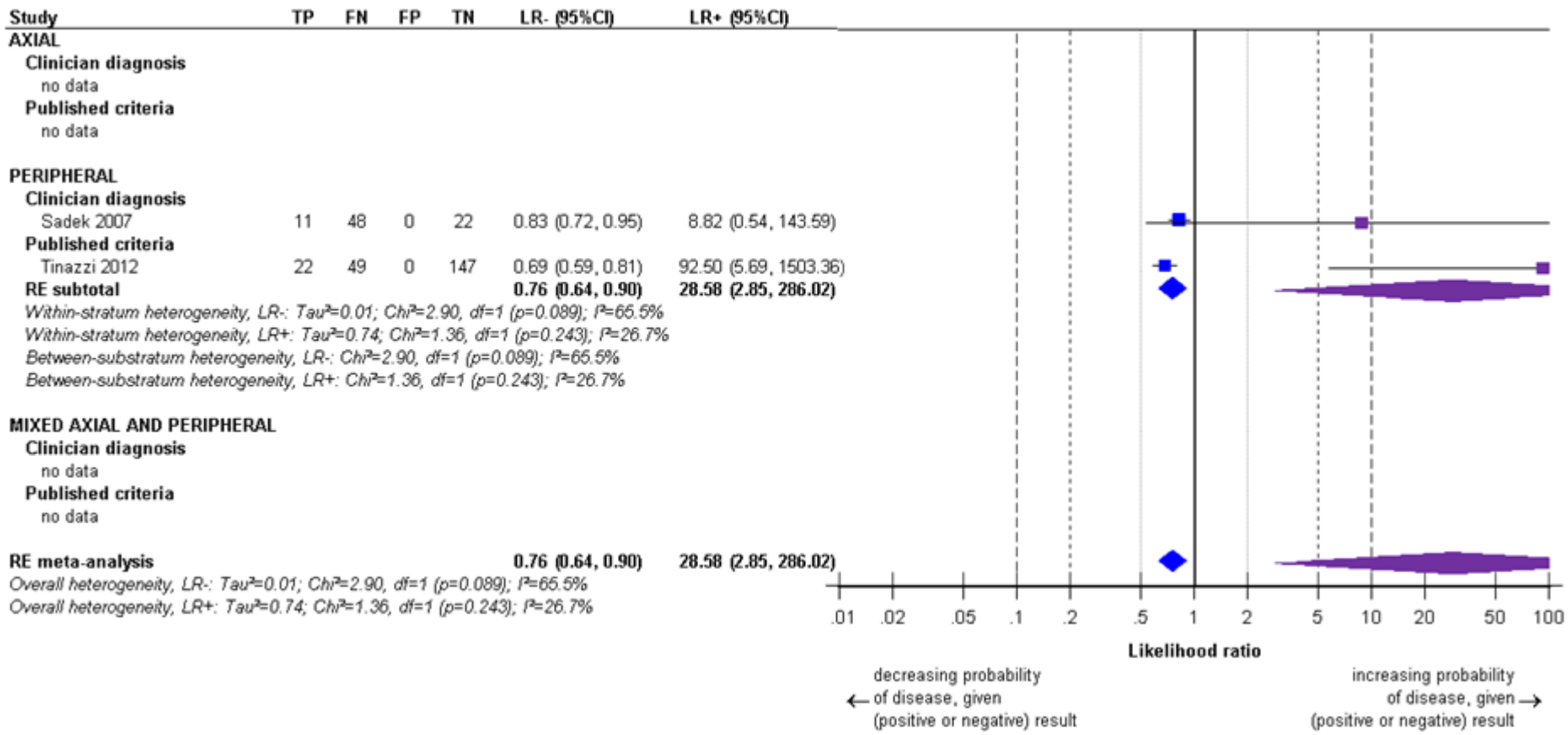


Figure 41: Oligoarthritis (in people with symptoms of peripheral arthritis) – forest plot: likelihood ratios

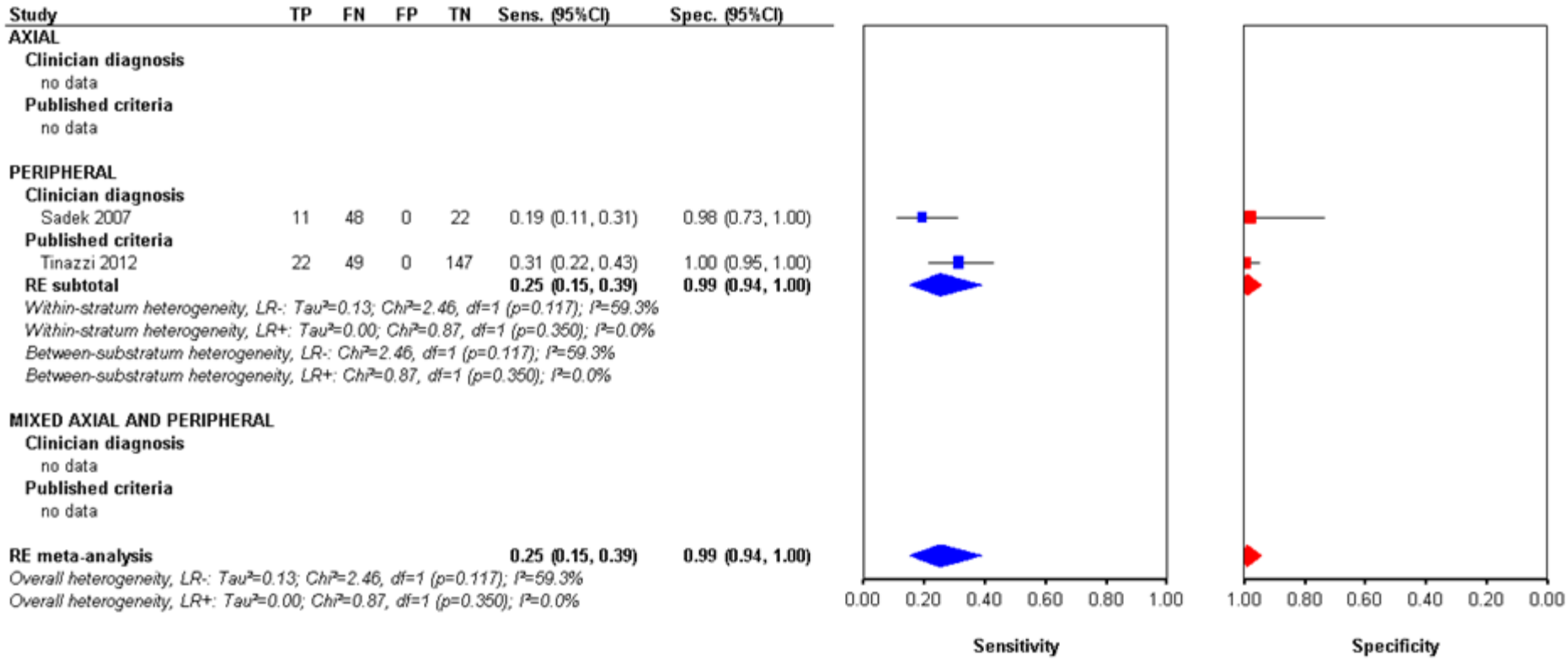


Figure 42: Oligoarthritis (in people with symptoms of peripheral arthritis) – forest plot: sensitivity and specificity

G.1.1.21 Nail disease

Table 22: Nail disease – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	5 studies[a]	Cross-sectional	Serious[b]	Serious[c]	Serious[d]	Serious[e]	3,568	1.60 (1.03, 2.47)	VERY LOW
LR-			Serious[b]	No serious	Serious[d]	No serious		0.76 (0.64, 0.91)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	5 studies[a]	Cross-sectional	Serious[b]	Serious[c]	Serious[d]	Serious[e]	3,568	1.60 (1.03, 2.47)	VERY LOW
LR-			Serious[b]	No serious	Serious[d]	No serious		0.76 (0.64, 0.91)	LOW

[a] *Haroon 2013; Tinazzi 2012; Wilson 2009; Yang 2011; You 2015*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

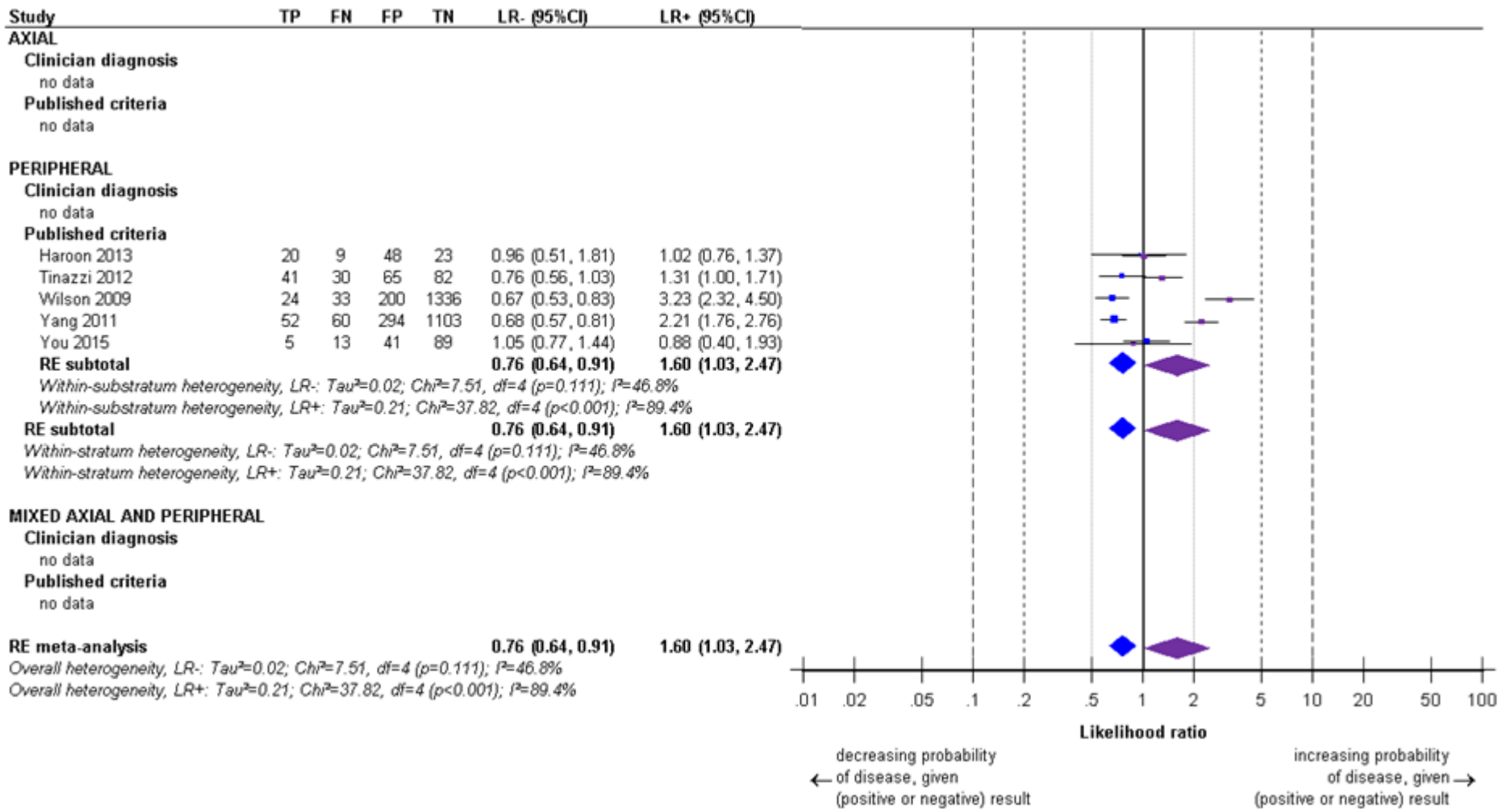


Figure 43: Nail disease – forest plot: likelihood ratios

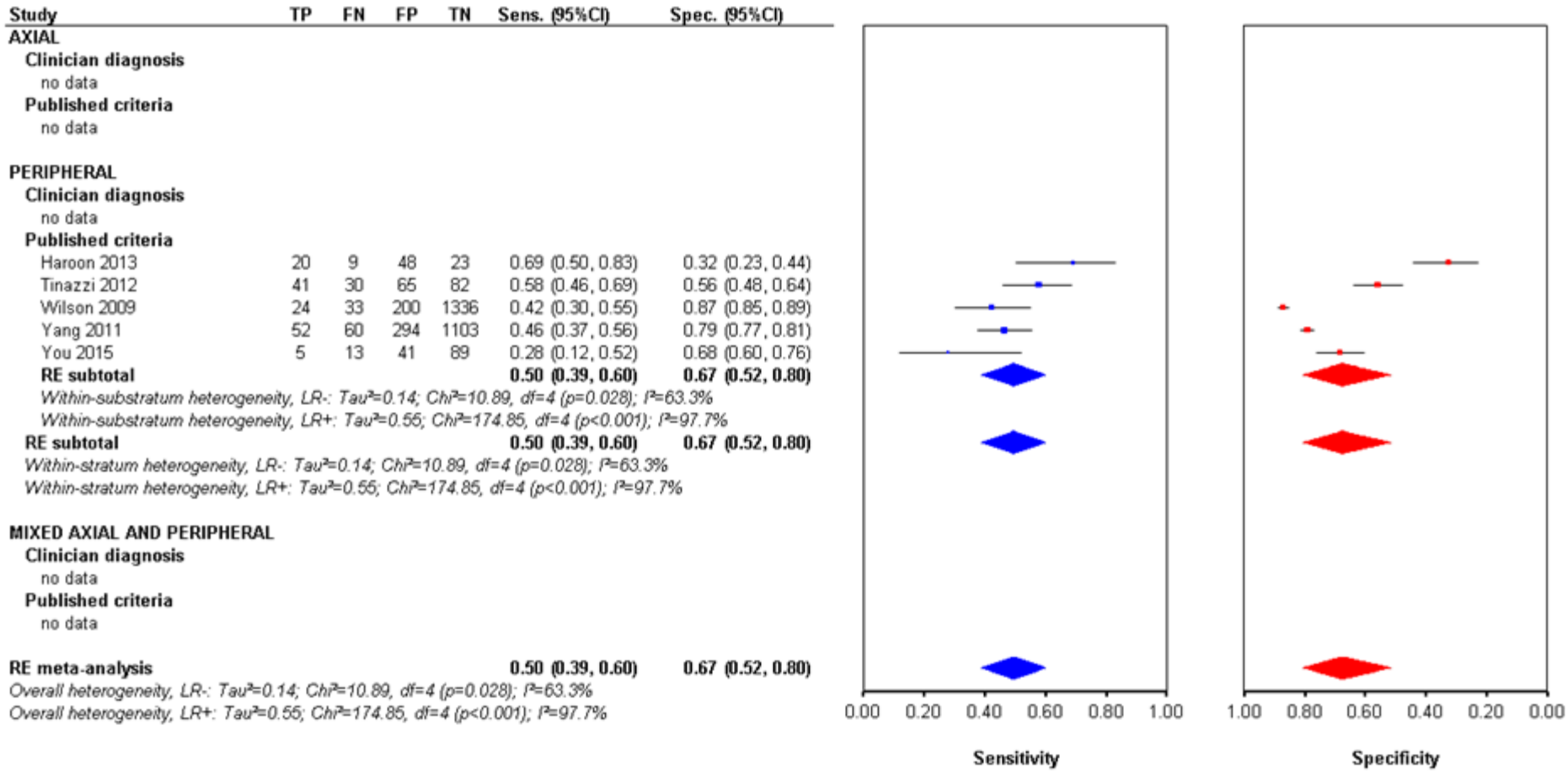


Figure 44: Nail disease – forest plot: sensitivity and specificity

G.1.1.22 Fatigue / malaise

Table 23: Fatigue / malaise – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	329	0.93 (0.70, 1.24)	MODERATE
LR-			Serious[c]	No serious	Serious[b]	No serious		1.14 (0.89, 1.45)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	329	0.93 (0.70, 1.24)	MODERATE
LR-			Serious[c]	No serious	Serious[b]	No serious		1.14 (0.89, 1.45)	LOW

[a] *Kvien 1996; Mattila 1998*
[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*

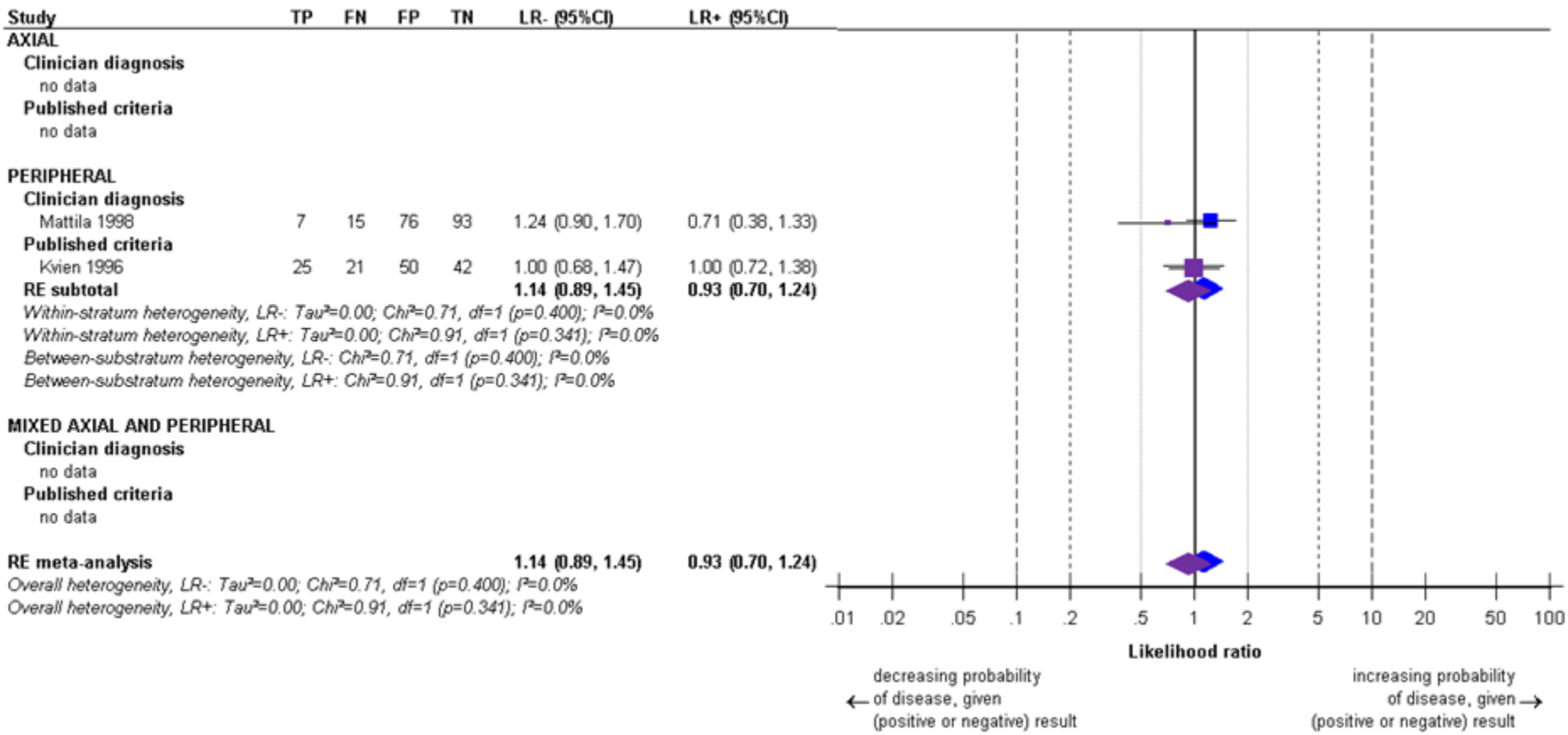


Figure 45: Fatigue / malaise – forest plot: likelihood ratios

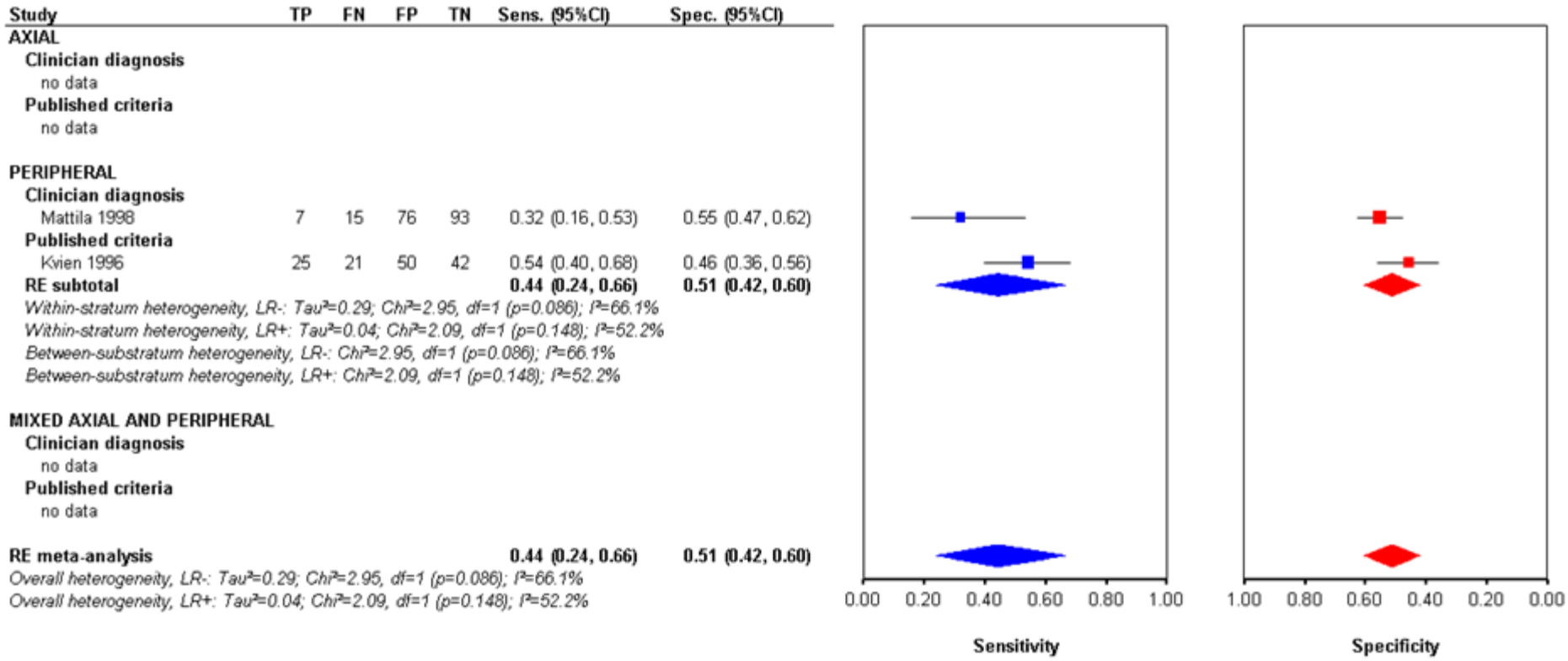


Figure 46: Fatigue / malaise – forest plot: sensitivity and specificity

Family history of spondyloarthritis

Table 24: Family history of spondyloarthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	6 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	2,908	1.63 (1.33, 1.98)	HIGH
LR-			No serious	No serious	No serious	No serious		0.91 (0.86, 0.96)	HIGH
PERIPHERAL									
LR+	2 studies[b]	Cross-sectional	No serious	Serious[c]	No serious	Serious[d]	666	5.35 (0.87, 32.86)	LOW
LR-			Serious[e]	Serious[c]	No serious	No serious		0.91 (0.84, 0.98)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	4 studies[f]	Cross-sectional	No serious	Serious[c]	No serious	Serious[d]	1,821	2.13 (1.13, 4.01)	LOW
LR-			Serious[e]	Serious[c]	Serious[g]	No serious		0.89 (0.79, 1.00)	VERY LOW
ALL EVIDENCE POOLED									
LR+	12 studies[h]	Cross-sectional	No serious	No serious	No serious	Serious[d]	5,395	1.81 (1.46, 2.23)	MODERATE
LR-			Serious[e]	Serious[c]	No serious	No serious		0.91 (0.87, 0.94)	LOW

[a] *Poddubnyy 2011; Sieper 2013; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *Rudwaleit 2011; Tey 2010*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[f] *D'Agostino 2011; Liao 2009; Salvarini 2001; Tomero 2014*
[g] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[h] *D'Agostino 2011; Liao 2009; Poddubnyy 2011; Rudwaleit 2011; Salvarini 2001; Sieper 2013; Tey 2010; Tomero 2014; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

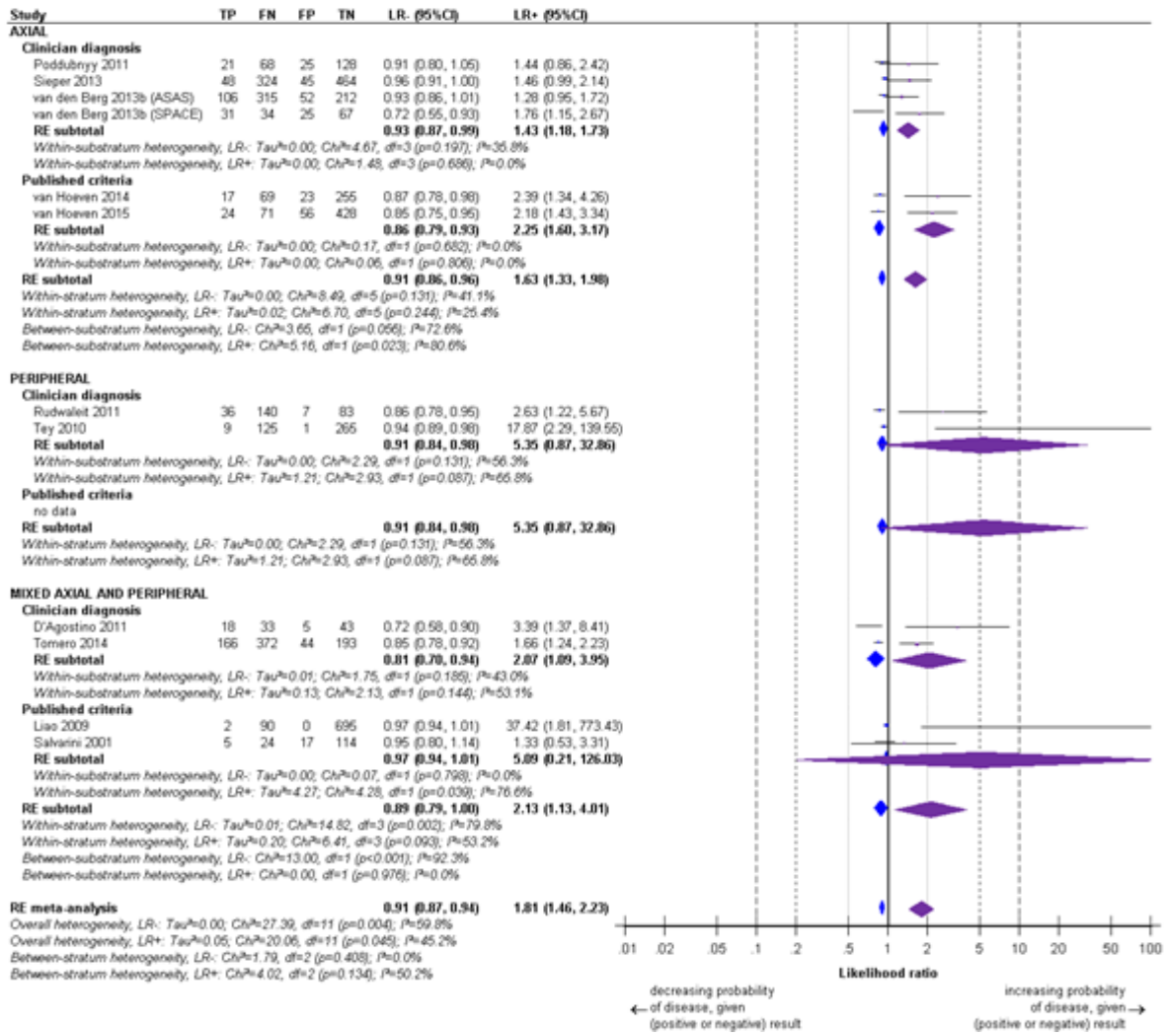

Study TP FN FP TN LR- (95%CI) LR+ (95%CI)
AXIAL
Clinician diagnosis
Poddubnyy 2011 21 68 25 128 0.91 (0.80, 1.05) 1.44 (0.86, 2.42)
Sieper 2013 48 324 45 464 0.96 (0.91, 1.00) 1.46 (0.99, 2.14)
van den Berg 2013b (ASAS) 106 315 52 212 0.93 (0.86, 1.01) 1.28 (0.95, 1.72)
van den Berg 2013b (SPACE) 31 34 25 67 0.72 (0.55, 0.93) 1.76 (1.15, 2.67)
RE subtotal 0.93 (0.87, 0.99) 1.43 (1.18, 1.73)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=4.67, df=3 (p=0.197); I²=35.8%
Within-substratum heterogeneity, LR+: Tau²=0.00; Ch²=1.48, df=3 (p=0.686); I²=0.0%
Published criteria
van Hoeven 2014 17 69 23 255 0.87 (0.78, 0.98) 2.39 (1.34, 4.26)
van Hoeven 2015 24 71 56 428 0.85 (0.75, 0.95) 2.18 (1.43, 3.34)
RE subtotal 0.86 (0.79, 0.93) 2.25 (1.60, 3.17)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=0.17, df=1 (p=0.682); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=0.00; Ch²=0.06, df=1 (p=0.806); I²=0.0%
RE subtotal 0.91 (0.86, 0.96) 1.63 (1.33, 1.98)
Within-stratum heterogeneity, LR-: Tau²=0.00; Ch²=8.49, df=5 (p=0.131); I²=41.1%
Within-stratum heterogeneity, LR+: Tau²=0.02; Ch²=6.70, df=5 (p=0.244); I²=25.4%
Between-substratum heterogeneity, LR-: Ch²=3.65, df=1 (p=0.056); I²=72.6%
Between-substratum heterogeneity, LR+: Ch²=5.16, df=1 (p=0.023); I²=80.6%
PERIPHERAL
Clinician diagnosis
Rudwaleit 2011 36 140 7 83 0.86 (0.78, 0.95) 2.63 (1.22, 5.67)
Tey 2010 9 125 1 265 0.94 (0.89, 0.98) 17.87 (2.29, 139.55)
RE subtotal 0.91 (0.84, 0.98) 5.35 (0.87, 32.86)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=2.29, df=1 (p=0.131); I²=56.3%
Within-substratum heterogeneity, LR+: Tau²=1.21; Ch²=2.93, df=1 (p=0.087); I²=65.8%
Published criteria
no data
RE subtotal 0.91 (0.84, 0.98) 5.35 (0.87, 32.86)
Within-stratum heterogeneity, LR-: Tau²=0.00; Ch²=2.29, df=1 (p=0.131); I²=56.3%
Within-stratum heterogeneity, LR+: Tau²=1.21; Ch²=2.93, df=1 (p=0.087); I²=65.8%
MIXED AXIAL AND PERIPHERAL
Clinician diagnosis
D'Agostino 2011 18 33 5 43 0.72 (0.58, 0.90) 3.39 (1.37, 8.41)
Tornero 2014 166 372 44 193 0.85 (0.78, 0.92) 1.66 (1.24, 2.23)
RE subtotal 0.81 (0.70, 0.94) 2.87 (1.09, 3.95)
Within-substratum heterogeneity, LR-: Tau²=0.01; Ch²=1.75, df=1 (p=0.186); I²=43.0%
Within-substratum heterogeneity, LR+: Tau²=0.13; Ch²=2.13, df=1 (p=0.144); I²=53.1%
Published criteria
Liao 2009 2 90 0 696 0.97 (0.94, 1.01) 37.42 (1.81, 773.43)
Salvarani 2001 5 24 17 114 0.95 (0.80, 1.14) 1.33 (0.53, 3.31)
RE subtotal 0.97 (0.94, 1.01) 5.09 (0.21, 126.03)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=0.07, df=1 (p=0.798); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=4.27; Ch²=4.28, df=1 (p=0.039); I²=76.6%
RE subtotal 0.89 (0.79, 1.00) 2.13 (1.13, 4.01)
Within-stratum heterogeneity, LR-: Tau²=0.01; Ch²=14.82, df=3 (p=0.002); I²=79.8%
Within-stratum heterogeneity, LR+: Tau²=0.20; Ch²=6.41, df=3 (p=0.093); I²=53.2%
Between-substratum heterogeneity, LR-: Ch²=13.00, df=1 (p<0.001); I²=92.3%
Between-substratum heterogeneity, LR+: Ch²=0.00, df=1 (p=0.976); I²=0.0%
RE meta-analysis 0.91 (0.87, 0.94) 1.81 (1.46, 2.23)
Overall heterogeneity, LR-: Tau²=0.00; Ch²=27.39, df=11 (p=0.004); I²=59.8%
Overall heterogeneity, LR+: Tau²=0.05; Ch²=20.06, df=11 (p=0.045); I²=45.2%
Between-stratum heterogeneity, LR-: Ch²=1.79, df=2 (p=0.408); I²=0.0%
Between-stratum heterogeneity, LR+: Ch²=4.02, df=2 (p=0.134); I²=50.2%
.01 .02 .05 .1 .2 .5 1 2 5 10 20 50 100
Likelihood ratio
decreasing probability of disease, given (positive or negative) result
increasing probability of disease, given (positive or negative) result

Figure 47: Family history of spondyloarthritis – forest plot: likelihood ratios

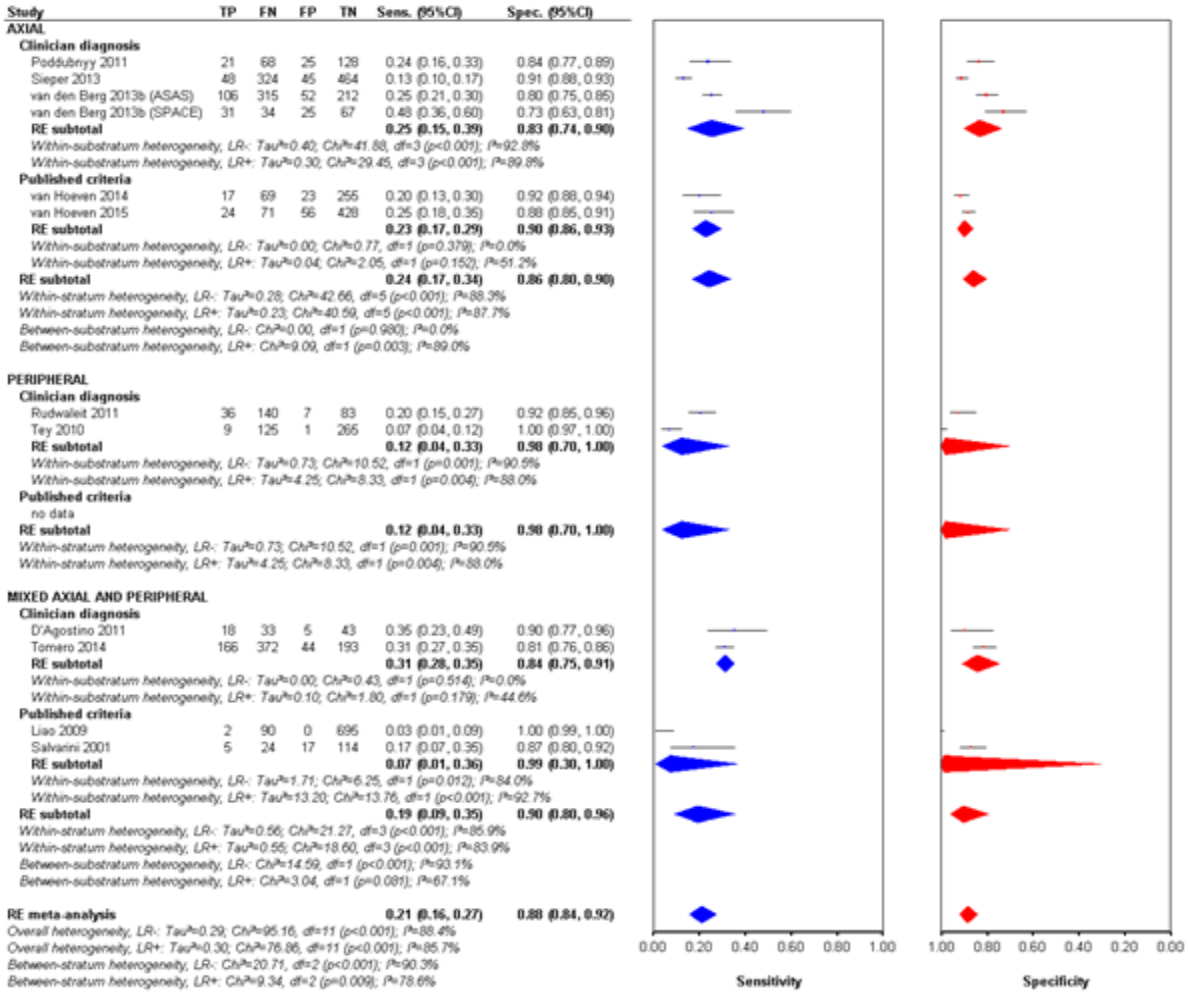

Study TP FN FP TN Sens. (95%CI) Spec. (95%CI)
AXIAL
Clinician diagnosis
Poddubnyy 2011 21 68 25 128 0.24 (0.16, 0.33) 0.84 (0.77, 0.89)
Sieper 2013 48 324 45 464 0.13 (0.10, 0.17) 0.91 (0.88, 0.93)
van den Berg 2013b (ASAS) 106 315 52 212 0.25 (0.21, 0.30) 0.80 (0.75, 0.85)
van den Berg 2013b (SPACE) 31 34 25 67 0.48 (0.36, 0.60) 0.73 (0.63, 0.81)
RE subtotal 0.25 (0.15, 0.39) 0.83 (0.74, 0.90)
Within-substratum heterogeneity, LR-: Tau²=0.40; Ch²=41.88, df=3 (p<0.001); I²=92.8%
Within-substratum heterogeneity, LR+: Tau²=0.30; Ch²=29.45, df=3 (p<0.001); I²=89.8%
Published criteria
van Hoeven 2014 17 69 23 255 0.20 (0.13, 0.30) 0.92 (0.88, 0.94)
van Hoeven 2015 24 71 56 428 0.25 (0.18, 0.35) 0.88 (0.85, 0.91)
RE subtotal 0.23 (0.17, 0.29) 0.90 (0.86, 0.93)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=0.77, df=1 (p=0.379); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=0.04; Ch²=2.05, df=1 (p=0.152); I²=51.2%
RE subtotal 0.24 (0.17, 0.34) 0.86 (0.80, 0.90)
Within-stratum heterogeneity, LR-: Tau²=0.28; Ch²=42.66, df=5 (p<0.001); I²=88.3%
Within-stratum heterogeneity, LR+: Tau²=0.23; Ch²=40.59, df=5 (p<0.001); I²=87.7%
Between-substratum heterogeneity, LR-: Ch²=0.00, df=1 (p=0.980); I²=0.0%
Between-substratum heterogeneity, LR+: Ch²=9.09, df=1 (p=0.003); I²=89.0%
PERIPHERAL
Clinician diagnosis
Rudwaleit 2011 36 140 7 83 0.20 (0.15, 0.27) 0.92 (0.85, 0.96)
Tey 2010 9 125 1 265 0.07 (0.04, 0.12) 1.00 (0.97, 1.00)
RE subtotal 0.12 (0.04, 0.33) 0.98 (0.70, 1.00)
Within-substratum heterogeneity, LR-: Tau²=0.73; Ch²=10.52, df=1 (p=0.001); I²=90.5%
Within-substratum heterogeneity, LR+: Tau²=4.25; Ch²=8.33, df=1 (p=0.004); I²=88.0%
Published criteria
no data
RE subtotal 0.12 (0.04, 0.33) 0.98 (0.70, 1.00)
Within-stratum heterogeneity, LR-: Tau²=0.73; Ch²=10.52, df=1 (p=0.001); I²=90.5%
Within-stratum heterogeneity, LR+: Tau²=4.25; Ch²=8.33, df=1 (p=0.004); I²=88.0%
MIXED AXIAL AND PERIPHERAL
Clinician diagnosis
D'Agostino 2011 18 33 5 43 0.35 (0.23, 0.49) 0.90 (0.77, 0.96)
Tomero 2014 166 372 44 193 0.31 (0.27, 0.35) 0.81 (0.76, 0.86)
RE subtotal 0.31 (0.28, 0.35) 0.84 (0.75, 0.91)
Within-substratum heterogeneity, LR-: Tau²=0.00; Ch²=0.43, df=1 (p=0.514); I²=0.0%
Within-substratum heterogeneity, LR+: Tau²=0.10; Ch²=1.80, df=1 (p=0.179); I²=44.6%
Published criteria
Liao 2009 2 90 0 696 0.03 (0.01, 0.09) 1.00 (0.99, 1.00)
Salvarini 2001 5 24 17 114 0.17 (0.07, 0.35) 0.87 (0.80, 0.92)
RE subtotal 0.07 (0.01, 0.36) 0.99 (0.30, 1.00)
Within-substratum heterogeneity, LR-: Tau²=1.71; Ch²=6.25, df=1 (p=0.012); I²=84.0%
Within-substratum heterogeneity, LR+: Tau²=13.20; Ch²=13.76, df=1 (p<0.001); I²=92.7%
RE subtotal 0.19 (0.09, 0.35) 0.90 (0.80, 0.96)
Within-stratum heterogeneity, LR-: Tau²=0.56; Ch²=21.27, df=3 (p<0.001); I²=85.9%
Within-stratum heterogeneity, LR+: Tau²=0.55; Ch²=18.60, df=3 (p<0.001); I²=83.9%
Between-substratum heterogeneity, LR-: Ch²=14.59, df=1 (p<0.001); I²=93.1%
Between-substratum heterogeneity, LR+: Ch²=3.04, df=1 (p=0.081); I²=67.1%
RE meta-analysis 0.21 (0.16, 0.27) 0.88 (0.84, 0.92)
Overall heterogeneity, LR-: Tau²=0.29; Ch²=95.16, df=11 (p<0.001); I²=88.4%
Overall heterogeneity, LR+: Tau²=0.30; Ch²=76.86, df=11 (p<0.001); I²=85.7%
Between-stratum heterogeneity, LR-: Ch²=20.71, df=2 (p<0.001); I²=90.3%
Between-stratum heterogeneity, LR+: Ch²=9.34, df=2 (p=0.009); I²=78.6%
0.00 0.20 0.40 0.60 0.80 1.00
Sensitivity
1.00 0.80 0.60 0.40 0.20 0.00
Specificity

Figure 48: Family history of spondyloarthritis – forest plot: sensitivity and specificity

G.1.1.23 Family history of psoriasis

Table 25: Family history of psoriasis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	1,909	1.34 (1.06, 1.70)	LOW
LR-			Serious[b]	No serious	Serious[c]	No serious		0.91 (0.84, 0.99)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	Serious[b]	No serious	Serious[c]	No serious	1,909	1.34 (1.06, 1.70)	LOW
LR-			Serious[b]	No serious	Serious[c]	No serious		0.91 (0.84, 0.99)	LOW

[a] *Tey 2010; Yang 2011*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*

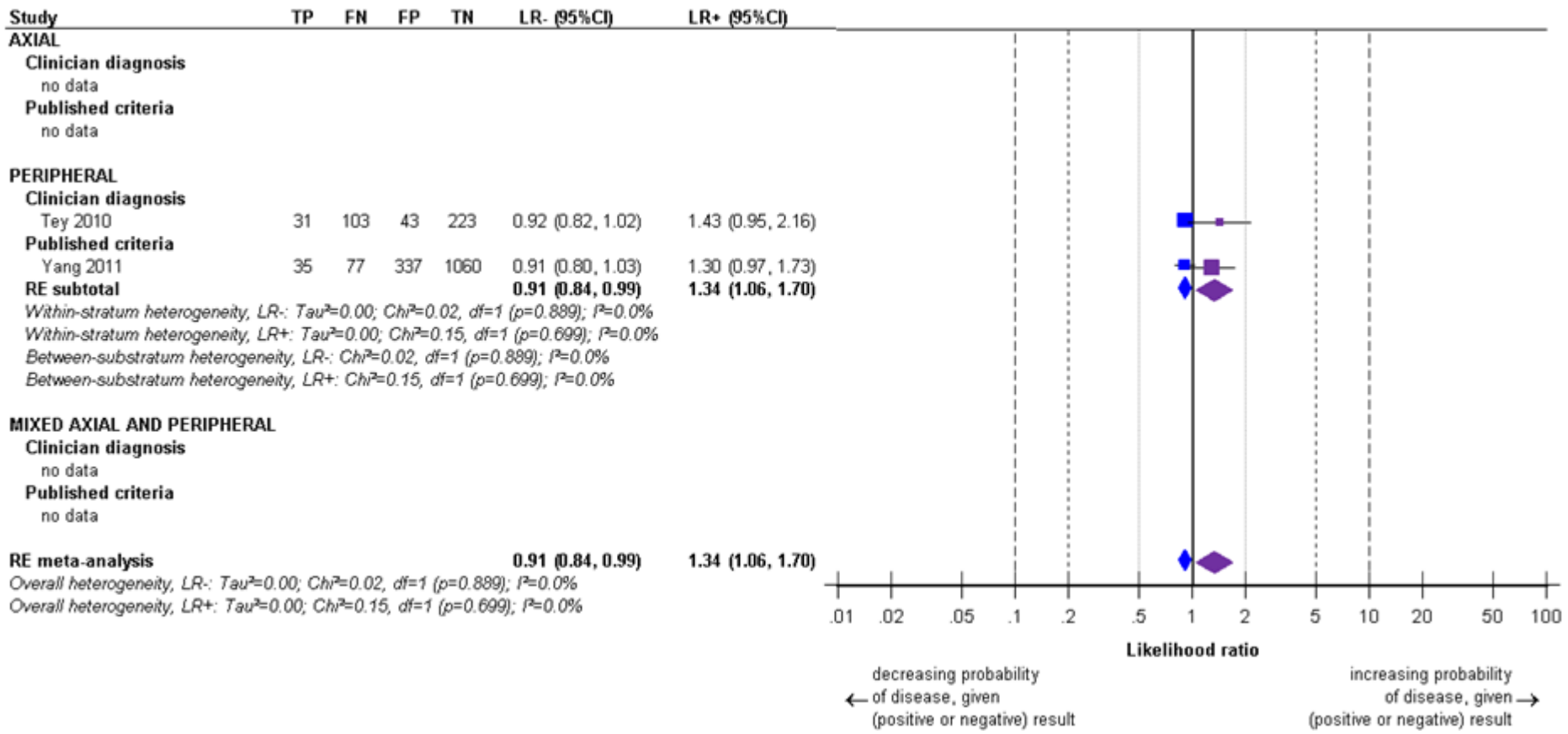


Figure 49: Family history of psoriasis – forest plot: likelihood ratios

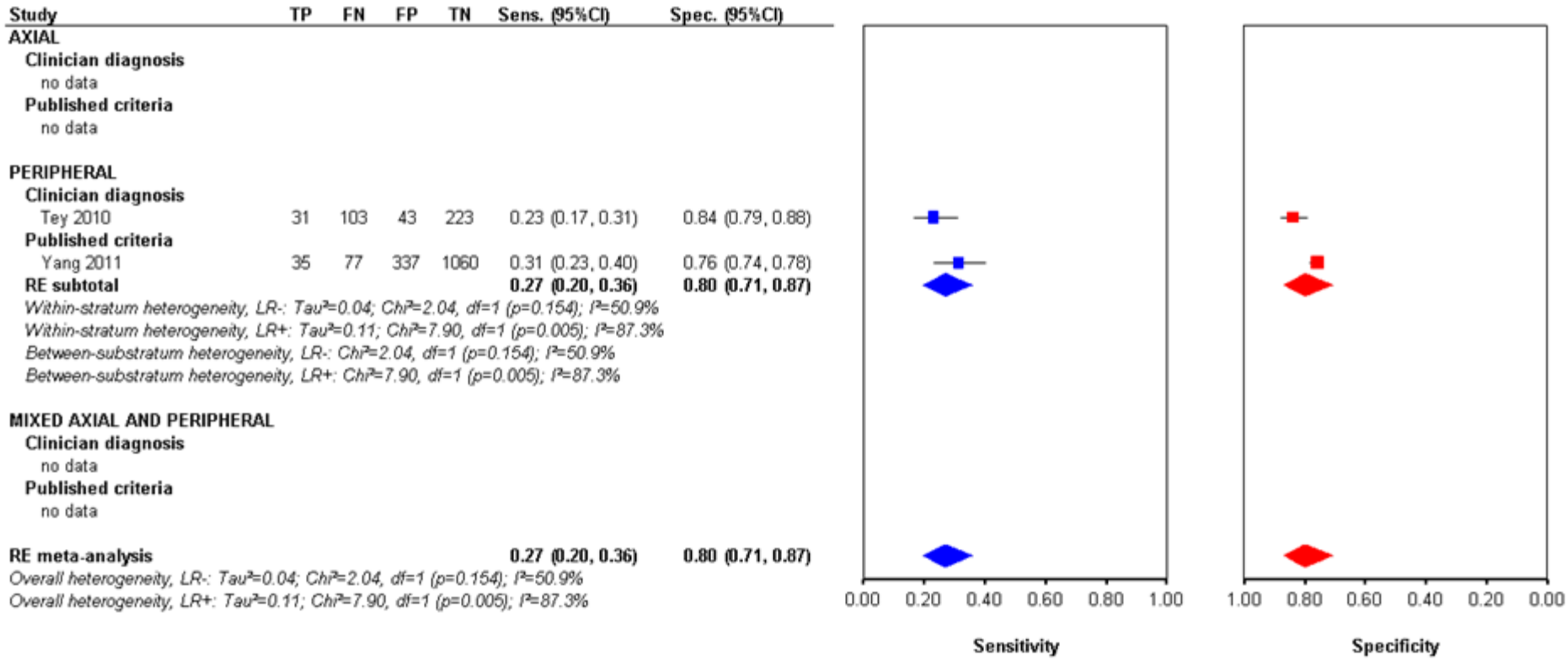


Figure 50: Family history of psoriasis – forest plot: sensitivity and specificity

G.1.1.24 Preceding infection

Table 26: Preceding infection – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	Serious[b]	842	1.77 (0.67, 4.70)	MODERATE
LR-			No serious	No serious	No serious	No serious		0.99 (0.97, 1.01)	HIGH
PERIPHERAL									
LR+	2 studies[c]	Cross-sectional	No serious	Serious[d]	Serious[e]	Serious[b]	638	3.80 (1.08, 13.33)	VERY LOW
LR-			No serious	Serious[d]	No serious	Serious[f]		0.63 (0.25, 1.55)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[g]	Cross-sectional	Serious[h]	No serious	No serious	Serious[b]	1,337	2.11 (1.01, 4.39)	LOW
LR-			No serious	Serious[d]	No serious	No serious		0.94 (0.87, 1.03)	MODERATE
ALL EVIDENCE POOLED									
LR+	7 studies[i]	Cross-sectional	No serious	Serious[d]	No serious	Serious[b]	2,817	2.71 (1.36, 5.38)	LOW
LR-			No serious	Serious[d]	No serious	No serious		0.96 (0.92, 1.00)	MODERATE

[a] *van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[c] *Kvien 1994; Rudwaleit 2011*
[d] *I2 ≥ 50%*
[e] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[g] *Granfors 1983; Hulsemann 1995; Tomero 2014*
[h] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[i] *Granfors 1983; Hulsemann 1995; Kvien 1994; Rudwaleit 2011; Tomero 2014; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	12	409	5	259	0.99 (0.97, 1.01)	1.50 (0.54, 4.22)
van den Berg 2013b (SPACE)	2	63	0	92	0.97 (0.92, 1.02)	7.05 (0.34, 144.36)
RE subtotal					**0.99 (0.97, 1.01)**	**1.77 (0.67, 4.70)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.69, df=1 (p=0.407); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.90, df=1 (p=0.343); I^2=0.0%						
Published criteria						
no data						
RE subtotal					**0.99 (0.97, 1.01)**	**1.77 (0.67, 4.70)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.69, df=1 (p=0.407); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.90, df=1 (p=0.343); I^2=0.0%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	10	166	3	87	0.98 (0.93, 1.03)	1.70 (0.48, 6.04)
Published criteria						
Kvien 1994	34	18	33	287	0.39 (0.27, 0.56)	6.34 (4.34, 9.26)
RE subtotal					**0.63 (0.25, 1.55)**	**3.80 (1.08, 13.33)**
Within-stratum heterogeneity, LR-: Tau^2=0.41; Chi^2=22.99, df=1 (p<0.001); I^2=95.7%						
Within-stratum heterogeneity, LR+: Tau^2=0.64; Chi^2=3.80, df=1 (p=0.051); I^2=73.7%						
Between-substratum heterogeneity, LR-: Chi^2=22.99, df=1 (p<0.001); I^2=95.7%						
Between-substratum heterogeneity, LR+: Chi^2=3.80, df=1 (p=0.051); I^2=73.7%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Granfors 1983	12	50	20	272	0.87 (0.76, 0.98)	2.83 (1.46, 5.47)
Hulsemann 1995	10	31	31	136	0.93 (0.77, 1.12)	1.31 (0.70, 2.46)
Tornero 2014	10	528	0	237	0.98 (0.97, 1.00)	9.27 (0.55, 157.59)
RE subtotal					**0.94 (0.87, 1.03)**	**2.11 (1.01, 4.39)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=4.17, df=2 (p=0.124); I^2=52.1%						
Within-substratum heterogeneity, LR+: Tau^2=0.19; Chi^2=3.90, df=2 (p=0.142); I^2=48.7%						
Published criteria						
no data						
RE subtotal					**0.94 (0.87, 1.03)**	**2.11 (1.01, 4.39)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=4.17, df=2 (p=0.124); I^2=52.1%						
Within-stratum heterogeneity, LR+: Tau^2=0.19; Chi^2=3.90, df=2 (p=0.142); I^2=48.7%						
RE meta-analysis					**0.96 (0.92, 1.00)**	**2.71 (1.36, 5.38)**
Overall heterogeneity, LR-: Tau^2=0.00; Chi^2=28.85, df=6 (p<0.001); I^2=79.2%						
Overall heterogeneity, LR+: Tau^2=0.51; Chi^2=23.74, df=6 (p<0.001); I^2=74.7%						
Between-stratum heterogeneity, LR-: Chi^2=1.00, df=2 (p=0.608); I^2=0.0%						
Between-stratum heterogeneity, LR+: Chi^2=15.14, df=2 (p<0.001); I^2=86.8%						

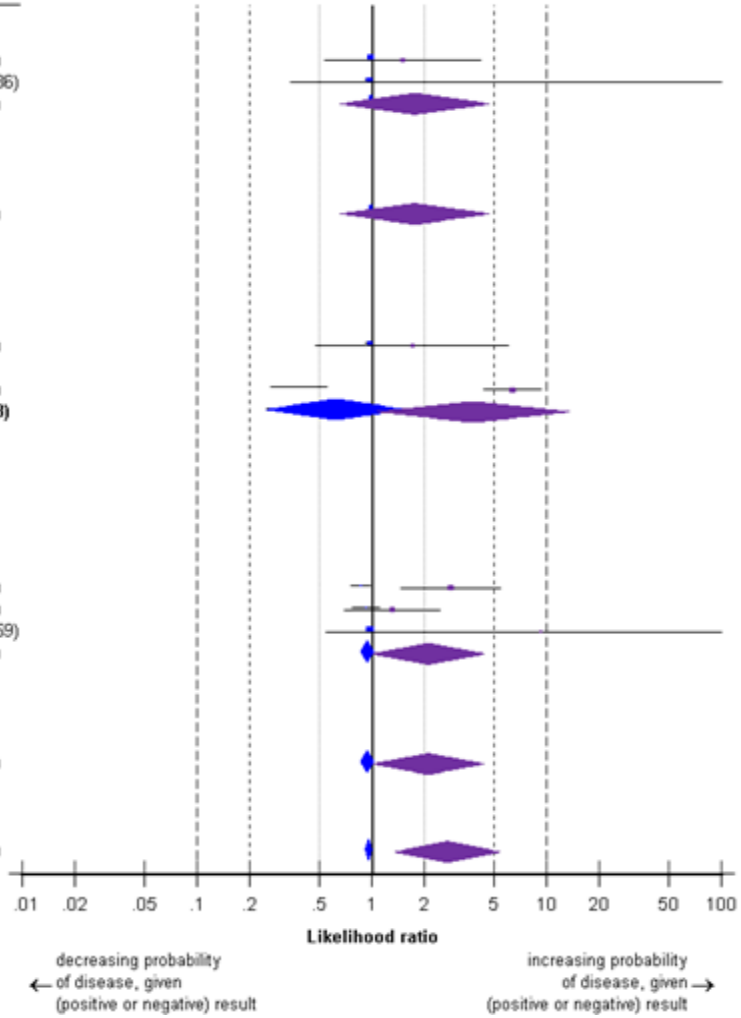


Figure 51: Preceding infection – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	12	409	5	259	0.03 (0.02, 0.05)	0.98 (0.96, 0.99)
van den Berg 2013b (SPACE)	2	63	0	92	0.04 (0.01, 0.12)	0.99 (0.92, 1.00)
RE subtotal					**0.03 (0.02, 0.05)**	**0.98 (0.96, 0.99)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.17, df=1 (p=0.678); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.73, df=1 (p=0.392); I^2=0.0%						
Published criteria						
no data						
RE subtotal					**0.03 (0.02, 0.05)**	**0.98 (0.96, 0.99)**
Within-stratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.17, df=1 (p=0.678); I^2=0.0%						
Within-stratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.73, df=1 (p=0.392); I^2=0.0%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	10	166	3	87	0.06 (0.03, 0.10)	0.97 (0.90, 0.99)
Published criteria						
Kvien 1994	34	18	33	287	0.65 (0.52, 0.77)	0.90 (0.86, 0.93)
RE subtotal					**0.25 (0.01, 0.91)**	**0.93 (0.82, 0.98)**
Within-stratum heterogeneity, LR-: Tau^2=5.84; Chi^2=62.15, df=1 (p<0.001); I^2=98.4%						
Within-stratum heterogeneity, LR+: Tau^2=0.54; Chi^2=3.83, df=1 (p=0.050); I^2=73.9%						
Between-substratum heterogeneity, LR-: Chi^2=62.15, df=1 (p<0.001); I^2=98.4%						
Between-substratum heterogeneity, LR+: Chi^2=3.83, df=1 (p=0.050); I^2=73.9%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Granfors 1983	12	50	20	272	0.19 (0.11, 0.31)	0.93 (0.90, 0.96)
Hulsemann 1995	10	31	31	136	0.24 (0.14, 0.40)	0.81 (0.75, 0.87)
Tornero 2014	10	528	0	237	0.02 (0.01, 0.04)	1.00 (0.97, 1.00)
RE subtotal					**0.10 (0.02, 0.40)**	**0.93 (0.80, 0.98)**
Within-substratum heterogeneity, LR-: Tau^2=2.34; Chi^2=44.66, df=2 (p<0.001); I^2=95.5%						
Within-substratum heterogeneity, LR+: Tau^2=0.92; Chi^2=22.44, df=2 (p<0.001); I^2=91.1%						
Published criteria						
no data						
RE subtotal					**0.10 (0.02, 0.40)**	**0.93 (0.80, 0.98)**
Within-stratum heterogeneity, LR-: Tau^2=2.34; Chi^2=44.66, df=2 (p<0.001); I^2=95.5%						
Within-stratum heterogeneity, LR+: Tau^2=0.92; Chi^2=22.44, df=2 (p<0.001); I^2=91.1%						
RE meta-analysis					**0.10 (0.03, 0.29)**	**0.95 (0.90, 0.98)**
Overall heterogeneity, LR-: Tau^2=3.01; Chi^2=165.25, df=6 (p<0.001); I^2=96.4%						
Overall heterogeneity, LR+: Tau^2=0.70; Chi^2=47.44, df=6 (p<0.001); I^2=87.4%						
Between-stratum heterogeneity, LR-: Chi^2=58.26, df=2 (p<0.001); I^2=96.6%						
Between-stratum heterogeneity, LR+: Chi^2=20.44, df=2 (p<0.001); I^2=90.2%						

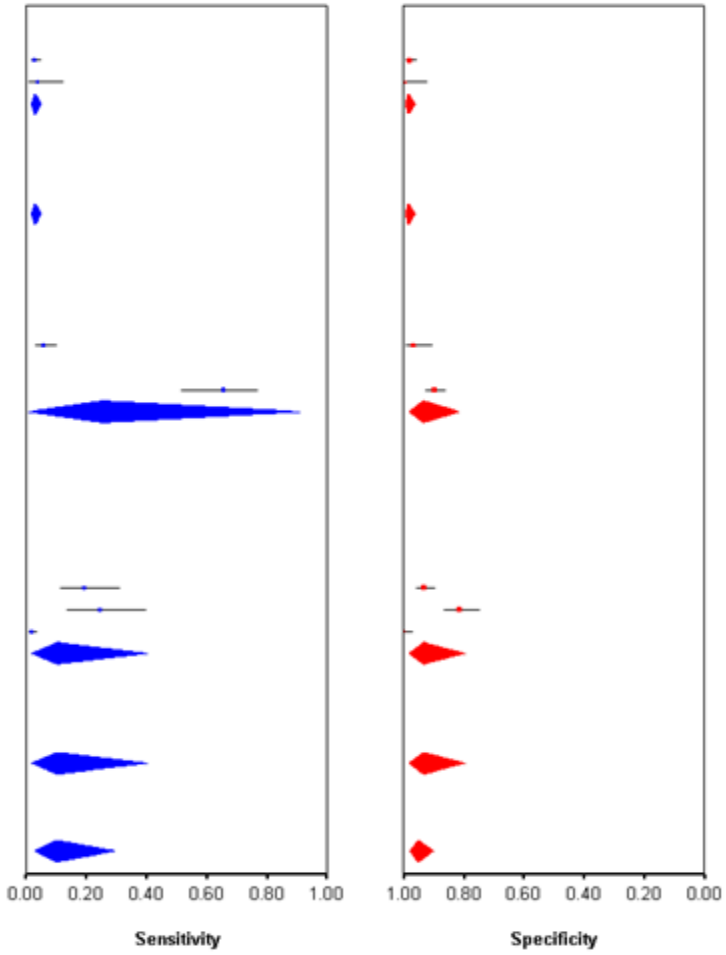


Figure 52: Preceding infection – forest plot: sensitivity and specificity

G.1.2 Indicators for referral

Review Question 12

- What are the indications (signs, risk factors, test or scan findings) for referral for specialist advice at initial diagnosis?

Table 27: Indicators for referral for suspected axial spondyloarthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
Sensitivity	Braun (2011)	Cohort study	No serious	No serious	No serious	Serious[a]	322[b]	See evidence table	MODERATE
Specificity			No serious	No serious	No serious	Serious[a]		See evidence table	MODERATE
AXIAL									
Sensitivity	Braun (2013)	Cohort study	No serious	No serious	No serious	Serious[a]	322[b]	See evidence table	MODERATE
Specificity			No serious	No serious	No serious	Serious[a]		See evidence table	MODERATE
AXIAL									
Sensitivity	van Houeven (2015a)	Cohort study	No serious	No serious	Serious[c]	Serious[a]	579[b]	See evidence table	LOW
Specificity			No serious	No serious	Serious[c]	Serious[a]		See evidence table	LOW
AXIAL									
Sensitivity	van Houeven (2015b)	Cohort study	No serious	No serious	Serious[c]	Serious[a]	579[b]	See evidence table	LOW
Specificity			No serious	No serious	Serious[c]	Serious[a]		See evidence table	LOW

[a] *Wide confidence intervals around sensitivity and specificity*
[b] *Total number with a confirmend diagnosis of either spondyloarthritis or not spondyloarthritis*
[c] *All participants in the study underwent imaging for sacroiliitis, which is not the case in the relevant UK population*

G.1.3 Comparative effectiveness of referral strategies

Review Question 6

- What is the comparative effectiveness of different referral strategies in diagnosing spondyloarthritis?

Table 28: Poddubnny 2011 referral strategies for axial spondyloarthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
Proportion of those referred diagnosed with axial spondyloarthritis									
Mean difference	Poddubnny (2011)	Cohort study	No serious	No serious	No serious	Serious[a]	560[b]	-5.1% (-13.1%, 3.1%)	MODERATE
Proportion of those referred diagnosed with possible axial spondyloarthritis									
Mean difference	Poddubnny (2011)	Cohort study	No serious	No serious	No serious	Serious[a]	560[b]	2.2% (-3.7%, 8.3%)	MODERATE
Proportion of those referred diagnosed as not having axial spondyloarthritis									
Mean difference	Poddubnny (2011)	Cohort study	No serious	No serious	No serious	Serious[a]	560[b]	2.9% (-5.4%, 11.1%)	MODERATE

[a] *No differences detected between referral strategies*
[b] *Total number of people referred through either strategy*

Table 29: Sieper 2013 referral strategies for axial spondyloarthritis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
Proportion of those referred diagnosed with axial spondyloarthritis									
Mean difference	Sieper (2013)	Cohort study	No serious	No serious	No serious	Serious[a]	1,049[b]	4.2% (-1.7%, 10.0%)	MODERATE
Proportion of those referred diagnosed with possible axial spondyloarthritis									
Mean difference	Sieper (2013)	Cohort study	No serious	No serious	No serious	Serious[a]	1,049[b]	-0.3% (-2.9%, 3.7%)	MODERATE
Proportion of those referred diagnosed as not having axial spondyloarthritis									
Mean difference	Sieper (2013)	Cohort study	No serious	No serious	No serious	Serious[a]	1,049[b]	3.9% (-2.2%, 9.9%)	MODERATE

[a] *No differences detected between referral strategies*
[b] *Total number of people referred through either strategy*

G.1.4 Obstacles to prompt diagnosis

Review Question 3

- What are the obstacles to a prompt diagnosis of spondyloarthritis?

G.1.4.1 Quality assessment

For the type of evidence included in this question (cross-sectional interview and survey based studies) the GRADE framework is not considered to be appropriate. A checklist developed by the British Medical Journal was used in the quality assessment of these studies.

Using this checklist all of the included studies were considered to be of very low quality. The studies included limited detail of any decisions on sampling, most were single-centre studies, they did not included any details of questions asked, any evidence of a piloting process, any validation of questionnaires/interview questions or provide details on who administered the questionnaires/interviews or any training they may have had.

G.1.4.2 BMJ checklist

(http://www.bmj.com/content/suppl/2004/05/27/328.7451.1312.DC1#e)

Research question and study design

- What information did the researchers seek to obtain?
- Was a questionnaire the most appropriate method and if not, what design might have been more appropriate?
- Were there any existing measures (questionnaires) that the researchers could have used? If so, why was a new one developed and was this justified?
- Were the views of consumers sought about the design, distribution, and administration of the questionnaire?
- Validity and reliability

Validity and reliability

- What claims for validity have been made, and are they justified? (In other words, what evidence is there that the instrument measures what it sets out to measure?)
- What claims for reliability have been made, and are they justified? (In other words, what evidence is there that the instrument provides stable responses over time and between researchers?)

Format

- Was the title of the questionnaire appropriate and if not, what were its limitations?What format did the questionnaire take, and were open and closed questions used appropriately?
- Were easy, non-threatening questions placed at the beginning of the measure and sensitive ones near the end?
- Was the questionnaire kept as brief as the study allowed?
- Did the questions make sense, and could the participants in the sample understand them? Were any questions ambiguous or overly complicated?

Instructions

- Did the questionnaire contain adequate instructions for completion—eg example answers, or an explanation of whether a ticked or written response was required?
- Were participants told how to return the questionnaire once completed?
- Did the questionnaire contain an explanation of the research, a summary of what would happen to the data, and a thank you message?

Piloting

- Was the questionnaire adequately piloted in terms of the method and means of administration, on people who were representative of the study population?
- How was the piloting exercise undertaken—what details are given?
- In what ways was the definitive instrument changed as a result of piloting?

Sampling

- What was the sampling frame for the definitive study and was it sufficiently large and representative?
- Was the instrument suitable for all participants and potential participants? In particular, did it take account of the likely range of physical/mental/cognitive abilities, language/literacy, understanding of numbers/scaling, and perceived threat of questions or questioner?

Distribution, administration and response

- How was the questionnaire distributed?
- How was the questionnaire administered?
- Were the response rates reported fully, including details of participants who were unsuitable for the research or refused to take part?
- Have any potential response biases been discussed?

Coding and analysis

- What sort of analysis was carried out and was this appropriate? (eg correct statistical tests for quantitative answers, qualitative analysis for open ended questions)
- What measures were in place to maintain the accuracy of the data, and were these adequate?
- Is there any evidence of data dredging—that is, analyses that were not hypothesis driven?
- Results
- What were the results and were all relevant data reported?
- Are quantitative results definitive (significant), and are relevant non-significant results also reported?
- Have qualitative results been adequately interpreted (e.g. using an explicit theoretical framework), and have any quotes been properly justified and contextualised?

Conclusions and discussion

- What do the results mean and have the researchers drawn an appropriate link between the data and their conclusions?
- Have the findings been placed within the wider body of knowledge in the field (eg via a comprehensive literature review), and are any recommendations justified?

G.1.5 Blood tests for spondyloarthritis

Review questions 7-9

- What is the diagnostic utility of a HLA B27 test for investigating suspected spondyloarthritis?
- What is the diagnostic utility of an erythrocyte sedimentation rate test for investigating suspected spondyloarthritis?
- What is the diagnostic utility of a C-reactive protein test for investigating suspected spondyloarthritis?

G.1.5.1 HLA-B27

Table 30: GRADE table for HLA-B27

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	13 studiesa	Cross-sectional	No serious	Seriousb	Seriousc	No serious	4,645	4.14 (3.09, 5.56)	LOW
LR-			No serious	Seriousb	Seriousc	Seriousd		0.37 (0.27, 0.50)	VERY LOW
PERIPHERAL									
LR+	7 studiese	Cross-sectional	No serious	Seriousb	Seriousc	Seriousf	1,005	3.51 (1.78, 6.90)	VERY LOW
LR-			Seriousg	Seriousb	Seriousc	Seriousd		0.66 (0.49, 0.87)	VERY LOW
MIXED AXIAL AND PERIPHERAL									
LR+	10 studiesh	Cross-sectional	Seriousg	Seriousb	Seriousc	No serious	2,475	2.98 (2.16, 4.11)	VERY LOW
LR-			Seriousg	Seriousb	Seriousc	Seriousd		0.50 (0.37, 0.69)	VERY LOW
ALL EVIDENCE POOLED									
LR+	30 studiesi	Cross-sectional	Seriousg	Seriousb	Seriousc	No serious	8,125	3.60 (2.95, 4.40)	VERY LOW
LR-			Seriousg	Seriousb	Seriousc	Seriousd		0.48 (0.40, 0.57)	VERY LOW

(a) Braun 2011; Davis 1978; Dougados 2011 (DESIR); Goie The 1985; Hermann 2009; Linssen 1983; Poddubnyy 2011; Sieper 2013; Song 2010; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)
(b) I2 ≥ 50%
(c) >33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)
(d) At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).
(e) Esdaile 1997; Kvien 1994; Kvien 1996; Mattila 1998; McColl 2000; Rohekar 2008; Rudwaleit 2011
(f) At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).

(g) >33.3% of weight in meta-analysis comes from studies with serious risk of bias
(h) Althoff 2009; Brandt 1999; D'Agostino 2011; Godfrin 2004 ; Granfors 1983; Hulsemann 1995; Hulsemann 1995; Liao 2009; Salvarini 2001; Tomero 2014
(i) Althoff 2009; Brandt 1999; Braun 2011; Davis 1978; Dougados 2011 (DESIR); D'Agostino 2011; Esdaile 1997; Godfrin 2004 ; Goie The 1985; Granfors 1983; Hermann 2009; Hulsemann 1995; Hulsemann 1995; Kvien 1994; Kvien 1996; Liao 2009; Linssen 1983; Mattila 1998; McColl 2000; Poddubnyy 2011; Rohekar 2008; Rudwaleit 2011; Salvarini 2001; Sieper 2013; Song 2010; Tomero 2014; van Hoeven 2014; van Hoeven 2015; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)

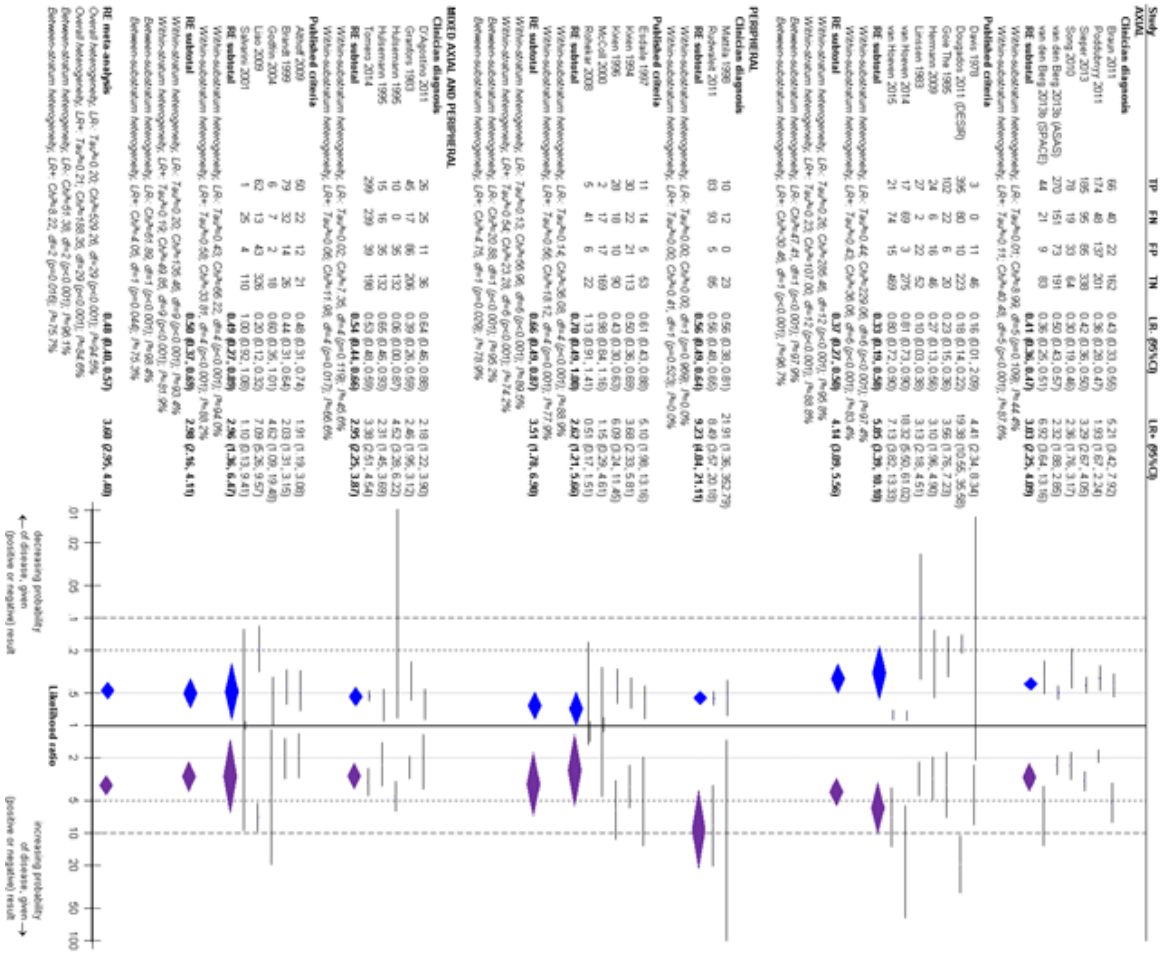


Figure 53 HLA-B27 – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Braun 2011	66	40	22	162	0.62 (0.53, 0.71)	0.88 (0.83, 0.92)
Poddubnyy 2011	174	48	137	201	0.78 (0.72, 0.83)	0.59 (0.54, 0.65)
Sieper 2013	185	95	85	338	0.66 (0.60, 0.71)	0.80 (0.76, 0.83)
Song 2010	78	19	33	64	0.80 (0.71, 0.87)	0.66 (0.56, 0.75)
van den Berg 2013b (ASAS)	270	151	73	191	0.64 (0.59, 0.69)	0.72 (0.67, 0.77)
van den Berg 2013b (SPACE)	44	21	9	83	0.68 (0.55, 0.78)	0.90 (0.82, 0.95)
RE subtotal					**0.70 (0.64, 0.76)**	**0.77 (0.67, 0.85)**
Within-substratum heterogeneity, LR-: Tau^2=0.09; Chi^2=22.31, df=5 (p<0.001); I^2=77.6%						
Within-substratum heterogeneity, LR+: Tau^2=0.36; Chi^2=74.64, df=5 (p<0.001); I^2=93.3%						
Published criteria						
Davis 1978	3	0	11	46	0.88 (0.27, 0.99)	0.80 (0.68, 0.89)
Dougados 2011 (DESIR)	395	80	10	223	0.83 (0.80, 0.86)	0.96 (0.92, 0.98)
Goie The 1985	102	22	6	20	0.82 (0.75, 0.88)	0.77 (0.57, 0.89)
Hermann 2009	24	6	16	46	0.80 (0.62, 0.91)	0.74 (0.62, 0.84)
Linssen 1983	27	2	22	52	0.93 (0.76, 0.98)	0.70 (0.59, 0.80)
van Hoeven 2014	17	69	3	275	0.20 (0.13, 0.30)	0.99 (0.97, 1.00)
van Hoeven 2015	21	74	15	469	0.22 (0.15, 0.32)	0.97 (0.95, 0.98)
RE subtotal					**0.69 (0.39, 0.88)**	**0.90 (0.77, 0.96)**
Within-substratum heterogeneity, LR-: Tau^2=2.40; Chi^2=195.02, df=6 (p<0.001); I^2=96.9%						
Within-substratum heterogeneity, LR+: Tau^2=1.62; Chi^2=97.67, df=6 (p<0.001); I^2=93.9%						
RE subtotal					**0.68 (0.57, 0.78)**	**0.85 (0.77, 0.90)**
Within-stratum heterogeneity, LR-: Tau^2=0.65; Chi^2=218.44, df=12 (p<0.001); I^2=94.5%						
Within-stratum heterogeneity, LR+: Tau^2=0.80; Chi^2=229.48, df=12 (p<0.001); I^2=94.8%						
Between-substratum heterogeneity, LR-: Chi^2=1.11, df=1 (p=0.291); I^2=10.2%						
Between-substratum heterogeneity, LR+: Chi^2=57.17, df=1 (p<0.001); I^2=98.3%						
PERIPHERAL						
Clinician diagnosis						
Mattila 1998	10	12	0	23	0.46 (0.27, 0.66)	0.98 (0.74, 1.00)
Rudwaleit 2011	83	93	5	85	0.47 (0.40, 0.55)	0.94 (0.87, 0.98)
RE subtotal					**0.47 (0.40, 0.54)**	**0.95 (0.89, 0.98)**
Within-substratum heterogeneity, LR-: Tau^2=0.00; Chi^2=0.02, df=1 (p=0.892); I^2=0.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.00; Chi^2=0.46, df=1 (p=0.498); I^2=0.0%						
Published criteria						
Esdaile 1997	11	14	5	53	0.44 (0.26, 0.63)	0.91 (0.81, 0.96)
Kvien 1994	30	22	21	113	0.58 (0.44, 0.70)	0.84 (0.77, 0.90)
Kvien 1996	28	18	10	90	0.61 (0.46, 0.74)	0.90 (0.82, 0.95)
McColl 2000	2	17	17	169	0.11 (0.03, 0.34)	0.91 (0.86, 0.94)
Rohekar 2008	5	41	6	22	0.11 (0.05, 0.24)	0.79 (0.60, 0.90)
RE subtotal					**0.35 (0.17, 0.58)**	**0.88 (0.84, 0.91)**
Within-substratum heterogeneity, LR-: Tau^2=1.00; Chi^2=30.67, df=4 (p<0.001); I^2=87.0%						
Within-substratum heterogeneity, LR+: Tau^2=0.06; Chi^2=6.38, df=4 (p=0.172); I^2=37.3%						
RE subtotal					**0.40 (0.28, 0.54)**	**0.89 (0.85, 0.92)**
Within-stratum heterogeneity, LR-: Tau^2=0.42; Chi^2=30.73, df=6 (p<0.001); I^2=80.5%						
Within-stratum heterogeneity, LR+: Tau^2=0.11; Chi^2=11.08, df=6 (p=0.086); I^2=45.9%						
Between-substratum heterogeneity, LR-: Chi^2=0.04, df=1 (p=0.843); I^2=0.0%						
Between-substratum heterogeneity, LR+: Chi^2=4.24, df=1 (p=0.039); I^2=76.4%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	26	25	11	36	0.51 (0.38, 0.64)	0.77 (0.62, 0.87)
Granfors 1983	45	17	86	206	0.73 (0.60, 0.82)	0.71 (0.65, 0.75)
Hulsemann 1995	10	0	35	132	0.95 (0.55, 1.00)	0.79 (0.72, 0.84)
Hulsemann 1995	15	16	35	132	0.48 (0.32, 0.65)	0.79 (0.72, 0.85)
Tomero 2014	299	239	39	198	0.56 (0.51, 0.60)	0.84 (0.78, 0.88)
RE subtotal					**0.59 (0.48, 0.69)**	**0.78 (0.72, 0.83)**
Within-substratum heterogeneity, LR-: Tau^2=0.13; Chi^2=11.59, df=4 (p=0.021); I^2=65.5%						
Within-substratum heterogeneity, LR+: Tau^2=0.08; Chi^2=13.18, df=4 (p=0.010); I^2=69.6%						
Published criteria						
Althoff 2009	50	22	12	21	0.69 (0.58, 0.79)	0.64 (0.46, 0.78)
Brandt 1999	79	32	14	26	0.71 (0.62, 0.79)	0.65 (0.49, 0.78)
Godfrin 2004	6	7	2	18	0.46 (0.22, 0.72)	0.90 (0.68, 0.97)
Liao 2009	62	13	43	326	0.83 (0.72, 0.90)	0.88 (0.85, 0.91)
Salvarini 2001	1	25	4	110	0.04 (0.01, 0.23)	0.96 (0.91, 0.99)
RE subtotal					**0.61 (0.42, 0.77)**	**0.84 (0.68, 0.93)**
Within-substratum heterogeneity, LR-: Tau^2=0.57; Chi^2=24.68, df=4 (p<0.001); I^2=83.8%						
Within-substratum heterogeneity, LR+: Tau^2=0.92; Chi^2=35.08, df=4 (p<0.001); I^2=88.6%						
RE subtotal					**0.62 (0.51, 0.71)**	**0.80 (0.74, 0.86)**
Within-stratum heterogeneity, LR-: Tau^2=0.32; Chi^2=61.41, df=9 (p<0.001); I^2=82.6%						
Within-stratum heterogeneity, LR+: Tau^2=0.28; Chi^2=59.66, df=9 (p<0.001); I^2=84.9%						
Between-substratum heterogeneity, LR-: Chi^2=15.14, df=1 (p<0.001); I^2=93.4%						
Between-substratum heterogeneity, LR+: Chi^2=11.41, df=1 (p<0.001); I^2=91.2%						
RE meta-analysis					**0.60 (0.52, 0.67)**	**0.85 (0.80, 0.88)**
Overall heterogeneity, LR-: Tau^2=0.55; Chi^2=368.47, df=29 (p<0.001); I^2=92.1%						
Overall heterogeneity, LR+: Tau^2=0.56; Chi^2=336.43, df=29 (p<0.001); I^2=91.4%						
Between-stratum heterogeneity, LR-: Chi^2=67.89, df=2 (p<0.001); I^2=97.1%						
Between-stratum heterogeneity, LR+: Chi^2=36.21, df=2 (p<0.001); I^2=94.5%						

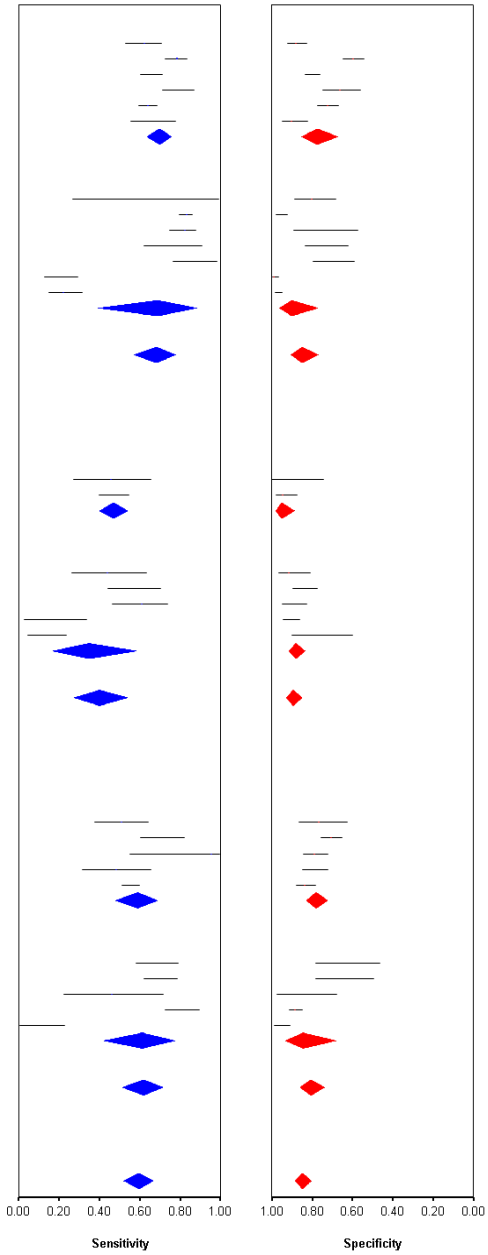


Figure 54: HLA-B27 – forest plot: sensitivity and specificity

G.1.5.2 ESR

Table 31: GRADE table for ESR

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 studya	Cross-sectional	Serious	n/a	Seriousb	Seriousc	92	1.72 (0.84, 3.53)	VERY LOW
LR-			Serious	n/a	Seriousb	No serious		0.83 (0.62, 1.09)	LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 studyd	Cross-sectional	No serious	n/a	No serious	No serious	775	3.52 (2.07, 6.01)	HIGH
LR-			No serious	n/a	No serious	No serious		0.84 (0.80, 0.89)	HIGH
ALL EVIDENCE POOLED									
LR+	2 studiese	Cross-sectional	Seriousf	Seriousg	Serioush	Seriousc	867	2.57 (1.28, 5.16)	VERY LOW
LR-			No serious	No serious	No serious	No serious		0.84 (0.80, 0.89)	HIGH

(a) Hermann 2009
(b) suboptimal reference standard (published classification criteria, rather than expert diagnosis)
(c) At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).
(d) Tomero 2014
(e) Hermann 2009; Tomero 2014
(f) >33.3% of weight in meta-analysis comes from studies with serious risk of bias
(g) I2 ≥ 50%
(h) >33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
Hermann 2009	10	20	12	50	0.83 (0.62, 1.09)	1.72 (0.84, 3.53)
RE subtotal					**0.83 (0.62, 1.09)**	**1.72 (0.84, 3.53)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Tomero 2014	112	426	14	223	0.84 (0.80, 0.89)	3.52 (2.07, 6.01)
Published criteria						
no data						
RE subtotal					**0.84 (0.80, 0.89)**	**3.52 (2.07, 6.01)**
RE meta-analysis					**0.84 (0.80, 0.89)**	**2.57 (1.28, 5.16)**

Overall heterogeneity, LR-: Tau²=0.00; Chi²=0.01, df=1 (p=0.903); I²=0.0%
Overall heterogeneity, LR+: Tau²=0.15; Chi²=2.46, df=1 (p=0.117); I²=59.4%
Between-stratum heterogeneity, LR-: Chi²=0.01, df=1 (p=0.903); I²=0.0%
Between-stratum heterogeneity, LR+: Chi²=2.46, df=1 (p=0.117); I²=59.4%

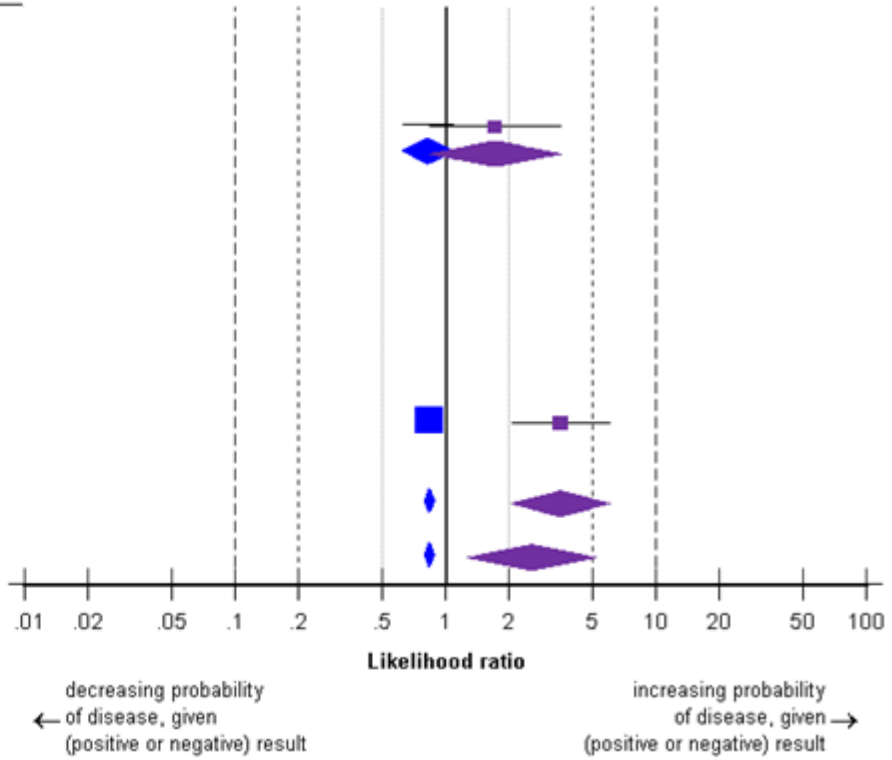


Figure 55 ESR – forest plot: likelihood ratios

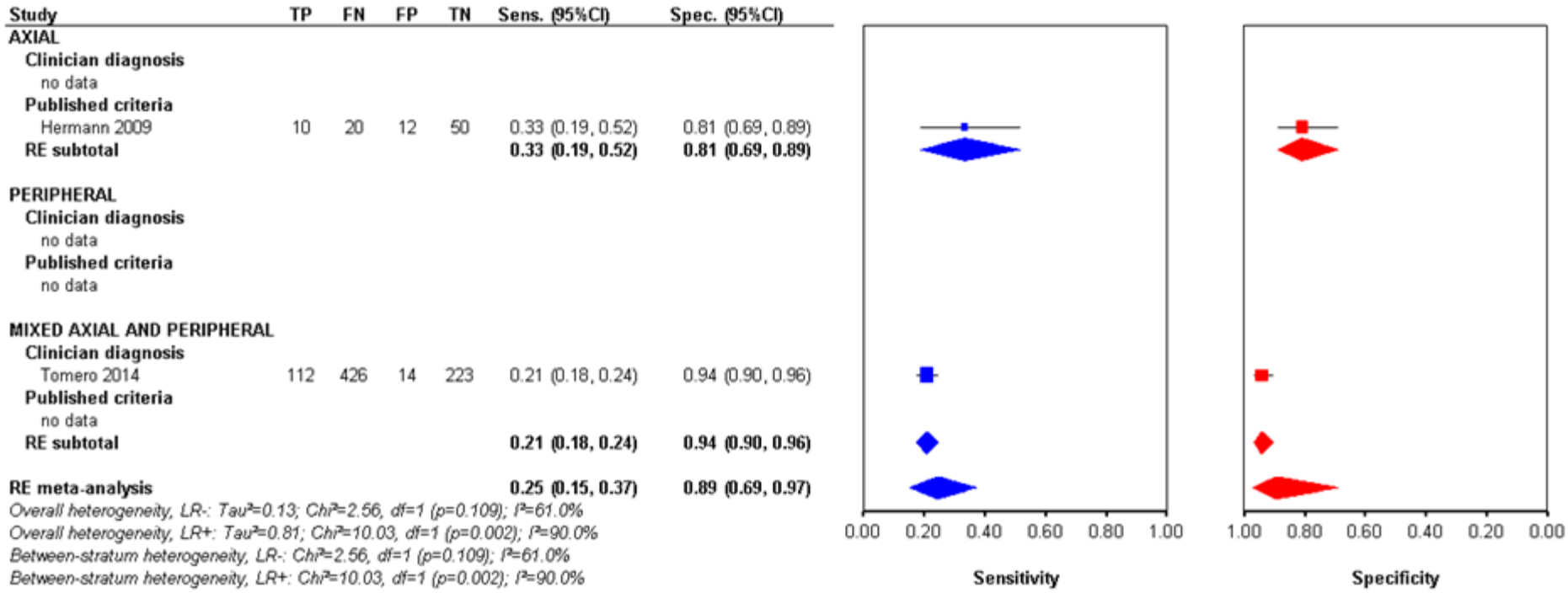


Figure 56 ESR – forest plot: sensitivity and specificity

G.1.5.3 CRP

Table 32 GRADE table for ESR

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	5 studiesa	Cross-sectional	No serious	Seriousb	Seriousc	Seriousd	2,389	1.88 (0.91, 3.87)	VERY LOW
LR-			No serious	Seriousb	Seriousc	No serious		0.94 (0.79, 1.12)	LOW
PERIPHERAL									
LR+	2 studiese	Cross-sectional	No serious	No serious	Seriousc	No serious	412	1.51 (1.17, 1.95)	MODERATE
LR-			No serious	No serious	No serious	Seriousf		0.65 (0.45, 0.93)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	1 studyg	Cross-sectional	No serious	n/a	No serious	No serious	775	1.24 (0.92, 1.67)	HIGH
LR-			No serious	n/a	No serious	No serious		0.94 (0.87, 1.02)	HIGH
ALL EVIDENCE POOLED									
LR+	8 studiesh	Cross-sectional	No serious	Seriousb	Seriousc	Seriousd	3,576	1.63 (1.11, 2.41)	VERY LOW
LR-			No serious	Seriousb	Seriousc	No serious		0.89 (0.78, 1.00)	LOW

(a) Dougados 2011 (DESIR); Hermann 2009; Rudwaleit 2009 (ASAS); van Hoeven 2014; van Hoeven 2015
(b) I2 ≥ 50%
(c) >33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)
(d) At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).
(e) Kvien 1996; Rudwaleit 2011
(f) At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).
(g) Tomero 2014
(h) Dougados 2011 (DESIR); Hermann 2009; Kvien 1996; Rudwaleit 2009 (ASAS); Rudwaleit 2011; Tomero 2014; van Hoeven 2014; van Hoeven 2015

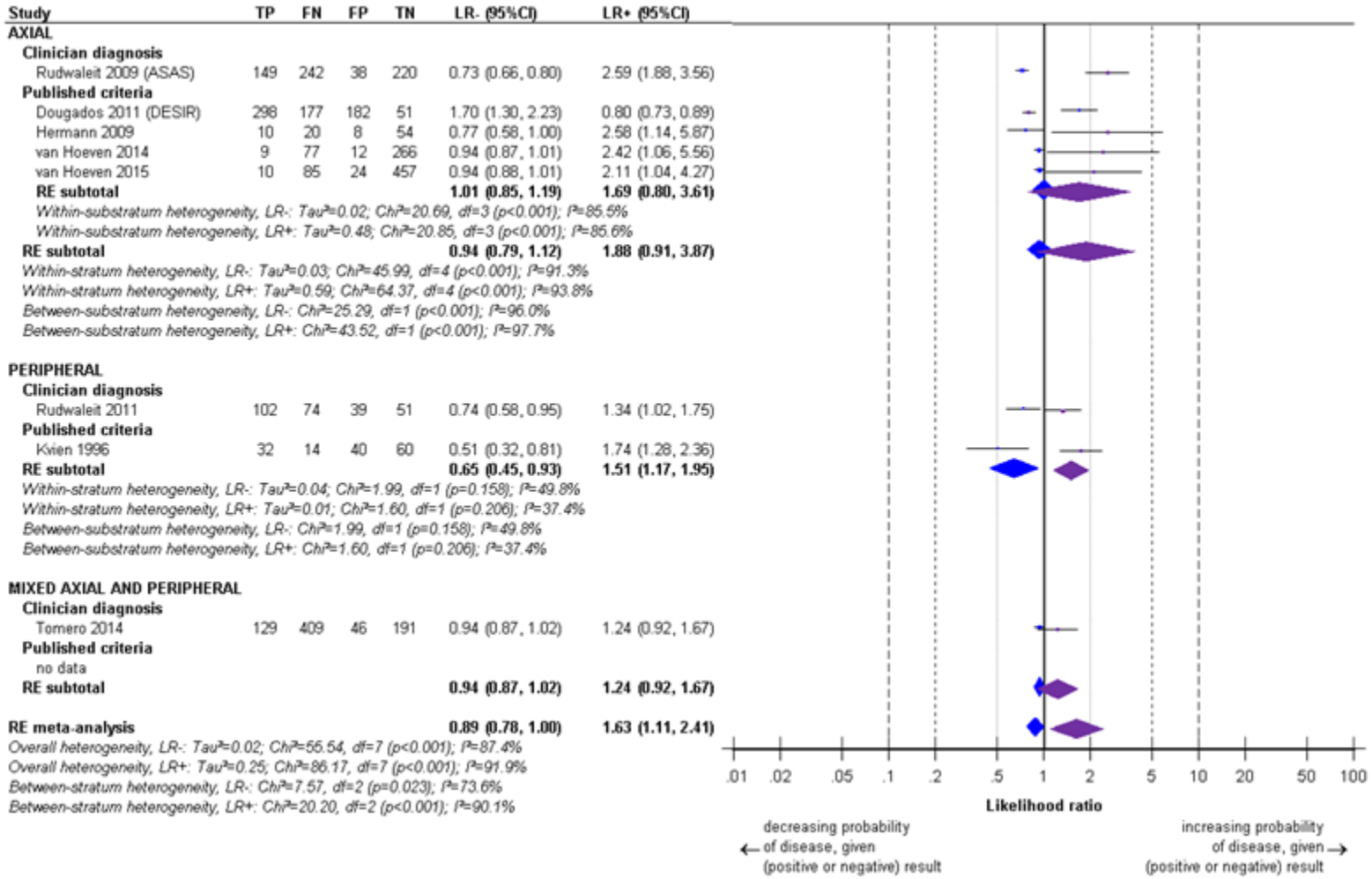


Figure 57 CRP – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Rudwaleit 2009 (ASAS)	149	242	38	220	0.38 (0.33, 0.43)	0.85 (0.80, 0.89)
Published criteria						
Dougados 2011 (DESIR)	298	177	182	51	0.63 (0.58, 0.67)	0.22 (0.17, 0.28)
Hermann 2009	10	20	8	54	0.33 (0.19, 0.52)	0.87 (0.76, 0.93)
van Hoeven 2014	9	77	12	266	0.10 (0.06, 0.19)	0.96 (0.93, 0.98)
van Hoeven 2015	10	85	24	457	0.11 (0.06, 0.18)	0.95 (0.93, 0.97)
RE subtotal					**0.25 (0.06, 0.62)**	**0.84 (0.30, 0.98)**
Within-substratum heterogeneity, LR-: Tau²=2.55; Chi²=109.36, df=3 (p<0.001); I²=97.3%						
Within-substratum heterogeneity, LR+: Tau²=6.47; Chi²=345.37, df=3 (p<0.001); I²=99.1%						
RE subtotal					**0.28 (0.14, 0.48)**	**0.84 (0.46, 0.97)**
Within-stratum heterogeneity, LR-: Tau²=0.94; Chi²=129.70, df=4 (p<0.001); I²=96.9%						
Within-stratum heterogeneity, LR+: Tau²=4.31; Chi²=366.25, df=4 (p<0.001); I²=98.9%						
Between-substratum heterogeneity, LR-: Chi²=20.34, df=1 (p<0.001); I²=95.1%						
Between-substratum heterogeneity, LR+: Chi²=20.88, df=1 (p<0.001); I²=95.2%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	102	74	39	51	0.58 (0.51, 0.65)	0.57 (0.46, 0.66)
Published criteria						
Kvien 1996	32	14	40	60	0.70 (0.55, 0.81)	0.60 (0.50, 0.69)
RE subtotal					**0.62 (0.51, 0.72)**	**0.58 (0.51, 0.65)**
Within-stratum heterogeneity, LR-: Tau²=0.06; Chi²=2.03, df=1 (p=0.154); I²=50.7%						
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=0.22, df=1 (p=0.642); I²=0.0%						
Between-substratum heterogeneity, LR-: Chi²=2.03, df=1 (p=0.154); I²=50.7%						
Between-substratum heterogeneity, LR+: Chi²=0.22, df=1 (p=0.642); I²=0.0%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Tomero 2014	129	409	46	191	0.24 (0.21, 0.28)	0.81 (0.75, 0.85)
Published criteria						
no data						
RE subtotal					**0.24 (0.21, 0.28)**	**0.81 (0.75, 0.85)**
RE meta-analysis					**0.36 (0.23, 0.51)**	**0.79 (0.56, 0.91)**
Overall heterogeneity, LR-: Tau²=0.79; Chi²=237.10, df=7 (p<0.001); I²=97.0%						
Overall heterogeneity, LR+: Tau²=2.31; Chi²=393.72, df=7 (p<0.001); I²=98.2%						
Between-stratum heterogeneity, LR-: Chi²=105.38, df=2 (p<0.001); I²=98.1%						
Between-stratum heterogeneity, LR+: Chi²=27.26, df=2 (p<0.001); I²=92.7%						

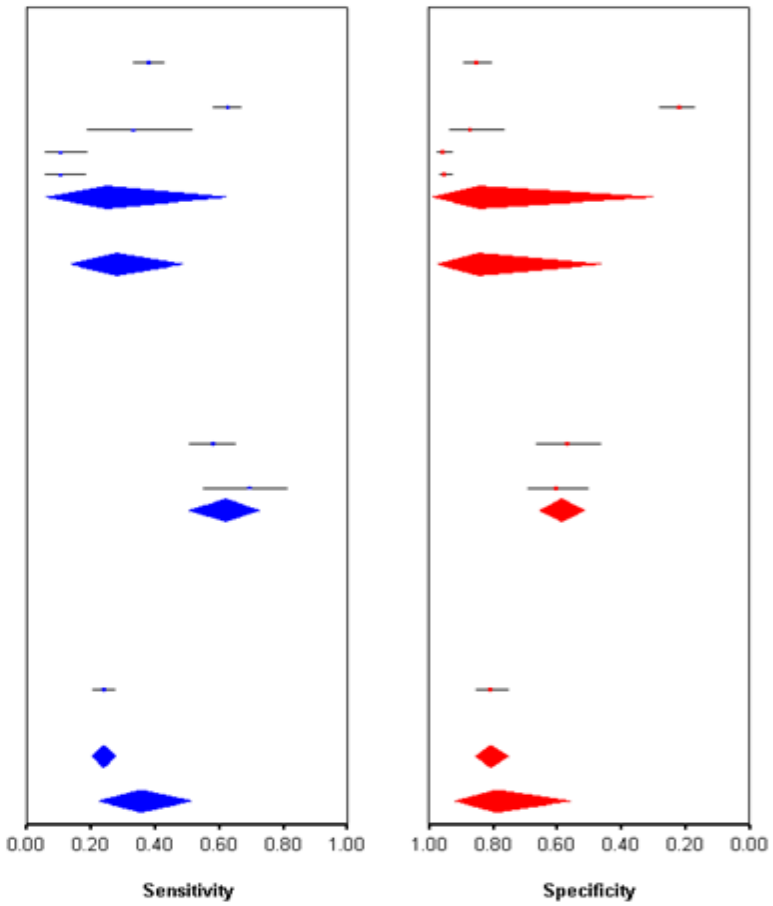


Figure 58 CRP – forest plot: sensitivity and specificity

G.1.6 Imaging for diagnosis of spondyloarthritis

Review Question 10

- What is the diagnostic utility of imaging (alone or in sequence) for investigating suspected spondyloarthritis?

G.1.6.1 X-ray

Sacroiliitis on x-ray

Table 33: Sacroiliitis on x-ray – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	3 studies[a]	Cross-sectional	No serious	Serious[b]	No serious	No serious	1,550	18.22 (4.12, 80.69)	MODERATE
LR-			No serious	Serious[b]	Serious[c]	No serious		0.72 (0.62, 0.85)	LOW
PERIPHERAL									
LR+	5 studies[d]	Cross-sectional	Serious[e]	Serious[b]	Serious[c]	No serious	751	6.84 (2.47, 18.89)	VERY LOW
LR-			No serious	Serious[b]	Serious[c]	No serious		0.75 (0.60, 0.94)	LOW
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[f]	Cross-sectional	No serious	n/a	No serious	No serious	775	89.64 (5.59, 1436.83)	HIGH
LR-			No serious	n/a	No serious	No serious		0.81 (0.78, 0.85)	HIGH
ALL EVIDENCE POOLED									
LR+	9 studies[g]	Cross-sectional	No serious	Serious[b]	No serious	No serious	3,076	10.15 (5.10, 20.23)	MODERATE
LR-			No serious	Serious[b]	Serious[c]	No serious		0.76 (0.68, 0.84)	LOW

[a] *Dougados 2011 (DESIR); van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *I2 ≥ 50%*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *Esdaile 1997; Rigby 1993; Rudwaleit 2011; Sadek 2007; You 2015*
[e] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*

[f] *Tomero 2014*

[g] *Dougados 2011 (DESIR); Esdaile 1997; Rigby 1993; Rudwaleit 2011; Sadek 2007; Tomero 2014; You 2015; van den Berg 2013b (ASAS) ; van den Berg 2013b (SPACE)*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	123	298	9	255	0.73 (0.69, 0.78)	8.57 (4.43, 16.57)
van den Berg 2013b (SPACE)	11	54	1	91	0.84 (0.75, 0.94)	15.57 (2.06, 117.64)
RE subtotal					**0.78 (0.68, 0.89)**	**9.08 (4.85, 16.98)**
Within-substratum heterogeneity, LR-: Tau²=0.01; Chi²=4.26, df=1 (p=0.039); I²=76.5%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.30, df=1 (p=0.582); I²=0.0%						
Published criteria						
Dougados 2011 (DESIR)	181	294	0	233	0.62 (0.58, 0.67)	178.45 (11.17, 2851.25)
RE subtotal					**0.72 (0.62, 0.85)**	**18.22 (4.12, 80.69)**
Within-stratum heterogeneity, LR-: Tau²=0.02; Chi²=23.37, df=2 (p<0.001); I²=91.4%						
Within-stratum heterogeneity, LR+: Tau²=0.98; Chi²=4.53, df=2 (p=0.104); I²=55.8%						
Between-substratum heterogeneity, LR-: Chi²=19.11, df=1 (p<0.001); I²=94.8%						
Between-substratum heterogeneity, LR+: Chi²=4.22, df=1 (p=0.040); I²=76.3%						
PERIPHERAL						
Clinician diagnosis						
Rigby 1993	28	2	17	165	0.07 (0.02, 0.28)	9.99 (6.29, 15.87)
Rudwaleit 2011	32	132	2	61	0.83 (0.76, 0.91)	6.15 (1.52, 24.89)
Sadek 2007	23	36	0	22	0.62 (0.50, 0.77)	18.02 (1.14, 284.54)
RE subtotal					**0.59 (0.38, 0.92)**	**9.68 (6.27, 14.93)**
Within-substratum heterogeneity, LR-: Tau²=0.11; Chi²=18.26, df=2 (p<0.001); I²=89.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.62, df=2 (p=0.734); I²=0.0%						
Published criteria						
Esdaile 1997	3	22	6	52	0.98 (0.83, 1.16)	1.16 (0.31, 4.27)
You 2015	4	14	1	129	0.78 (0.61, 1.00)	28.89 (3.42, 244.35)
RE subtotal					**0.89 (0.72, 1.11)**	**5.16 (0.22, 119.43)**
Within-substratum heterogeneity, LR-: Tau²=0.01; Chi²=2.16, df=1 (p=0.141); I²=53.8%						
Within-substratum heterogeneity, LR+: Tau²=4.35; Chi²=6.34, df=1 (p=0.012); I²=84.2%						
RE subtotal					**0.75 (0.60, 0.94)**	**6.84 (2.47, 18.89)**
Within-stratum heterogeneity, LR-: Tau²=0.05; Chi²=23.55, df=4 (p<0.001); I²=83.0%						
Within-stratum heterogeneity, LR+: Tau²=0.76; Chi²=11.15, df=4 (p=0.025); I²=64.1%						
Between-substratum heterogeneity, LR-: Chi²=3.13, df=1 (p=0.077); I²=68.0%						
Between-substratum heterogeneity, LR+: Chi²=4.19, df=1 (p=0.041); I²=76.1%						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
Tomero 2014	101	437	0	237	0.81 (0.78, 0.85)	89.64 (5.59, 1436.83)
Published criteria						
no data						
RE subtotal					**0.81 (0.78, 0.85)**	**89.64 (5.59, 1436.83)**
RE meta-analysis					**0.76 (0.68, 0.84)**	**10.15 (5.10, 20.23)**
Overall heterogeneity, LR-: Tau²=0.02; Chi²=73.91, df=8 (p<0.001); I²=89.2%						
Overall heterogeneity, LR+: Tau²=0.48; Chi²=18.75, df=8 (p=0.016); I²=57.3%						
Between-stratum heterogeneity, LR-: Chi²=26.99, df=2 (p<0.001); I²=92.6%						
Between-stratum heterogeneity, LR+: Chi²=3.07, df=2 (p=0.216); I²=34.8%						

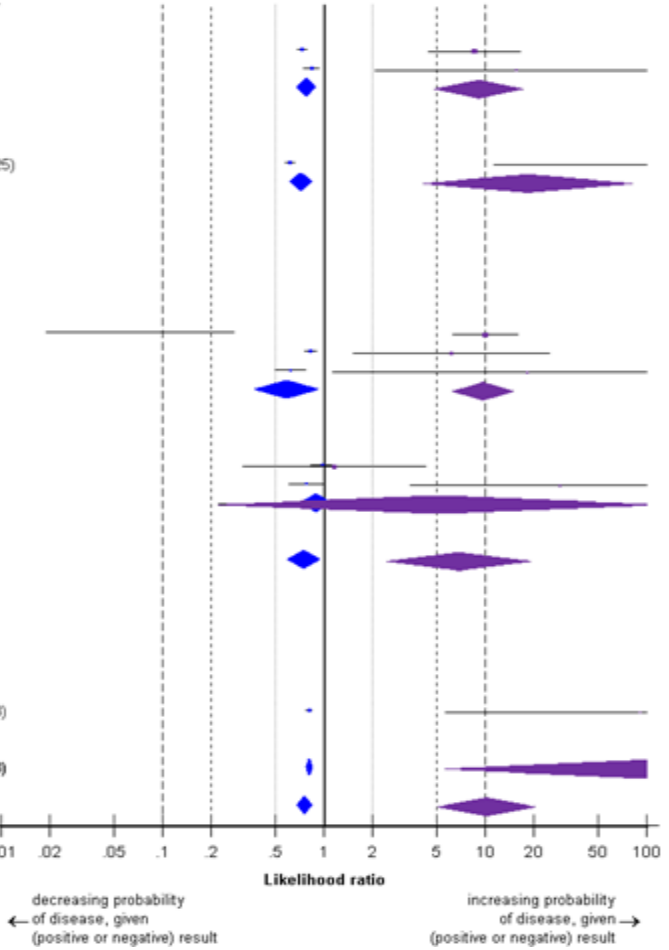


Figure 59: Sacroiliitis on x-ray – forest plot: likelihood ratios

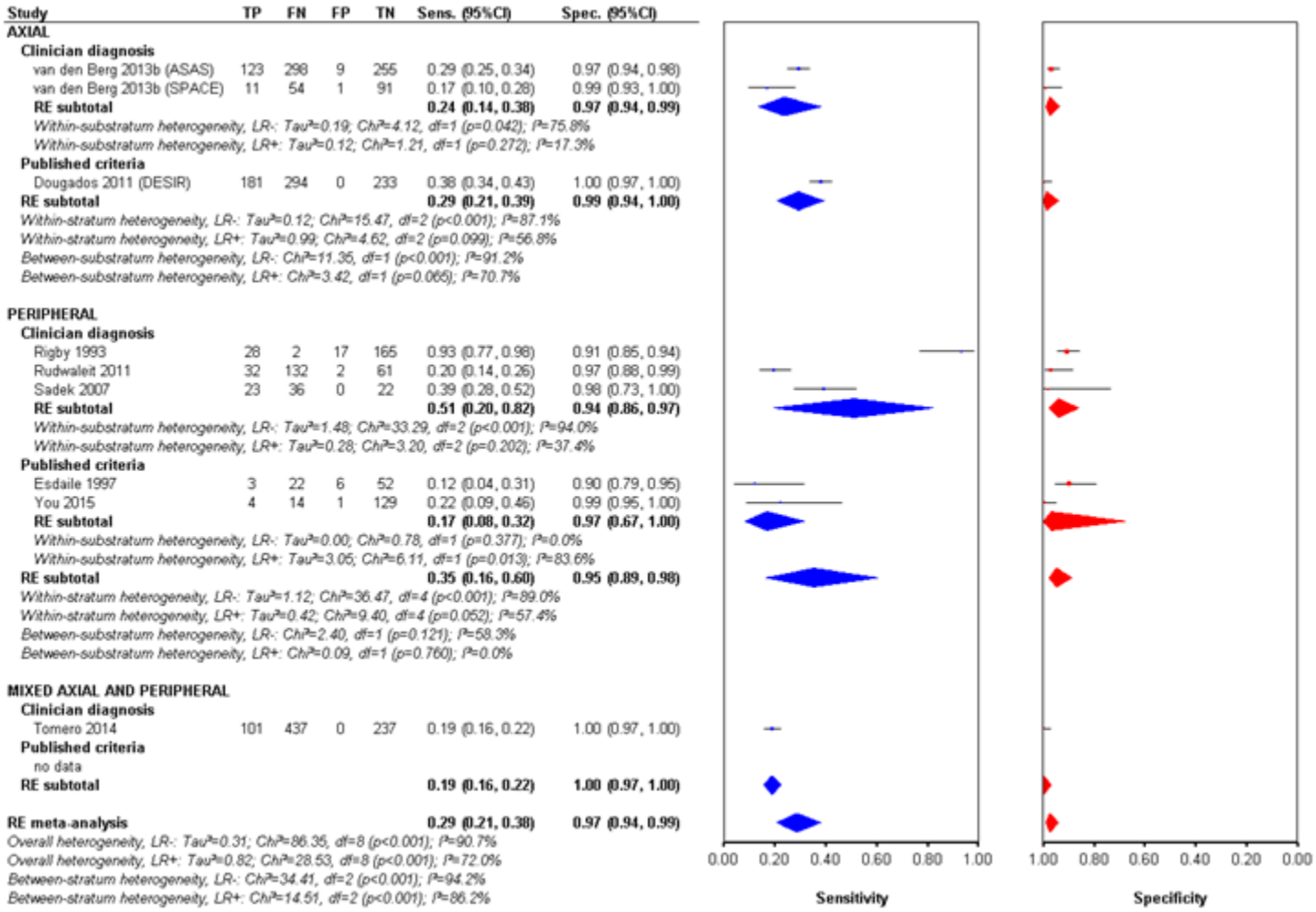


Figure 60 Sacroiliitis on x-ray – forest plot: sensitivity and specificity

Finger or toe pathology on x-ray

Table 34: Finger or toe pathology on x-ray – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	Serious[b]	52	10.57 (0.66, 169.08)	MODERATE
LR-			No serious	n/a	No serious	No serious		0.71 (0.56, 0.90)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	Serious[b]	52	10.57 (0.66, 169.08)	MODERATE
LR-			No serious	n/a	No serious	No serious		0.71 (0.56, 0.90)	HIGH

[a] *De Simone 2011*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

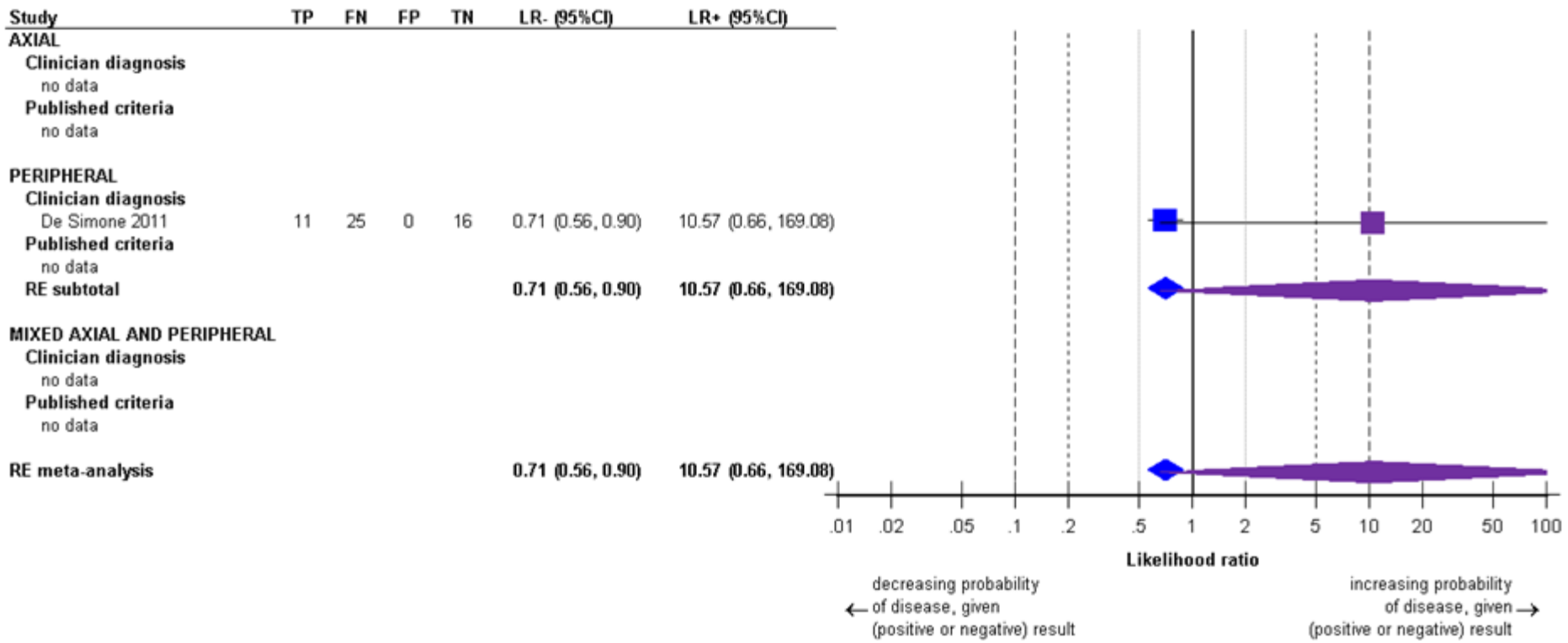


Figure 61: Finger or toe pathology on x-ray – forest plot: likelihood ratios

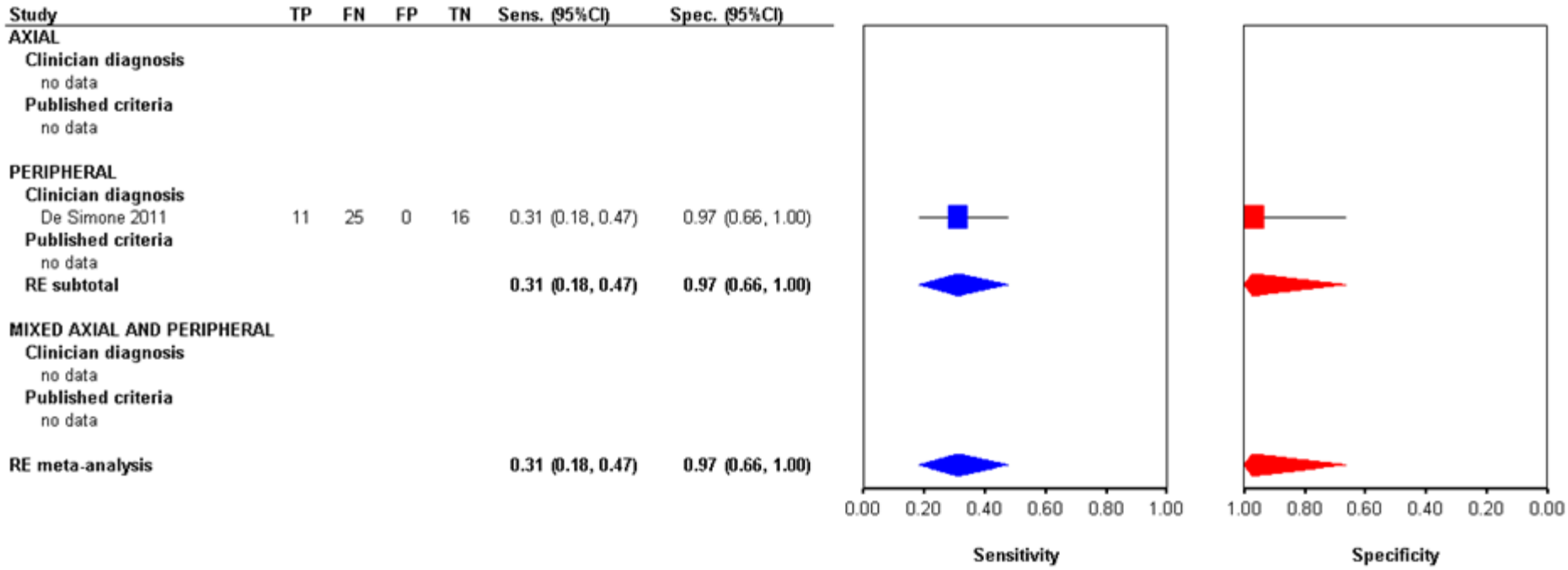


Figure 62: Finger or toe pathology on x-ray – forest plot: sensitivity and specificity

Enthesitis on x-ray

Table 35: Enthesitis on x-ray – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	Serious[b]	81	1.57 (0.92, 2.69)	MODERATE
LR-			No serious	n/a	No serious	Serious[c]		0.60 (0.37, 0.98)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[d]	Cross-sectional	No serious	n/a	Serious[e]	Serious[b]	33	25.50 (1.60, 407.29)	LOW
LR-			No serious	n/a	Serious[e]	Serious[c]		0.40 (0.21, 0.77)	LOW
ALL EVIDENCE POOLED									
LR+	2 studies[f]	Cross-sectional	No serious	Serious[g]	Serious[h]	V. serious[i]	114	4.49 (0.32, 63.10)	VERY LOW
LR-			No serious	No serious	Serious[h]	Serious[c]		0.52 (0.35, 0.77)	LOW

[a] *Sadek 2007*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[d] *Godfrin 2004*
[e] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[f] *Godfrin 2004 ; Sadek 2007*
[g] *I2 ≥ 50%*
[h] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[i] *At a 95% confidence level, data are consistent with meaningful predictive value in either direction and no predictive value at all (i.e. 95% CI for LR+ spans both 0.5 and 2).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
PERIPHERAL						
Clinician diagnosis						
Sadek 2007	38	21	9	13	0.60 (0.37, 0.98)	1.57 (0.92, 2.69)
Published criteria						
no data						
RE subtotal					**0.60 (0.37, 0.98)**	**1.57 (0.92, 2.69)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
Godfrin 2004	8	5	0	20	0.40 (0.21, 0.77)	25.50 (1.60, 407.29)
RE subtotal					**0.40 (0.21, 0.77)**	**25.50 (1.60, 407.29)**
RE meta-analysis					**0.52 (0.35, 0.77)**	**4.49 (0.32, 63.10)**

Overall heterogeneity, LR-: Tau²=0.00; Chi²=0.94, df=1 (p=0.333); I²=0.0%
Overall heterogeneity, LR+: Tau²=2.84; Chi²=3.74, df=1 (p=0.053); I²=73.3%
Between-stratum heterogeneity, LR-: Chi²=0.94, df=1 (p=0.333); I²=0.0%
Between-stratum heterogeneity, LR+: Chi²=3.74, df=1 (p=0.053); I²=73.3%

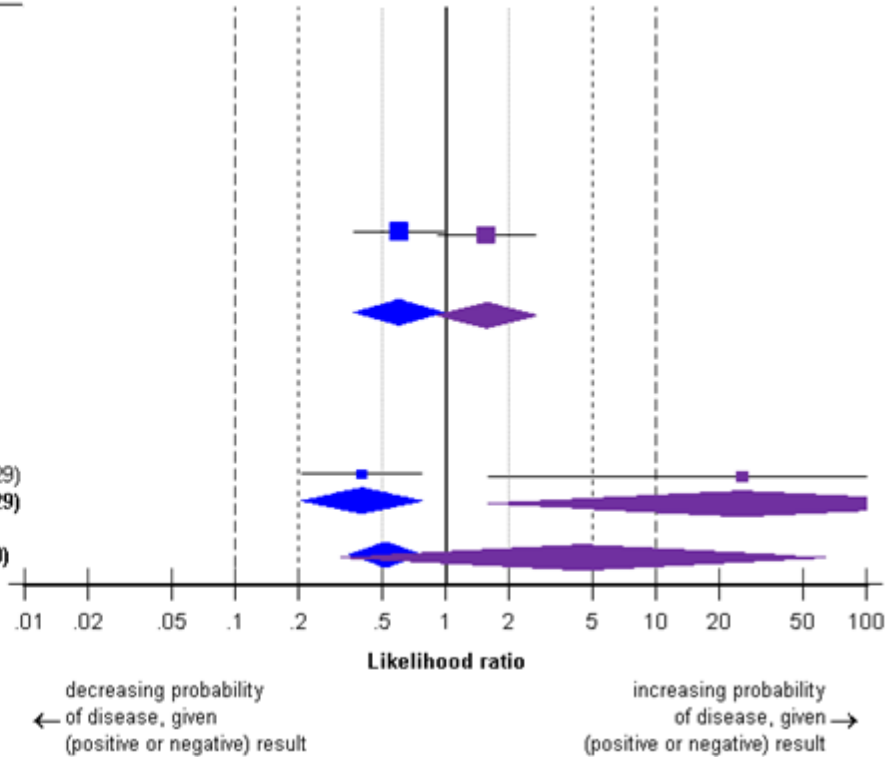


Figure 63: Enthesitis on x-ray – forest plot: likelihood ratios

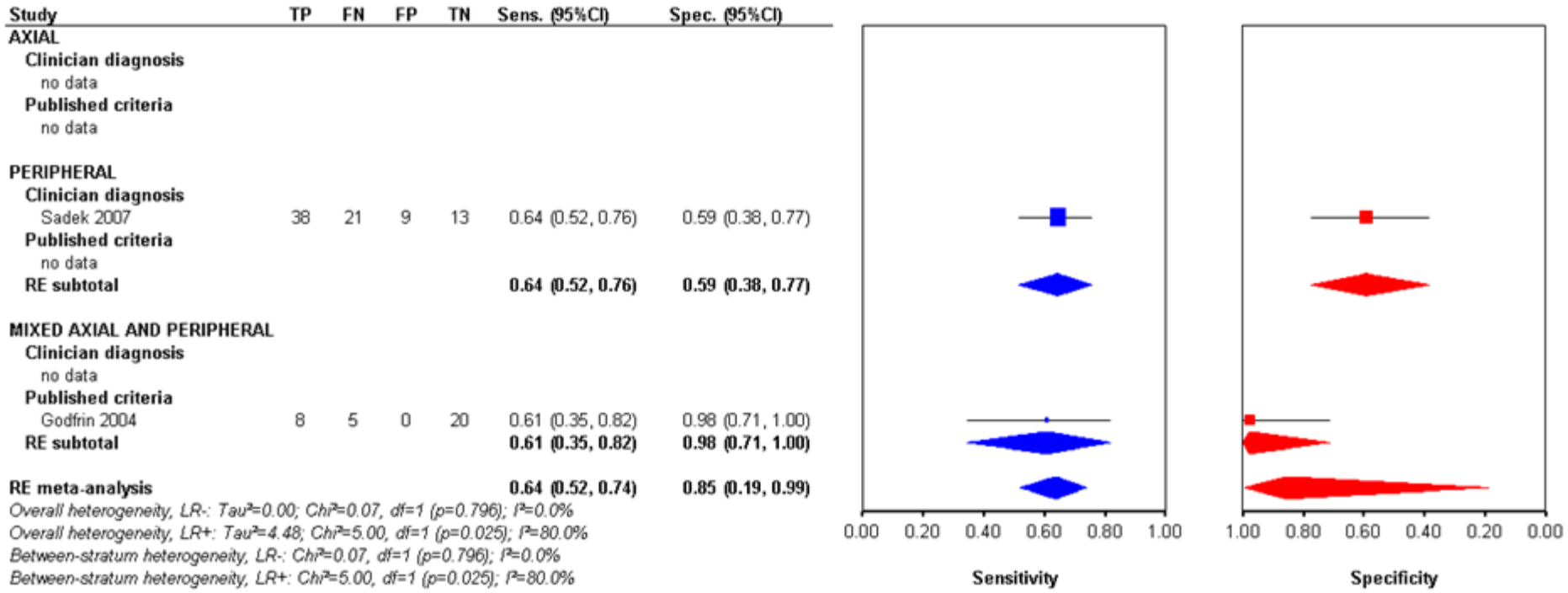


Figure 64: Enthesitis on x-ray – forest plot: sensitivity and specificity

G.1.6.2 MRI

Sacroiliitis on MRI

Table 36: Sacroiliitis on MRI – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	3 studies[a]	Cross-sectional	No serious	Serious[b]	No serious	No serious	1,550	41.49 (7.72, 223.02)	MODERATE
LR-			No serious	No serious	Serious[c]	No serious		0.54 (0.50, 0.57)	MODERATE
PERIPHERAL									
LR+	1 study[d]	Cross-sectional	No serious	n/a	No serious	Serious[e]	60	9.71 (0.64, 148.17)	MODERATE
LR-			No serious	n/a	No serious	Serious[f]		0.59 (0.44, 0.77)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[g]	Cross-sectional	No serious	n/a	No serious	Serious[e]	73	4.07 (1.28, 12.97)	MODERATE
LR-			No serious	n/a	No serious	No serious		0.70 (0.54, 0.91)	HIGH
ALL EVIDENCE POOLED									
LR+	5 studies[h]	Cross-sectional	No serious	Serious[b]	No serious	No serious	1,683	16.96 (5.29, 54.40)	MODERATE
LR-			No serious	No serious	Serious[c]	No serious		0.55 (0.51, 0.59)	MODERATE

[a] *Dougados 2011 (DESIR); van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *I2 ≥ 50%*
[c] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[d] *Rudwaleit 2011*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[g] *D'Agostino 2011*
[h] *Dougados 2011 (DESIR); D'Agostino 2011; Rudwaleit 2011; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

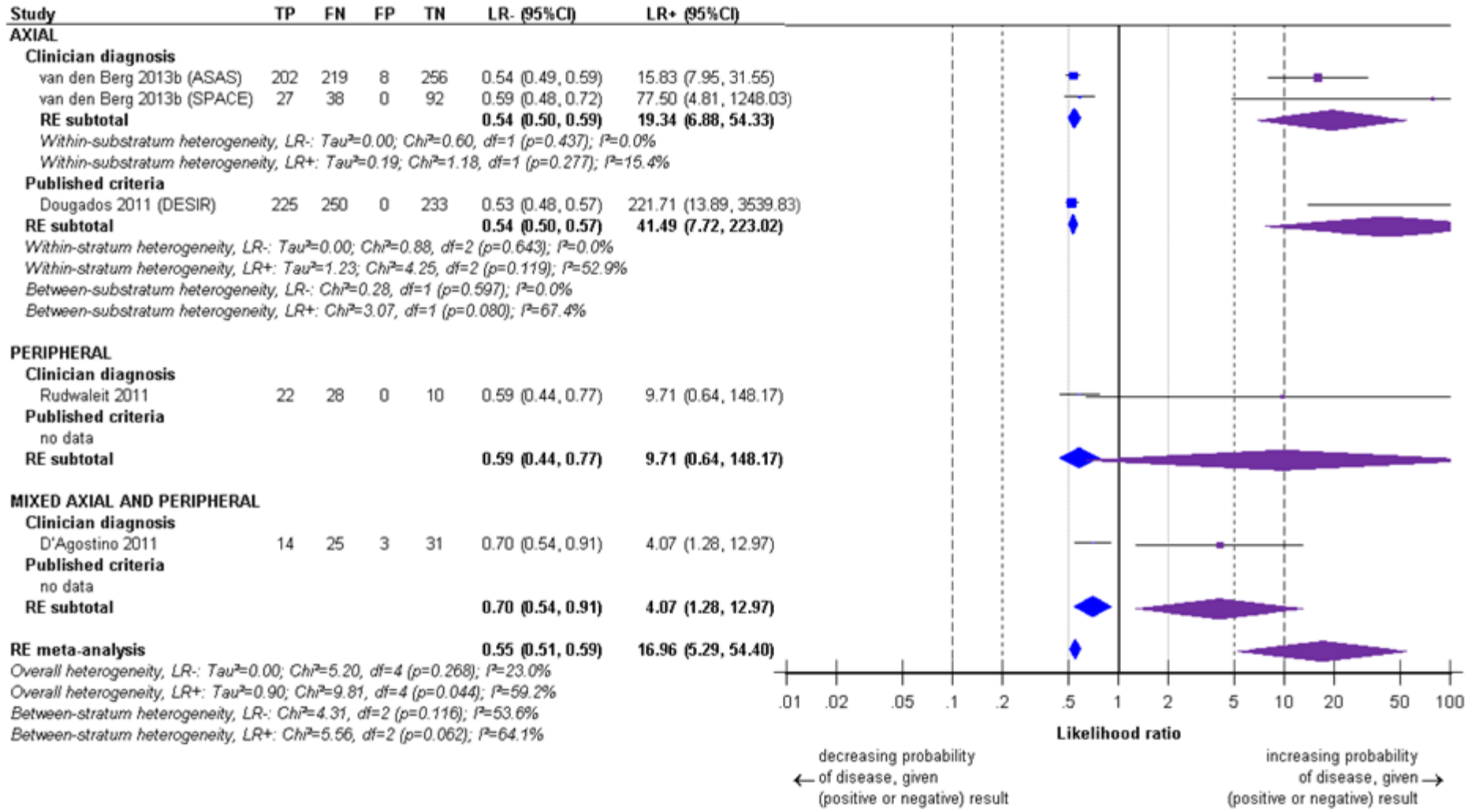


Figure 65 Sacroiliitis on MRI – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	202	219	8	256	0.48 (0.43, 0.53)	0.97 (0.94, 0.98)
van den Berg 2013b (SPACE)	27	38	0	92	0.42 (0.30, 0.54)	0.99 (0.92, 1.00)
RE subtotal					**0.47 (0.43, 0.52)**	**0.98 (0.92, 0.99)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.91, df=1 (p=0.340); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.47; Chi²=1.44, df=1 (p=0.230); I²=30.5%						
Published criteria						
Dougados 2011 (DESIR)	225	250	0	233	0.47 (0.43, 0.52)	1.00 (0.97, 1.00)
RE subtotal					**0.47 (0.44, 0.50)**	**0.99 (0.94, 1.00)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.92, df=2 (p=0.633); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=1.40; Chi²=4.56, df=2 (p=0.102); I²=56.1%						
Between-substratum heterogeneity, LR-: Chi²=0.01, df=1 (p=0.941); I²=0.0%						
Between-substratum heterogeneity, LR+: Chi²=3.12, df=1 (p=0.077); I²=67.9%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	22	28	0	10	0.44 (0.31, 0.58)	0.95 (0.55, 1.00)
Published criteria						
no data						
RE subtotal					**0.44 (0.31, 0.58)**	**0.95 (0.55, 1.00)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	14	25	3	31	0.36 (0.23, 0.52)	0.91 (0.76, 0.97)
Published criteria						
no data						
RE subtotal					**0.36 (0.23, 0.52)**	**0.91 (0.76, 0.97)**
RE meta-analysis					**0.47 (0.44, 0.50)**	**0.97 (0.92, 0.99)**
Overall heterogeneity, LR-: Tau²=0.00; Chi²=2.97, df=4 (p=0.563); I²=0.0%						
Overall heterogeneity, LR+: Tau²=0.74; Chi²=8.60, df=4 (p=0.072); I²=53.5%						
Between-stratum heterogeneity, LR-: Chi²=2.05, df=2 (p=0.358); I²=2.6%						
Between-stratum heterogeneity, LR+: Chi²=4.04, df=2 (p=0.133); I²=50.5%						

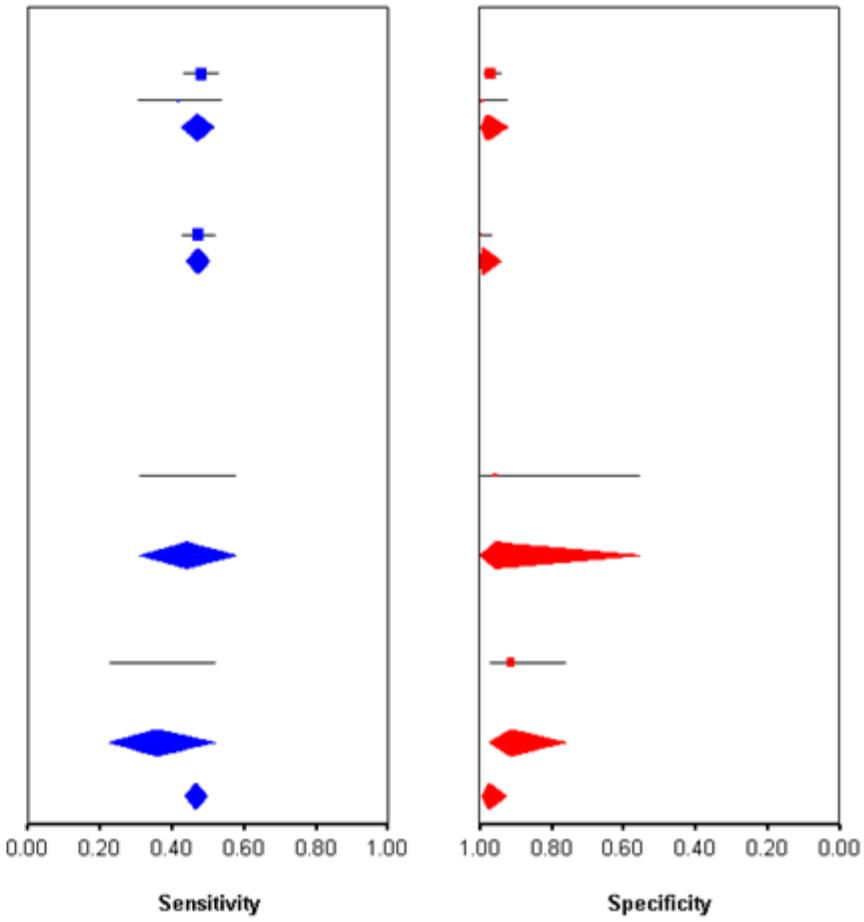


Figure 66 Sacroiliitis on MRI – forest plot: sensitivity and specificity

Spinal features on MRI

Table 37 Spinal features on MRI – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	708	2.70 (1.76, 4.13)	LOW
LR-			No serious	n/a	Serious[b]	No serious		0.82 (0.77, 0.88)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	708	2.70 (1.76, 4.13)	LOW
LR-			No serious	n/a	Serious[b]	No serious		0.82 (0.77, 0.88)	MODERATE

[a] *Dougados 2011 (DESIR)*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
Dougados 2011 (DESIR)	121	354	22	211	0.82 (0.77, 0.88)	2.70 (1.76, 4.13)
RE subtotal					**0.82 (0.77, 0.88)**	**2.70 (1.76, 4.13)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.82 (0.77, 0.88)**	**2.70 (1.76, 4.13)**

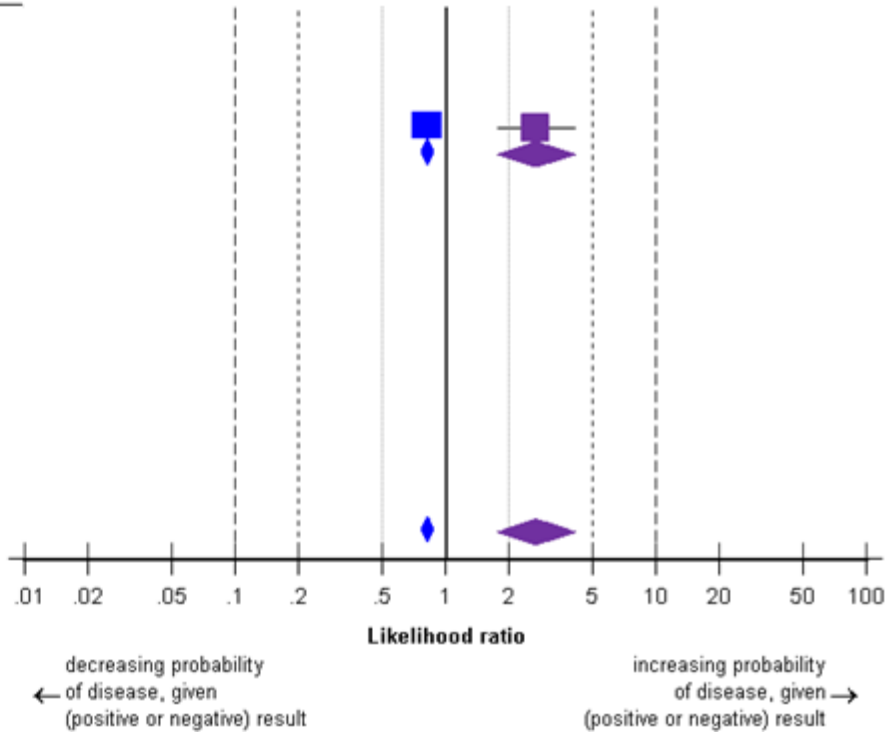


Figure 67 Spinal features on MRI – forest plot: likelihood ratios

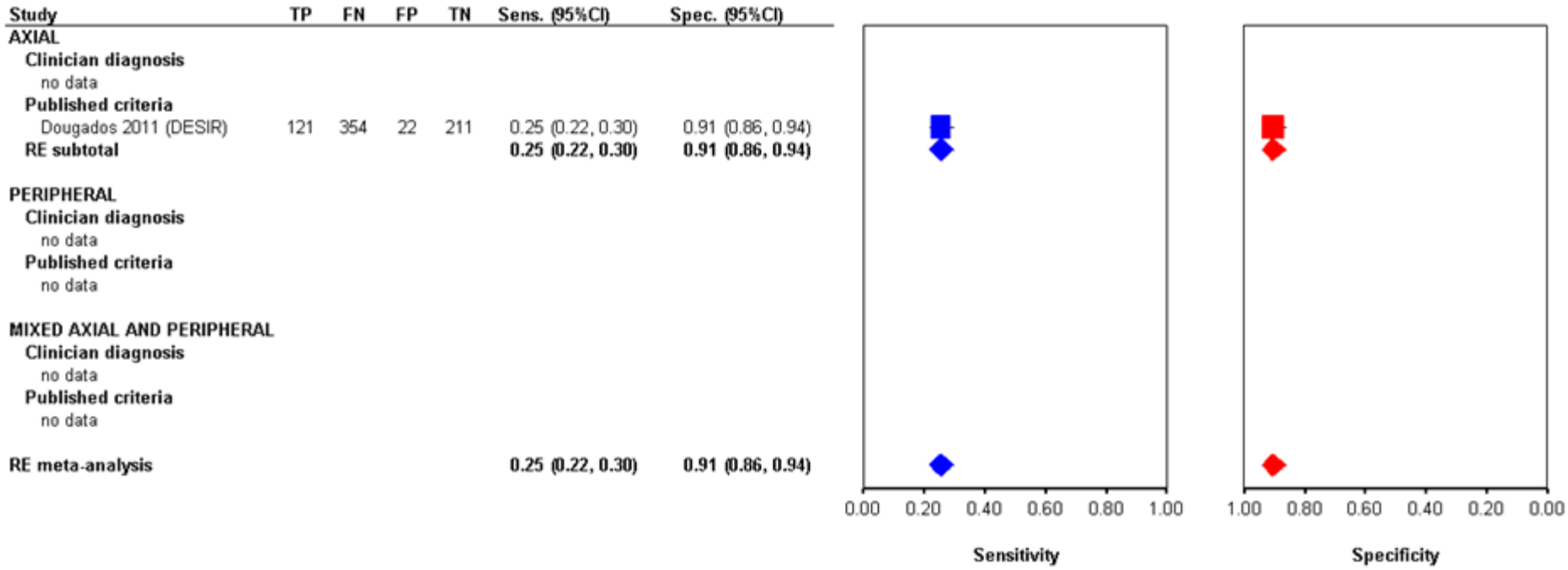


Figure 68 Spinal features on MRI – forest plot: sensitivity and specificity

Enthesitis on MRI

Table 38: Enthesitis on MRI – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	33	4.62 (1.53, 13.93)	LOW
LR-			No serious	n/a	Serious[b]	Serious[d]		0.36 (0.16, 0.84)	LOW
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	Serious[b]	Serious[c]	33	4.62 (1.53, 13.93)	LOW
LR-			No serious	n/a	Serious[b]	Serious[d]		0.36 (0.16, 0.84)	LOW

[a] *Godfrin 2004*
[b] *suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

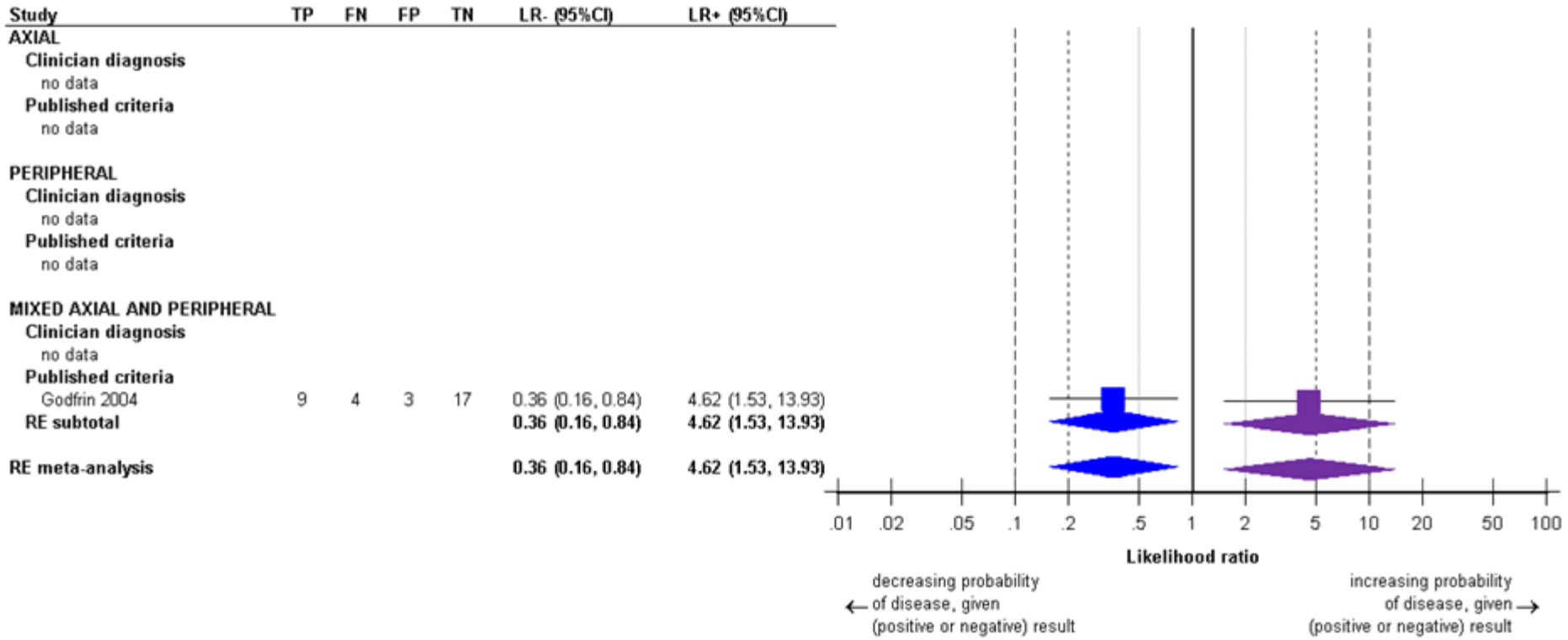


Figure 69 Enthesitis on MRI – forest plot: likelihood ratios

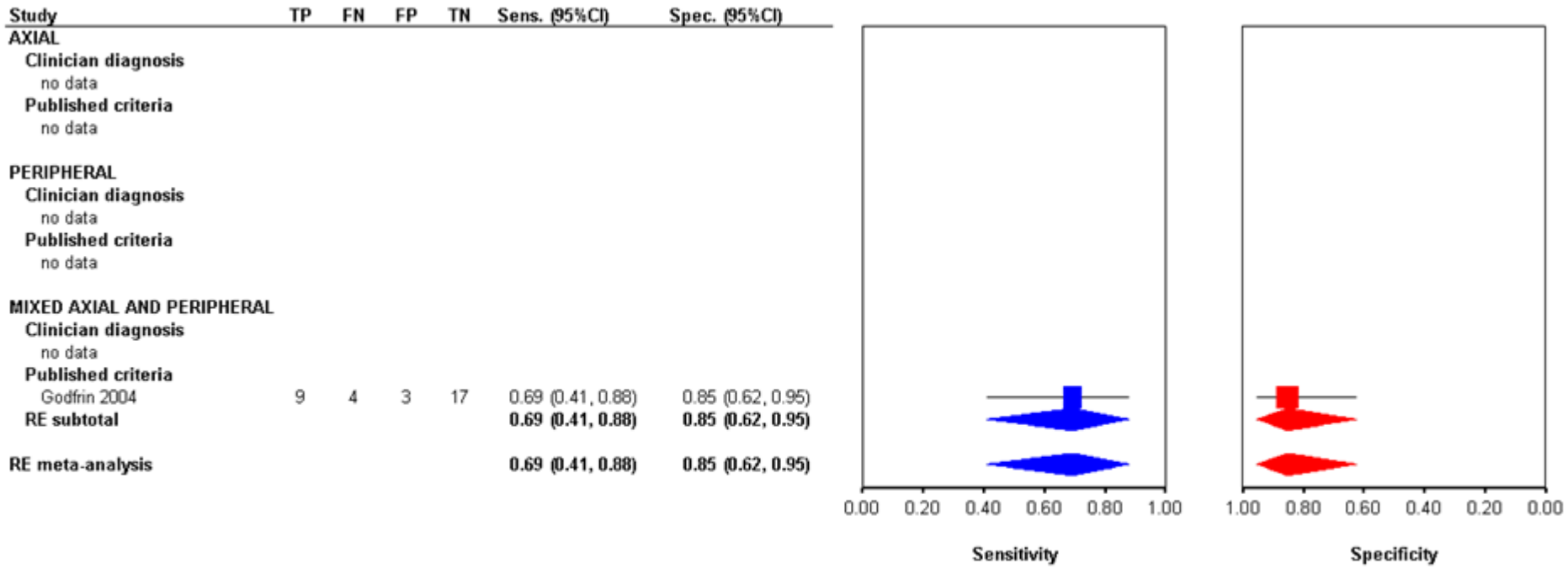


Figure 70 Enthesitis on MRI – forest plot: sensitivity and specificity

G.1.6.3 Ultrasound

Finger or toe pathology on ultrasound

Table 39: Finger or toe pathology on ultrasound – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	52	33.54 (2.19, 514.79)	HIGH
LR-			No serious	n/a	No serious	No serious		0.01 (0.00, 0.22)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	52	33.54 (2.19, 514.79)	HIGH
LR-			No serious	n/a	No serious	No serious		0.01 (0.00, 0.22)	HIGH

[a] *De Simone 2011*

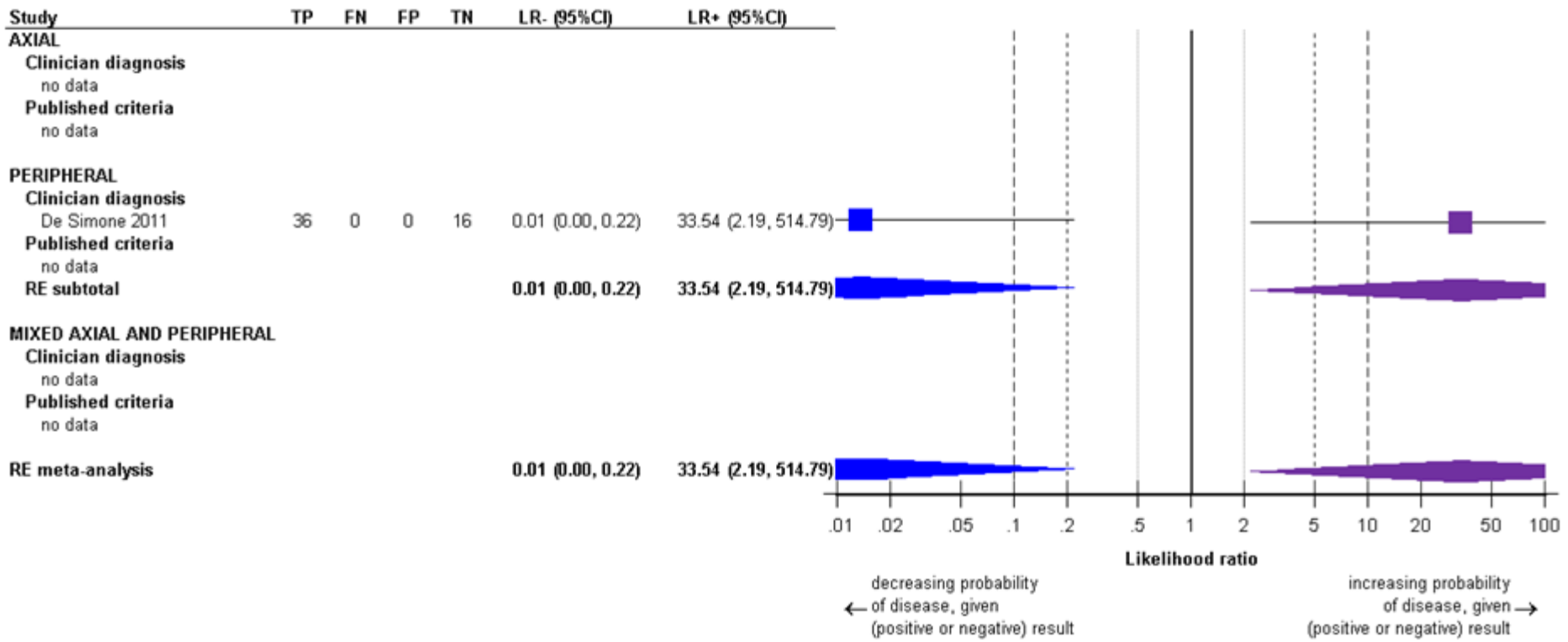


Figure 71 Finger or toe pathology on ultrasound – forest plot: likelihood ratios

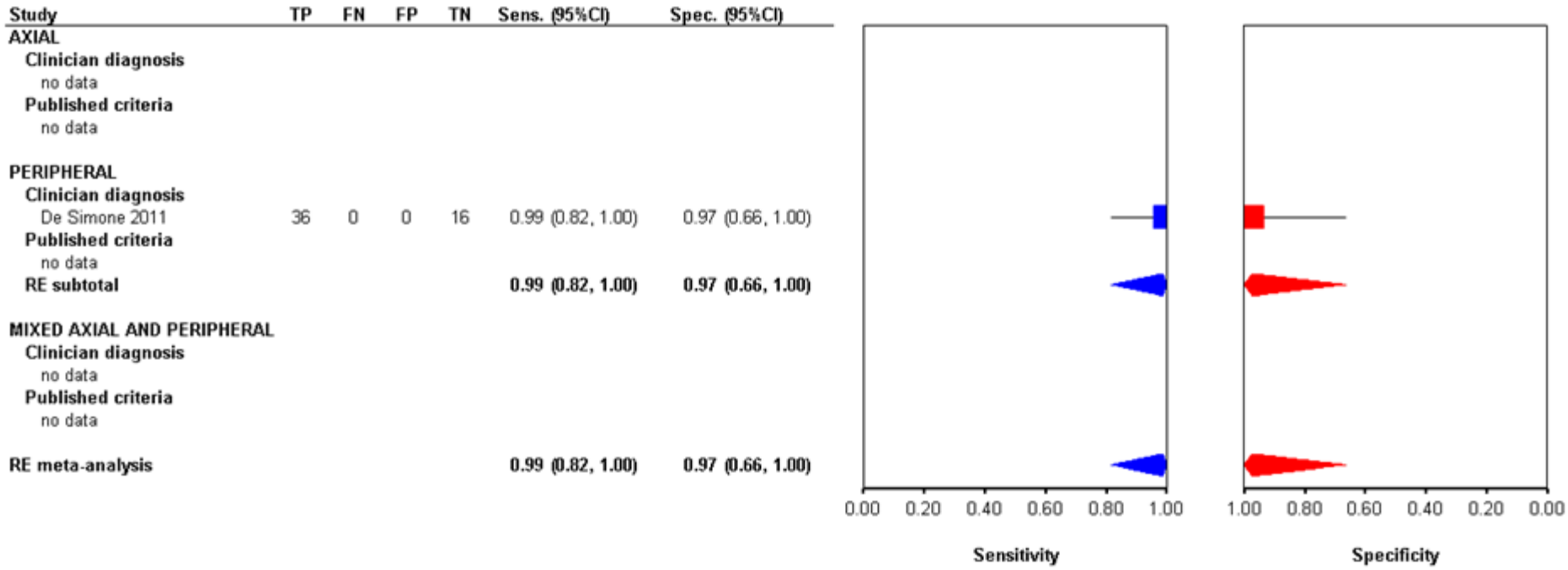


Figure 72 Finger or toe pathology on ultrasound – forest plot: sensitivity and specificity

Finger or toe pathology on power Doppler ultrasound

Table 40 Finger or toe pathology on power Doppler ultrasound – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	Serious[b]	52	2.15 (1.12, 4.13)	MODERATE
LR-			No serious	n/a	No serious	Serious[c]		0.31 (0.14, 0.67)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	Serious[b]	52	2.15 (1.12, 4.13)	MODERATE
LR-			No serious	n/a	No serious	Serious[c]		0.31 (0.14, 0.67)	MODERATE

[a] *De Simone 2011*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

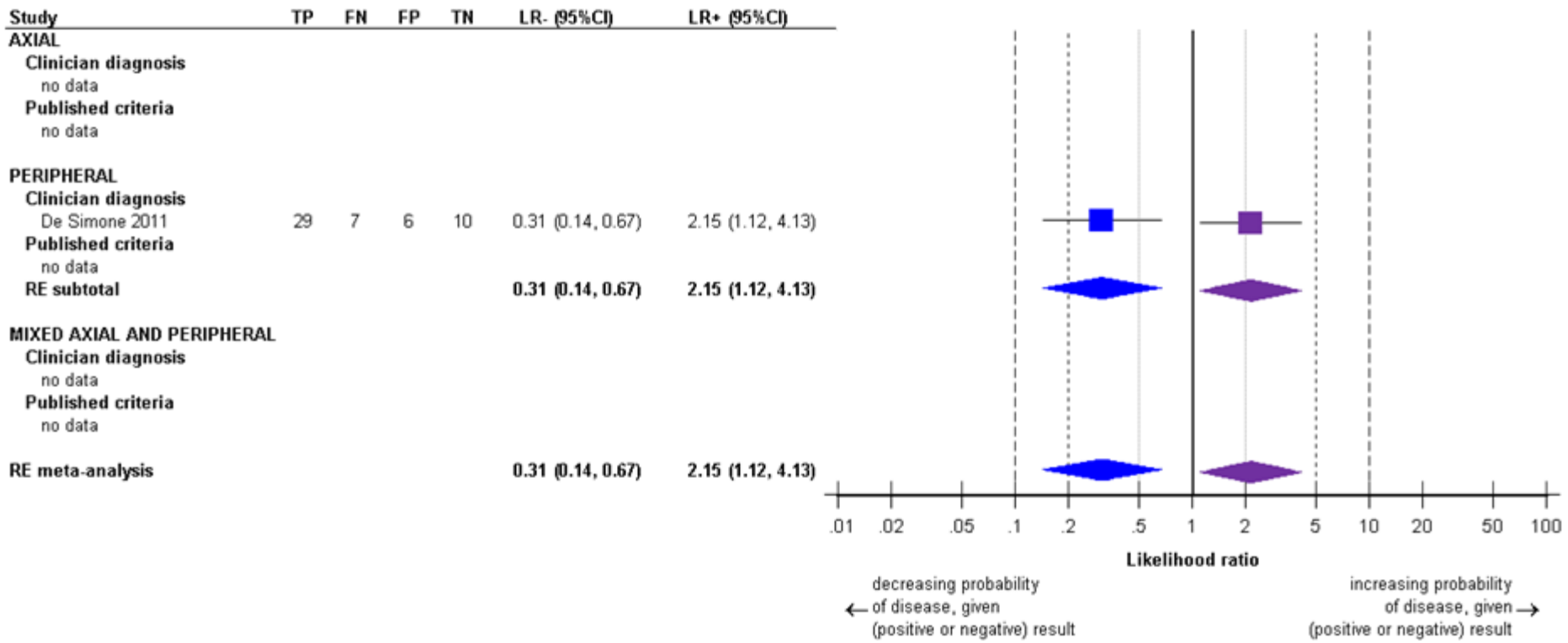


Figure 73 Finger or toe pathology on power Doppler ultrasound – forest plot: likelihood ratios

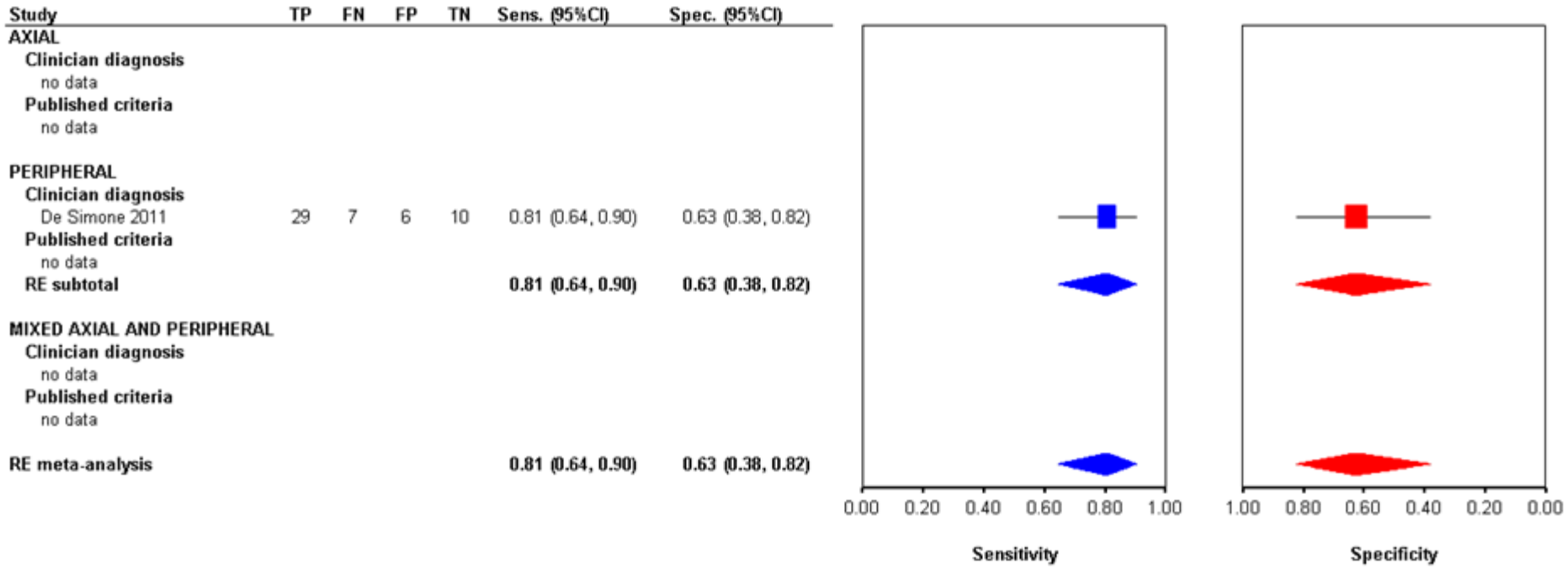


Figure 74 Finger or toe pathology on power Doppler ultrasound – forest plot: sensitivity and specificity

Enthesitis on power Doppler ultrasound

Table 41 Enthesitis on power Doppler ultrasound – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	99	1.43 (1.11, 1.84)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.35 (0.16, 0.75)	MODERATE
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	99	1.43 (1.11, 1.84)	HIGH
LR-			No serious	n/a	No serious	Serious[b]		0.35 (0.16, 0.75)	MODERATE

[a] *D'Agostino 2011*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	44	7	29	19	0.35 (0.16, 0.75)	1.43 (1.11, 1.84)
Published criteria						
no data						
RE subtotal					**0.35 (0.16, 0.75)**	**1.43 (1.11, 1.84)**
RE meta-analysis					**0.35 (0.16, 0.75)**	**1.43 (1.11, 1.84)**

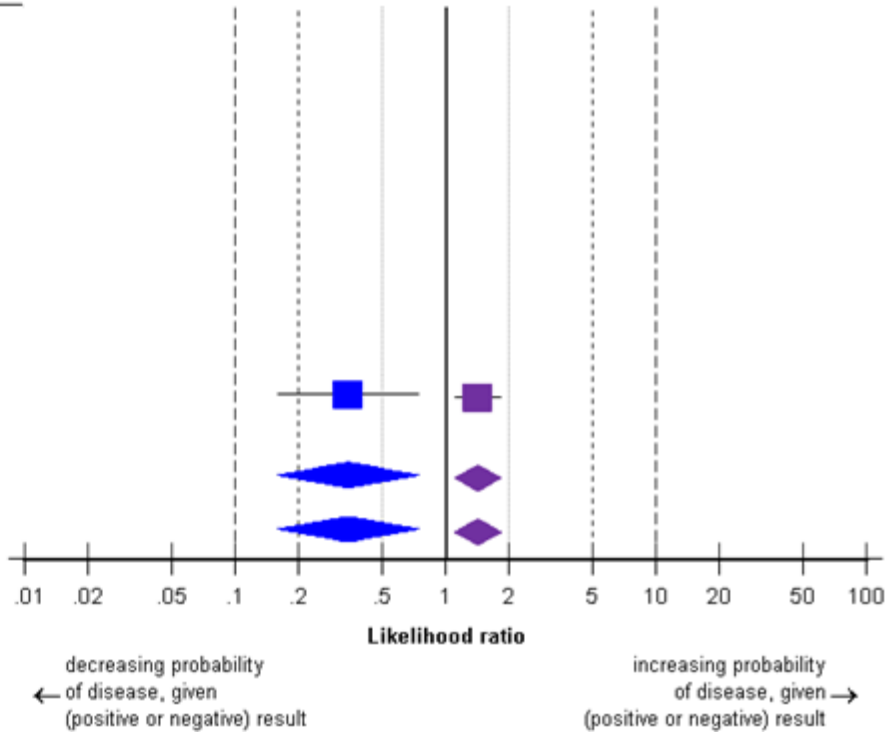


Figure 75 Enthesitis on power Doppler ultrasound – forest plot: likelihood ratios

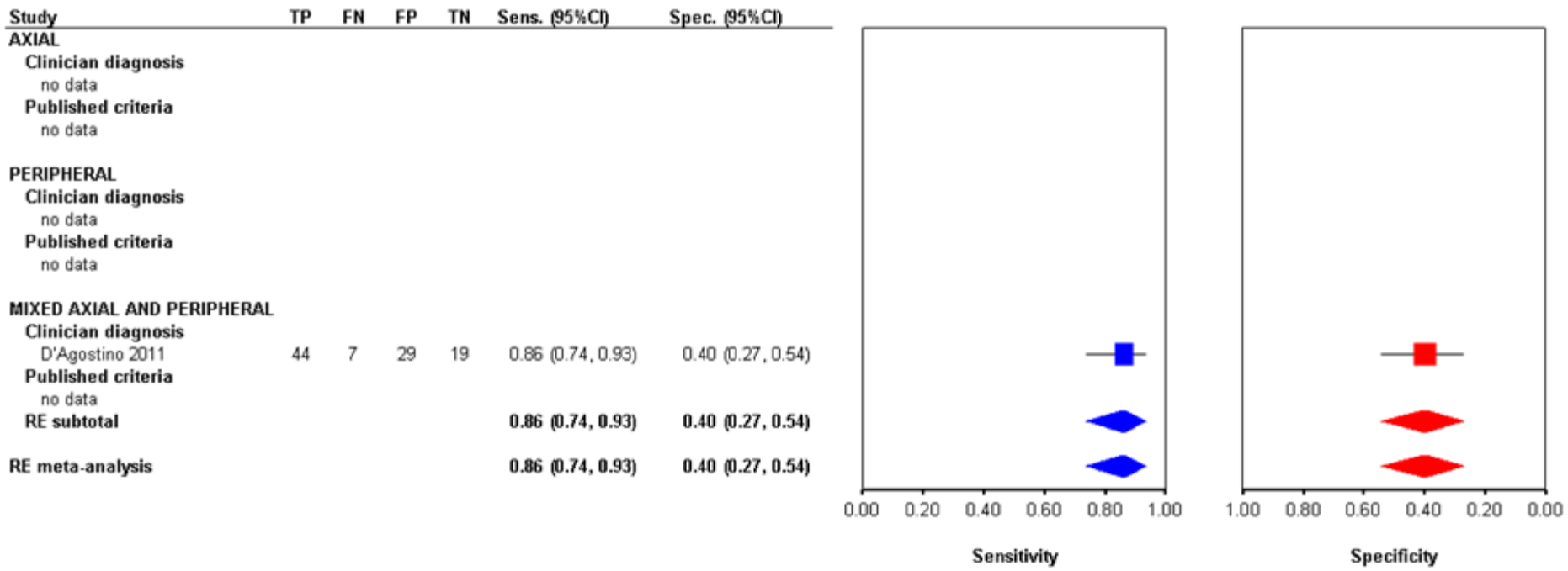


Figure 76: Enthesitis on power Doppler ultrasound – forest plot: sensitivity and specificity

G.1.6.4 Scintigraphy

Sacroiliitis on scintigraphy

Table 42: Sacroiliitis on scintigraphy – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	194	1.31 (1.02, 1.68)	MODERATE
LR-			Serious	n/a	No serious	Serious[b]		0.69 (0.50, 0.97)	LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	194	1.31 (1.02, 1.68)	MODERATE
LR-			Serious	n/a	No serious	Serious[b]		0.69 (0.50, 0.97)	LOW

[a] *Song 2010*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Song 2010	63	34	48	49	0.69 (0.50, 0.97)	1.31 (1.02, 1.68)
Published criteria						
no data						
RE subtotal					**0.69 (0.50, 0.97)**	**1.31 (1.02, 1.68)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.69 (0.50, 0.97)**	**1.31 (1.02, 1.68)**

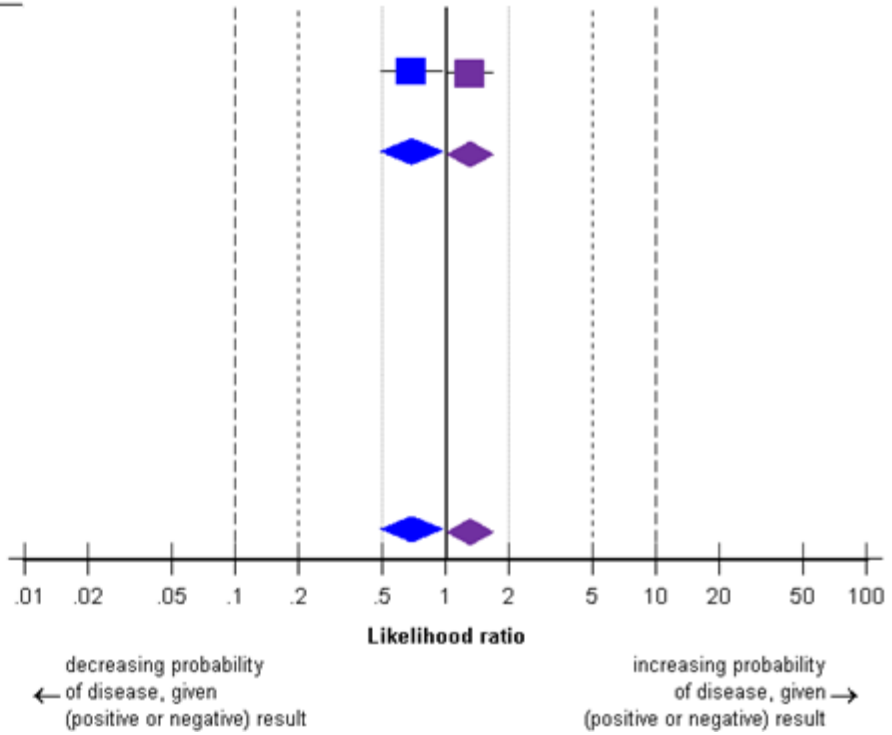


Figure 77: Sacroiliitis on scintigraphy – forest plot: likelihood ratios

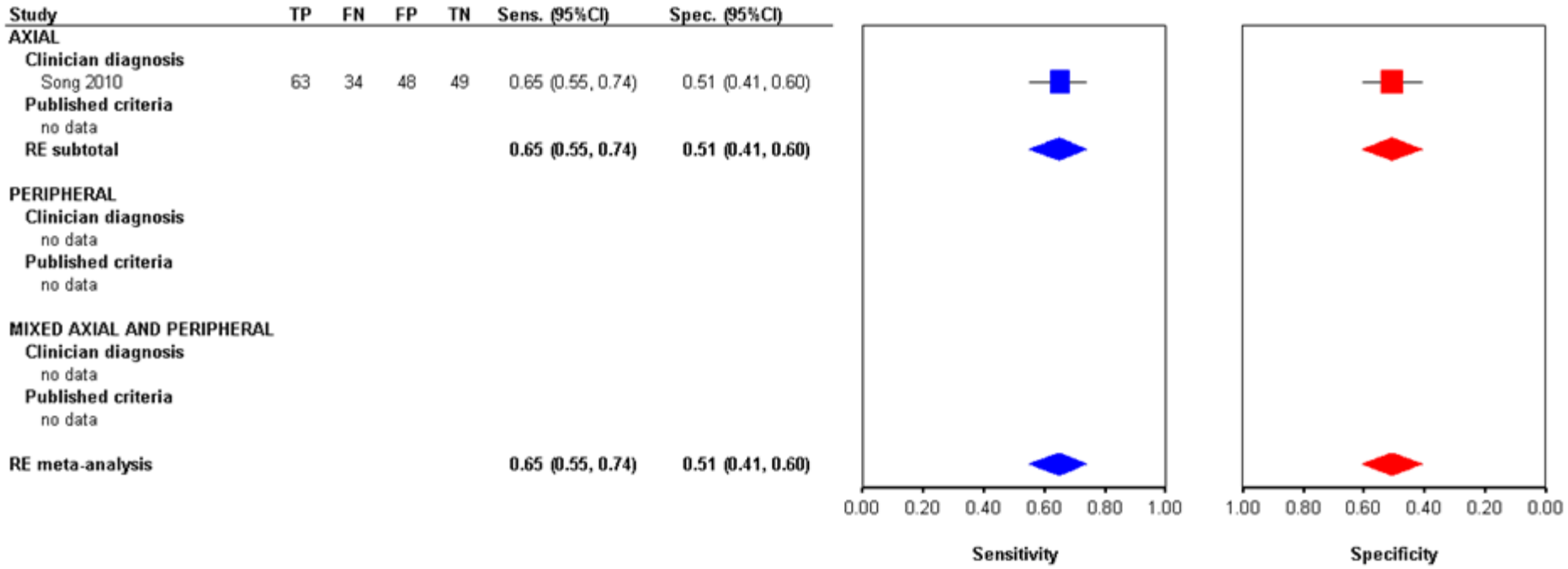


Figure 78: Sacroiliitis on scintigraphy – forest plot: sensitivity and specificity

G.1.7 Information gathering to improve early diagnosis

Review Question 5

- What is the usefulness of information gathering (for example family history, self-report questionnaires, and screening criteria) in improving early diagnosis of spondyloarthritis?

None

G.1.8 Diagnostic risk scores and models

Review Question 4

- What is the diagnostic utility of a risk assessment score for identifying spondyloarthritis?

G.1.8.1 Amor criteria

Original Amor criteria

Table 43: Original Amor criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,357	1.97 (0.80, 4.84)	VERY LOW
LR-			Serious[k]	No serious	No serious	No serious		0.39 (0.34, 0.46)	MODERATE
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	No serious	n/a	No serious	No serious	266	15.85 (3.97, 63.33)	HIGH
LR-			No serious	n/a	No serious	No serious		0.66 (0.59, 0.74)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[f]	Cross-sectional	No serious	Serious[l]	Serious[g]	Serious[m]	907	3.03 (1.36, 6.78)	VERY LOW
LR-			No serious	No serious	No serious	Serious[h]		0.47 (0.42, 0.53)	MODERATE
ALL EVIDENCE POOLED									
LR+	6 studies[i]	Cross-sectional	No serious	Serious[l]	No serious	Serious[m]	2,530	2.98 (1.68, 5.31)	LOW
LR-			No serious	Serious[l]	No serious	Serious[q]		0.47 (0.37, 0.59)	LOW
AXIAL									
LR+	2 studies[j]	Cross-sectional	Serious[k]	Serious[l]	No serious	Serious[m]	1,357	1.97 (0.80, 4.84)	VERY LOW
LR-			Serious[k]	No serious	No serious	No serious		0.39 (0.34, 0.46)	MODERATE
PERIPHERAL									
LR+	1 study[n]	Cross-sectional	No serious	n/a	No serious	No serious	266	15.85 (3.97, 63.33)	HIGH
LR-			No serious	n/a	No serious	No serious		0.66 (0.59, 0.74)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[o]	Cross-sectional	No serious	Serious[l]	Serious[p]	Serious[m]	907	3.03 (1.36, 6.78)	VERY LOW
LR-			No serious	No serious	No serious	Serious[q]		0.47 (0.42, 0.53)	MODERATE

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
ALL EVIDENCE POOLED									
LR+	6 studies[r]	Cross-sectional	No serious	Serious[l]	No serious	Serious[m]	2,530	2.98 (1.68, 5.31)	LOW
LR-			No serious	Serious[l]	No serious	Serious[q]		0.47 (0.37, 0.59)	LOW

[j] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS)*
[k] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[l] *I2 ≥ 50%*
[m] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[n] *Rudwaleit 2011*
[o] *D'Agostino 2011; Godfrin 2004 ; Tomero 2014*
[p] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[q] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[r] *Dougados 2015 (DESIR); D'Agostino 2011; Godfrin 2004 ; Rudwaleit 2009 (ASAS); Rudwaleit 2011; Tomero 2014*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Dougados 2015 (DESIR)	282	37	275	114	0.40 (0.28, 0.56)	1.25 (1.16, 1.35)
Rudwaleit 2009 (ASAS)	271	120	57	201	0.39 (0.33, 0.46)	3.14 (2.47, 3.98)
RE subtotal					**0.39 (0.34, 0.46)**	**1.97 (0.80, 4.84)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.00, df=1 (p=0.981); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.41; Chi²=51.98, df=1 (p<0.001); I²=98.1%						
Published criteria						
no data						
RE subtotal					**0.39 (0.34, 0.46)**	**1.97 (0.80, 4.84)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.00, df=1 (p=0.981); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.41; Chi²=51.98, df=1 (p<0.001); I²=98.1%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	62	114	2	88	0.66 (0.59, 0.74)	15.85 (3.97, 63.33)
Published criteria						
no data						
RE subtotal					**0.66 (0.59, 0.74)**	**15.85 (3.97, 63.33)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	31	20	6	42	0.45 (0.31, 0.64)	4.86 (2.23, 10.61)
Tomero 2014	317	221	33	204	0.48 (0.43, 0.53)	4.23 (3.06, 5.85)
RE subtotal					**0.47 (0.43, 0.53)**	**4.32 (3.20, 5.83)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.11, df=1 (p=0.743); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.10, df=1 (p=0.747); I²=0.0%						
Published criteria						
Godfrin 2004	13	0	13	7	0.10 (0.01, 1.61)	1.50 (1.07, 2.10)
RE subtotal					**0.47 (0.42, 0.53)**	**3.03 (1.36, 6.78)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=1.31, df=2 (p=0.520); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.44; Chi²=21.43, df=2 (p<0.001); I²=90.7%						
Between-substratum heterogeneity, LR-: Chi²=1.20, df=1 (p=0.273); I²=16.8%						
Between-substratum heterogeneity, LR+: Chi²=21.33, df=1 (p<0.001); I²=95.3%						
RE meta-analysis					**0.47 (0.37, 0.59)**	**2.98 (1.68, 5.31)**
Overall heterogeneity, LR-: Tau²=0.05; Chi²=35.36, df=5 (p<0.001); I²=85.9%						
Overall heterogeneity, LR+: Tau²=0.44; Chi²=117.69, df=5 (p<0.001); I²=95.8%						
Between-stratum heterogeneity, LR-: Chi²=34.05, df=2 (p<0.001); I²=94.1%						
Between-stratum heterogeneity, LR+: Chi²=44.28, df=2 (p<0.001); I²=95.5%						

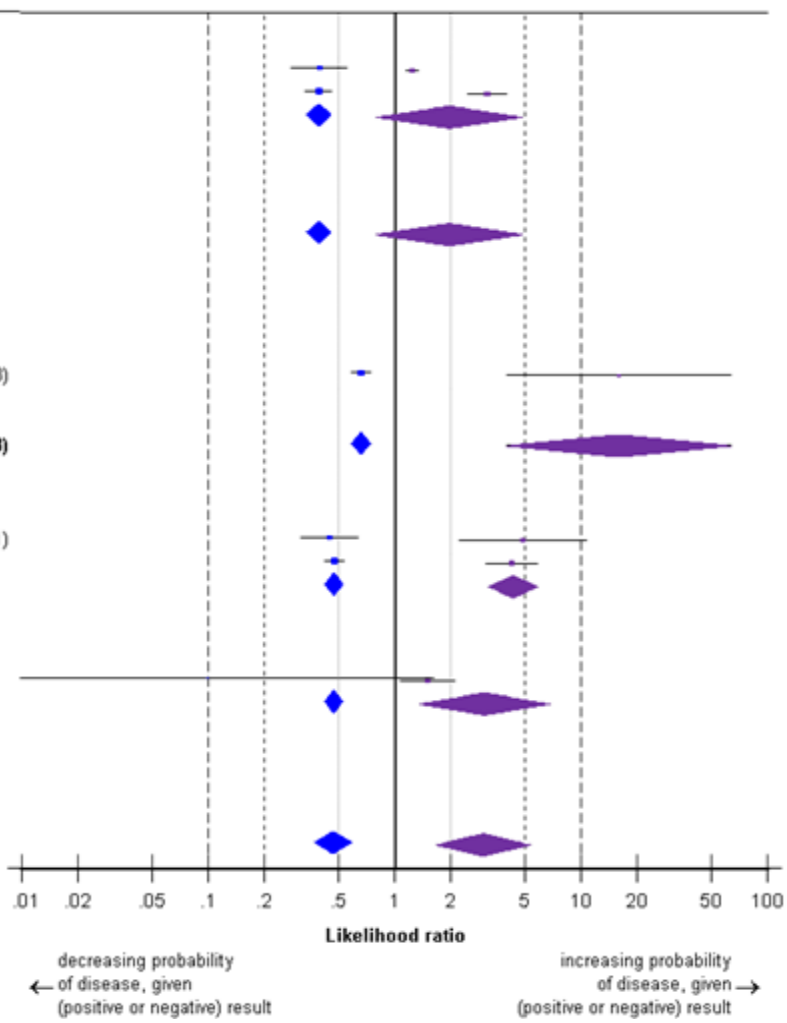


Figure 79: Original Amor criteria – forest plot: likelihood ratios

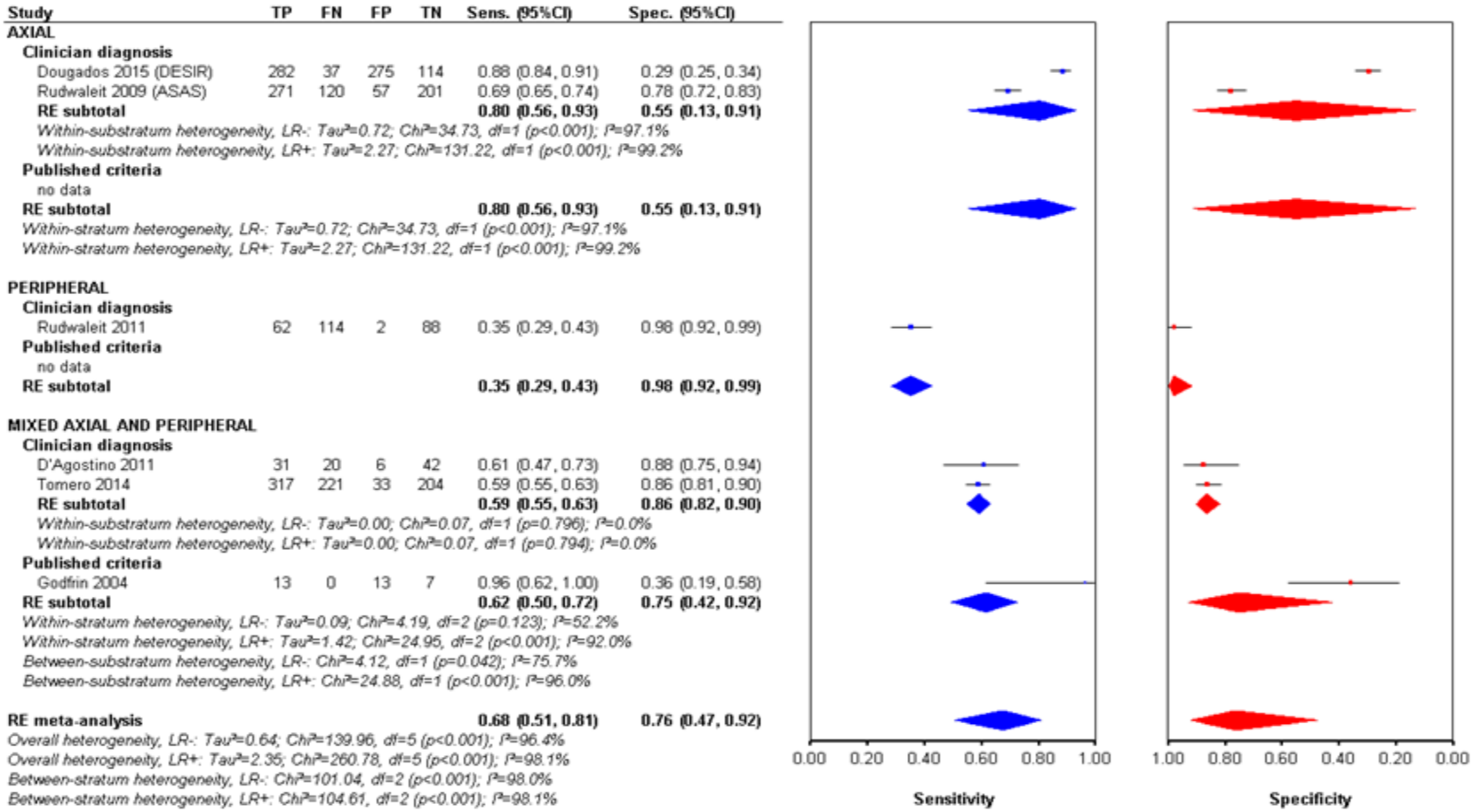


Figure 80: Original Amor criteria – forest plot: sensitivity and specificity

Modified Amor criteria

Table 44: Modified Amor criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,357	2.16 (0.76, 6.09)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	No serious		0.26 (0.18, 0.39)	LOW
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	No serious	n/a	No serious	No serious	266	17.90 (4.49, 71.31)	HIGH
LR-			No serious	n/a	No serious	No serious		0.62 (0.54, 0.70)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	3 studies[f]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,623	3.44 (1.30, 9.12)	VERY LOW
LR-			No serious	Serious[c]	No serious	Serious[g]		0.36 (0.17, 0.74)	LOW

[a] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS)*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *Rudwaleit 2011*
[f] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS); Rudwaleit 2011*
[g] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

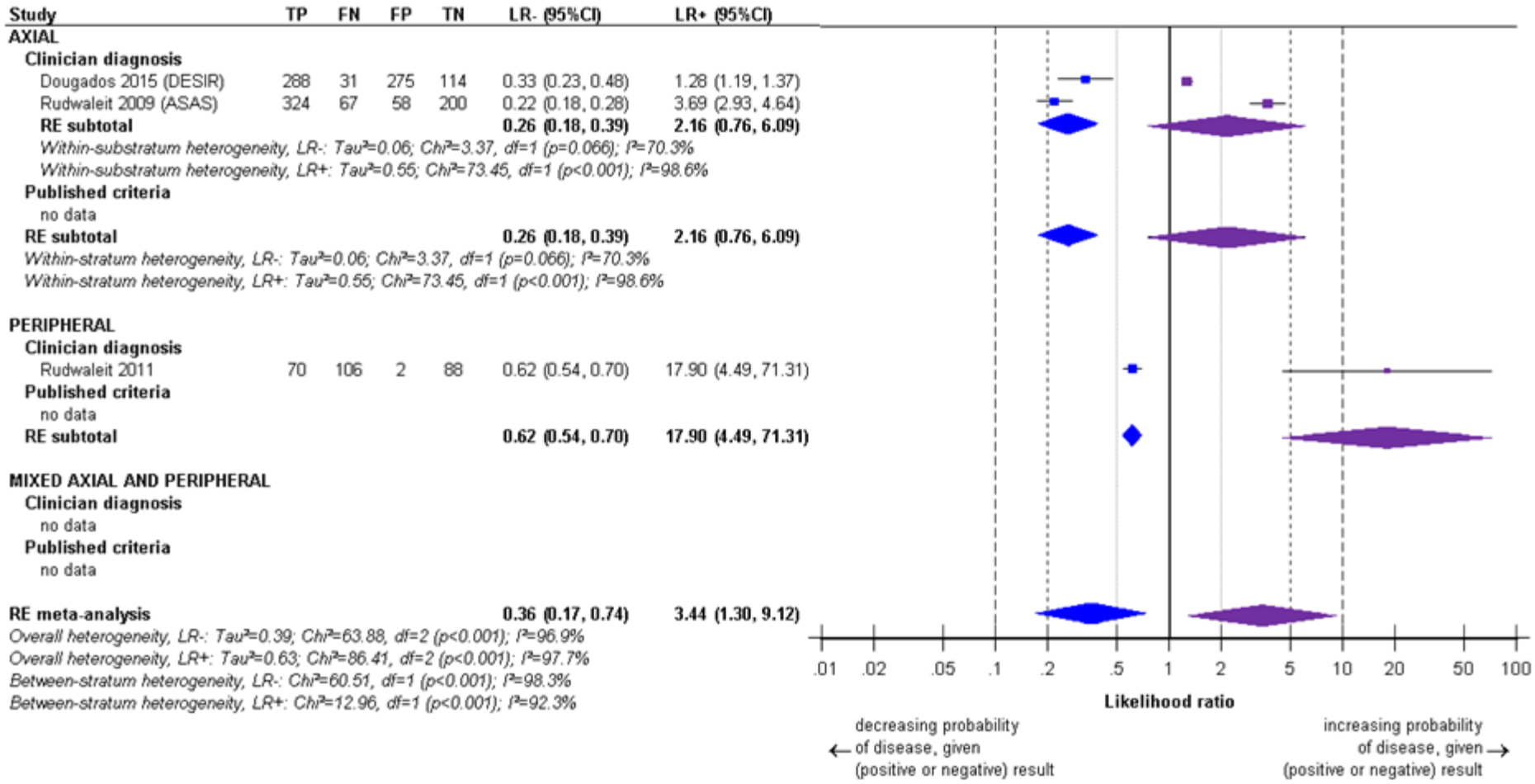


Figure 81: Modified Amor criteria – forest plot: likelihood ratios

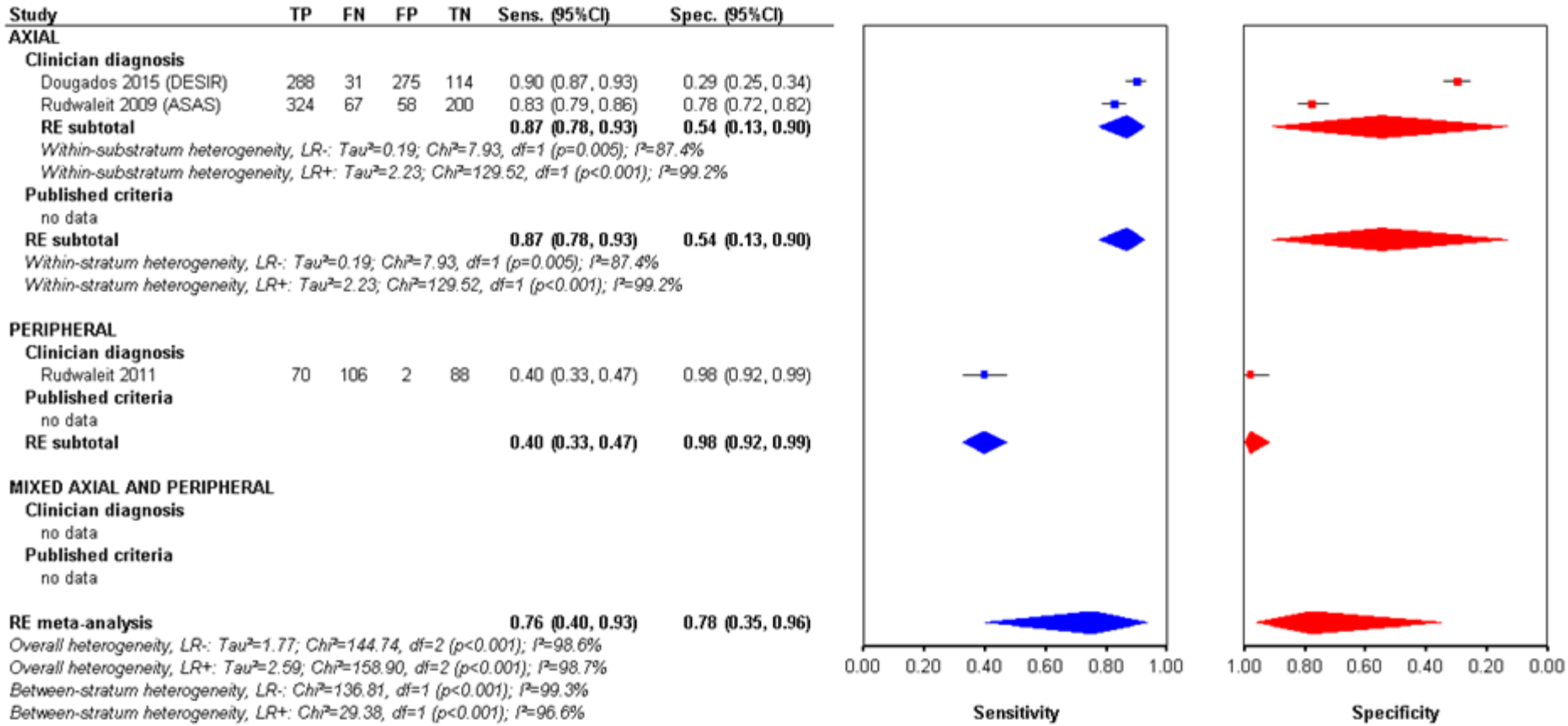


Figure 82: Modified Amor criteria – forest plot: sensitivity and specificity

G.1.8.2 ASAS axial criteria

Table 45: ASAS axial criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,357	2.71 (0.72, 10.12)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[e]		0.30 (0.14, 0.66)	VERY LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[f]	Cross-sectional	No serious	n/a	No serious	Serious[d]	43	3.26 (1.29, 8.23)	MODERATE
LR-			No serious	n/a	No serious	Serious[e]		0.43 (0.24, 0.79)	MODERATE
ALL EVIDENCE POOLED									
LR+	3 studies[g]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,400	2.85 (0.98, 8.35)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[e]		0.33 (0.18, 0.62)	VERY LOW

[a] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS)*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[f] *D'Agostino 2011*
[g] *Dougados 2015 (DESIR); D'Agostino 2011; Rudwaleit 2009 (ASAS)*

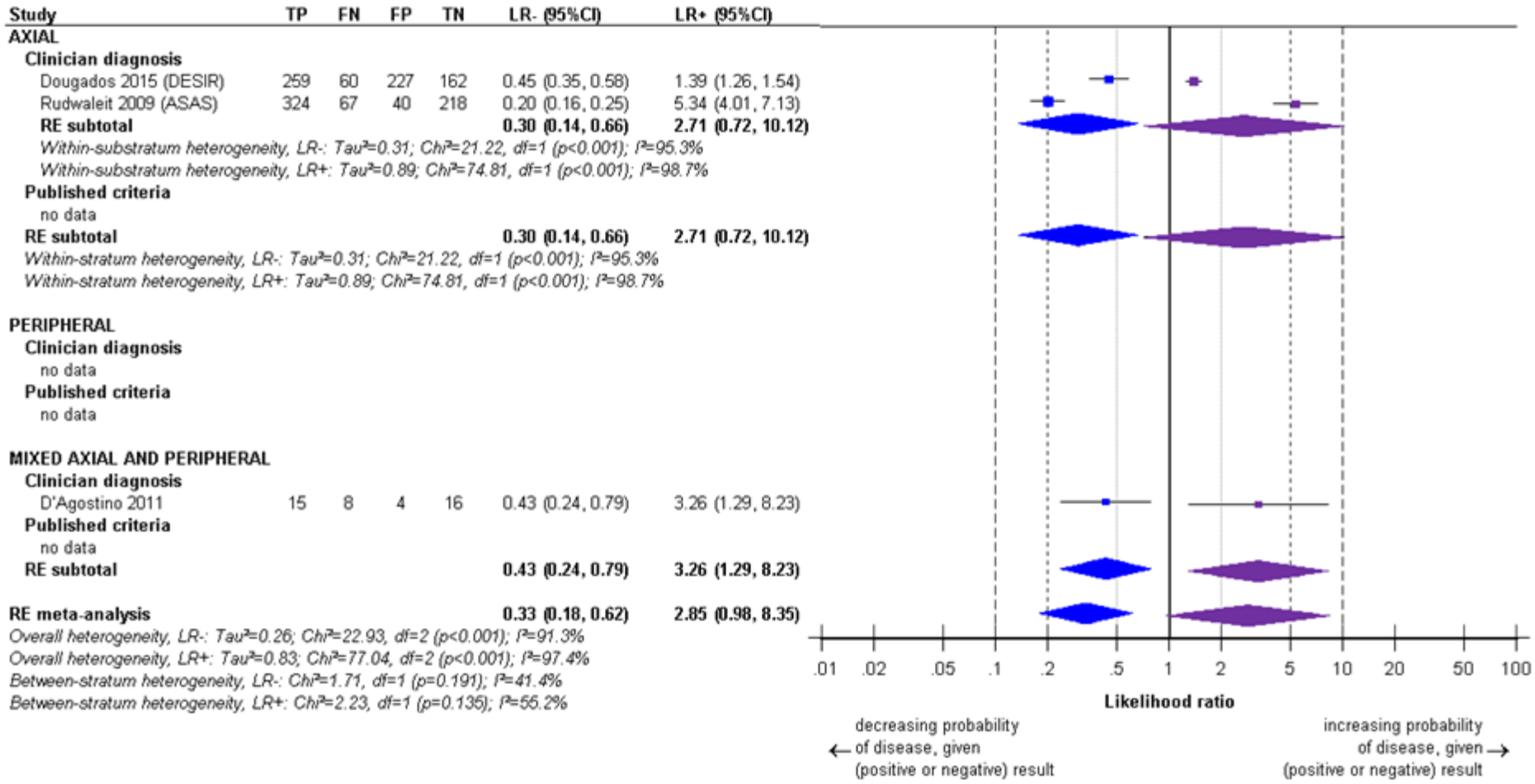


Figure 83: ASAS axial criteria – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Dougados 2015 (DESIR)	259	60	227	162	0.81 (0.77, 0.85)	0.42 (0.37, 0.47)
Rudwaleit 2009 (ASAS)	324	67	40	218	0.83 (0.79, 0.86)	0.84 (0.80, 0.88)
RE subtotal					**0.82 (0.79, 0.85)**	**0.66 (0.21, 0.94)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.33, df=1 (p=0.563); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=2.05; Chi²=102.90, df=1 (p<0.001); I²=99.0%						
Published criteria						
no data						
RE subtotal					**0.82 (0.79, 0.85)**	**0.66 (0.21, 0.94)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.33, df=1 (p=0.563); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=2.05; Chi²=102.90, df=1 (p<0.001); I²=99.0%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	15	8	4	16	0.65 (0.44, 0.82)	0.80 (0.57, 0.92)
Published criteria						
no data						
RE subtotal					**0.65 (0.44, 0.82)**	**0.80 (0.57, 0.92)**
RE meta-analysis					**0.81 (0.75, 0.85)**	**0.71 (0.33, 0.92)**
Overall heterogeneity, LR-: Tau²=0.04; Chi²=4.31, df=2 (p=0.116); I²=53.6%						
Overall heterogeneity, LR+: Tau²=1.92; Chi²=107.31, df=2 (p<0.001); I²=98.1%						
Between-stratum heterogeneity, LR-: Chi²=3.97, df=1 (p=0.046); I²=74.8%						
Between-stratum heterogeneity, LR+: Chi²=4.41, df=1 (p=0.036); I²=77.3%						

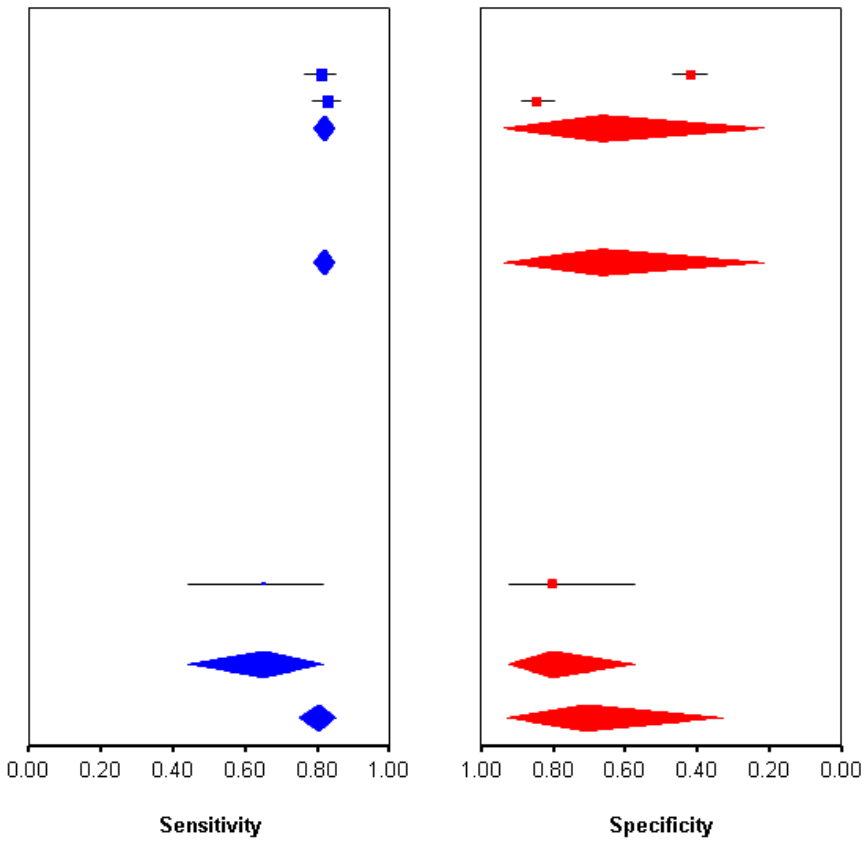


Figure 84: ASAS axial criteria – forest plot: sensitivity and specificity

ASAS axial criteria (imaging 'arm' only)

Table 46: ASAS axial criteria (imaging 'arm' only) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	649	24.41 (11.72, 50.87)	HIGH
LR-			No serious	n/a	No serious	No serious		0.35 (0.30, 0.40)	HIGH
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	649	24.41 (11.72, 50.87)	HIGH
LR-			No serious	n/a	No serious	No serious		0.35 (0.30, 0.40)	HIGH

[a]*Rudwaleit 2009 (ASAS)*

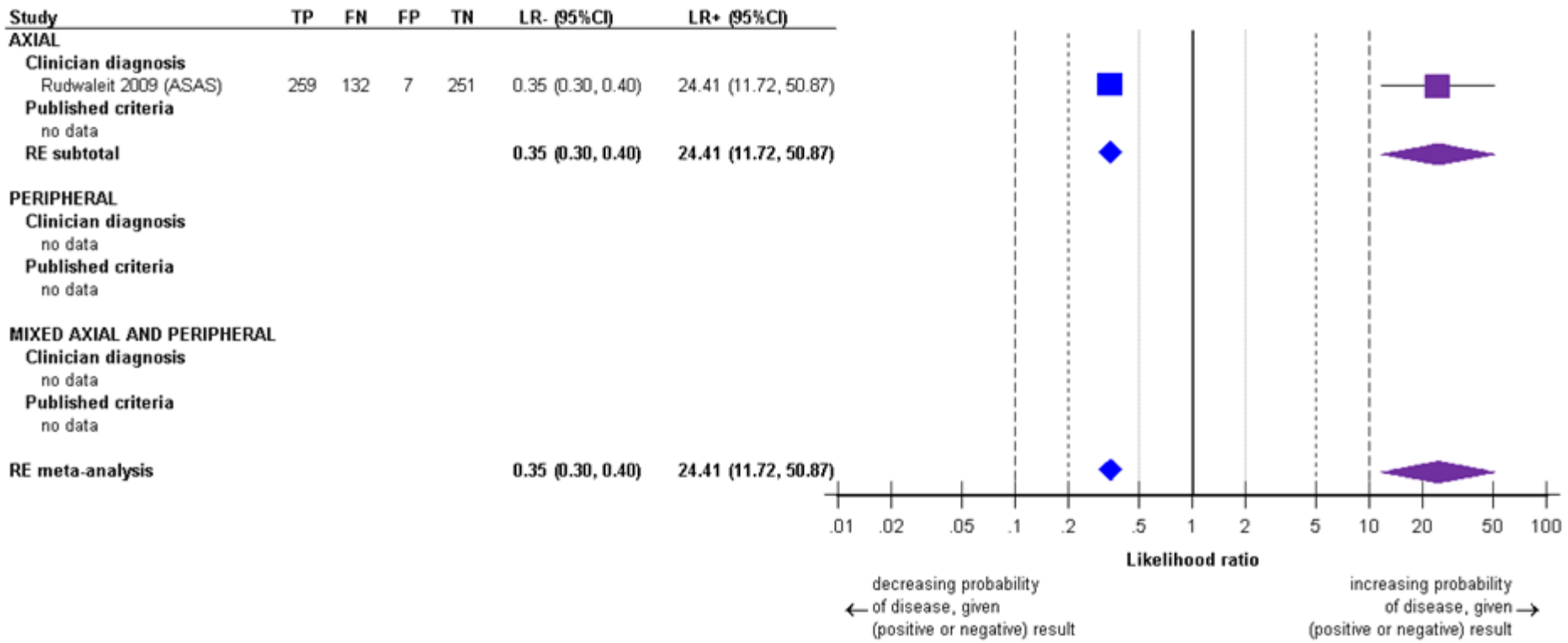


Figure 85: ASAS axial criteria (imaging 'arm' only) – forest plot: likelihood ratios

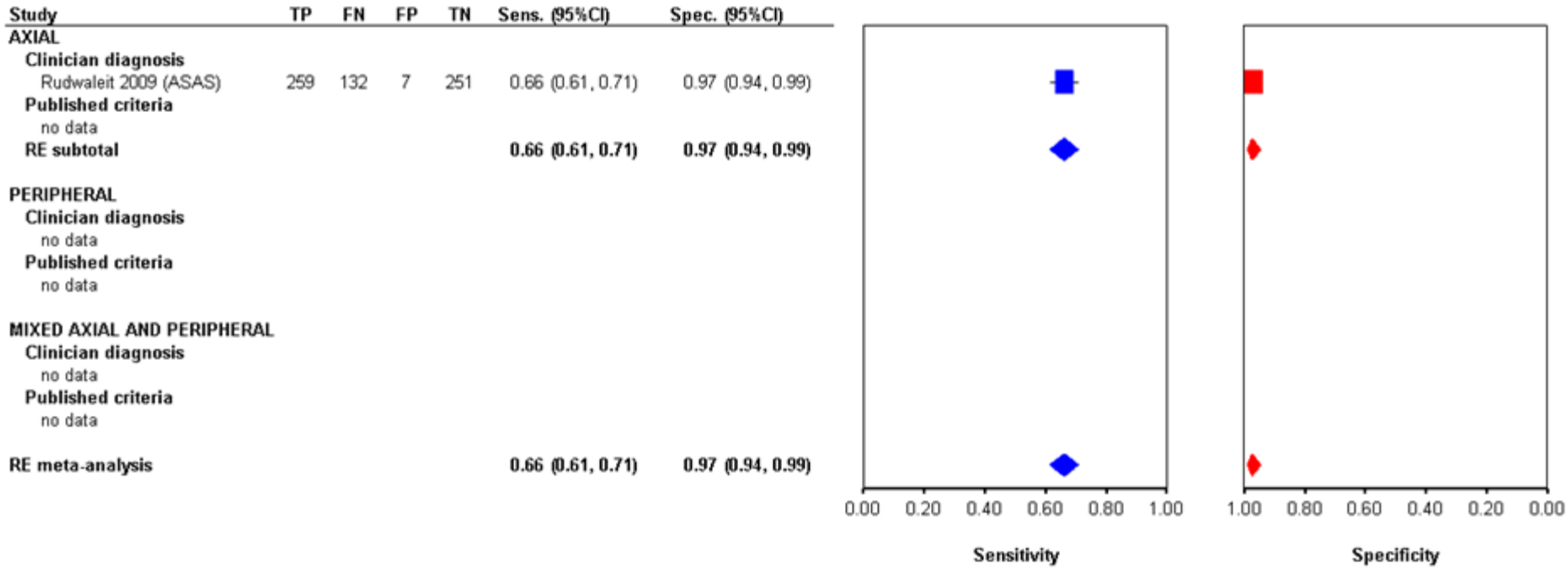


Figure 86: ASAS axial criteria (imaging 'arm' only) – forest plot: sensitivity and specificity

G.1.8.3 Berlin algorithm

Original Berlin algorithm

Table 47: Original Berlin algorithm – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	842	3.30 (2.65, 4.11)	HIGH
LR-			No serious	No serious	No serious	No serious		0.43 (0.38, 0.50)	HIGH
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	1 study[b]	Cross-sectional	No serious	n/a	No serious	Serious[c]	43	3.04 (1.19, 7.76)	MODERATE
LR-			No serious	n/a	No serious	Serious[d]		0.49 (0.28, 0.85)	MODERATE
ALL EVIDENCE POOLED									
LR+	3 studies[e]	Cross-sectional	No serious	No serious	No serious	No serious	885	3.29 (2.65, 4.07)	HIGH
LR-			No serious	No serious	No serious	No serious		0.44 (0.38, 0.50)	HIGH

[a] *van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*
[b] *D'Agostino 2011*
[c] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[e] *D'Agostino 2011; van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

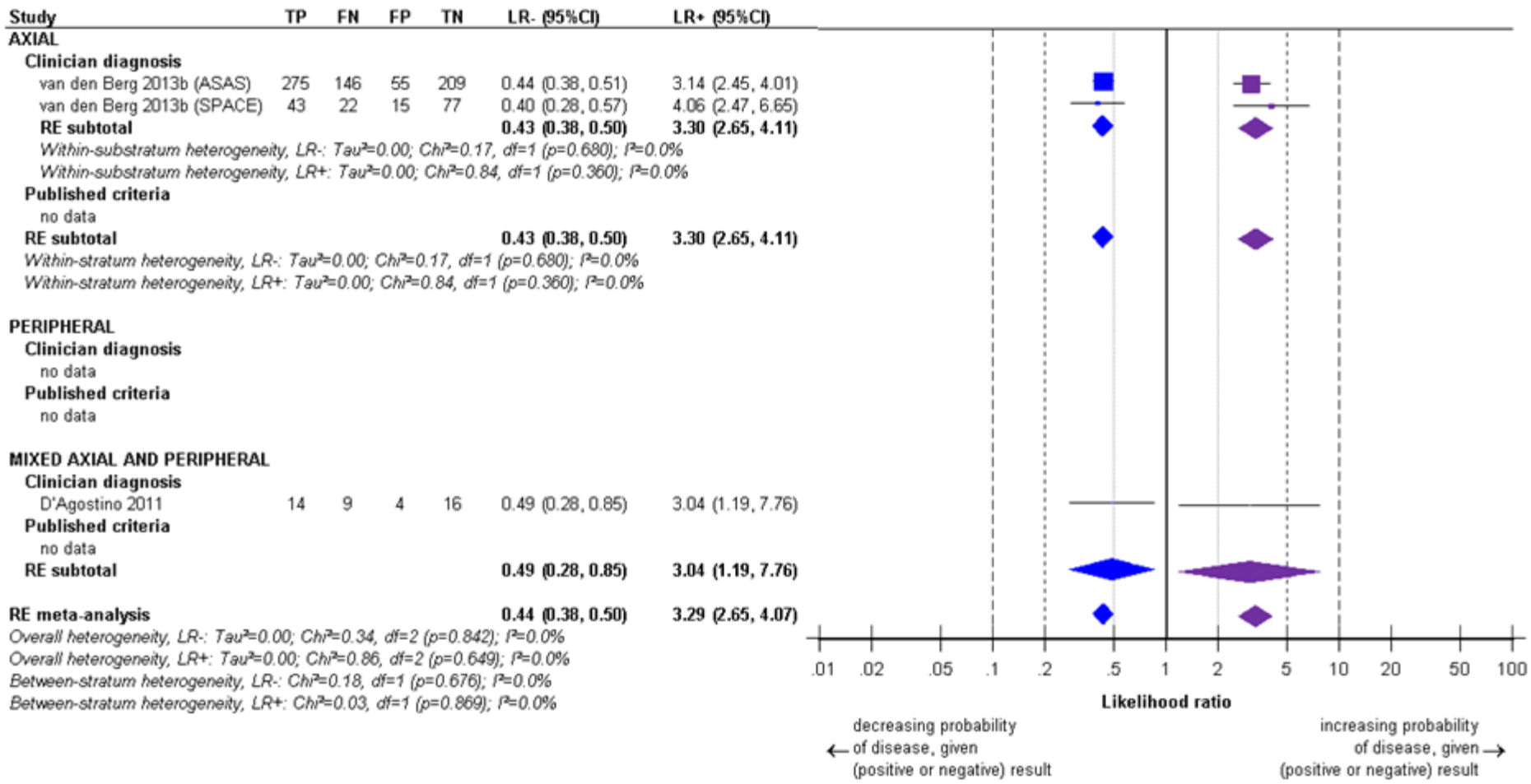


Figure 87: Original Berlin algorithm – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
van den Berg 2013b (ASAS)	275	146	55	209	0.65 (0.61, 0.70)	0.79 (0.74, 0.84)
van den Berg 2013b (SPACE)	43	22	15	77	0.66 (0.54, 0.77)	0.84 (0.75, 0.90)
RE subtotal					**0.65 (0.61, 0.70)**	**0.80 (0.76, 0.84)**
Within-substratum heterogeneity, LR-: Tau²=0.00; Chi²=0.02, df=1 (p=0.895); I²=0.0%						
Within-substratum heterogeneity, LR+: Tau²=0.00; Chi²=0.88, df=1 (p=0.348); I²=0.0%						
Published criteria						
no data						
RE subtotal					**0.65 (0.61, 0.70)**	**0.80 (0.76, 0.84)**
Within-stratum heterogeneity, LR-: Tau²=0.00; Chi²=0.02, df=1 (p=0.895); I²=0.0%						
Within-stratum heterogeneity, LR+: Tau²=0.00; Chi²=0.88, df=1 (p=0.348); I²=0.0%						
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	14	9	4	16	0.61 (0.40, 0.78)	0.80 (0.57, 0.92)
Published criteria						
no data						
RE subtotal					**0.61 (0.40, 0.78)**	**0.80 (0.57, 0.92)**
RE meta-analysis					**0.65 (0.61, 0.69)**	**0.80 (0.76, 0.84)**
Overall heterogeneity, LR-: Tau²=0.00; Chi²=0.22, df=2 (p=0.897); I²=0.0%						
Overall heterogeneity, LR+: Tau²=0.00; Chi²=0.88, df=2 (p=0.643); I²=0.0%						
Between-stratum heterogeneity, LR-: Chi²=0.20, df=1 (p=0.654); I²=0.0%						
Between-stratum heterogeneity, LR+: Chi²=0.00, df=1 (p=0.978); I²=0.0%						

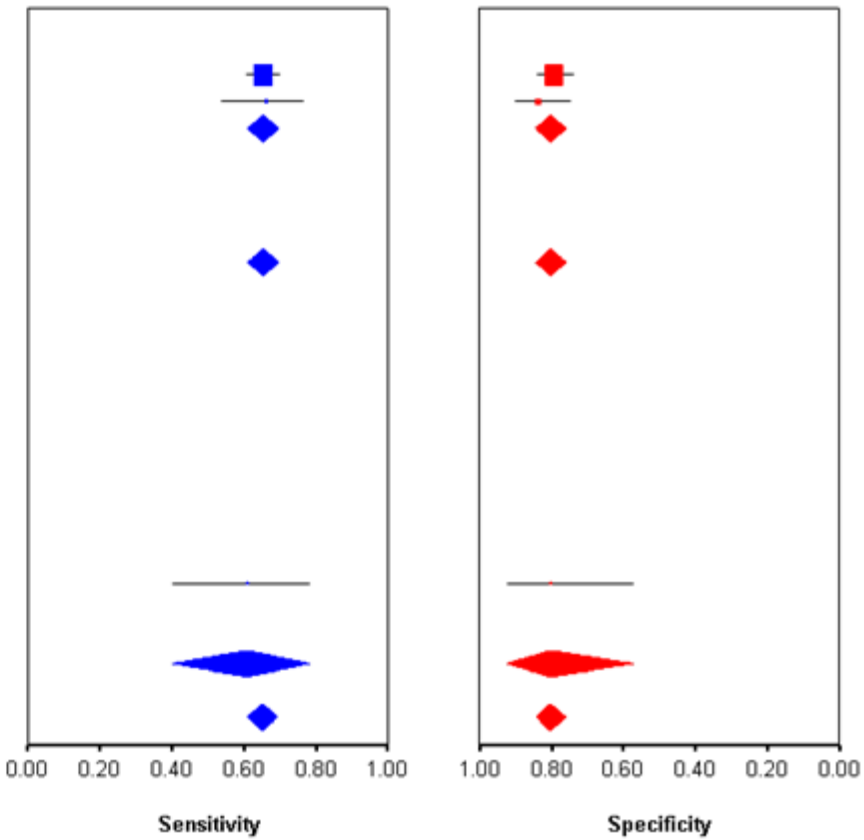


Figure 88: Original Berlin algorithm – forest plot: sensitivity and specificity

Berlin algorithm -- modification #1

Table 48: Berlin algorithm -- modification #1 – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	842	2.91 (2.43, 3.49)	HIGH
LR-			No serious	No serious	No serious	No serious		0.31 (0.26, 0.37)	HIGH
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	842	2.91 (2.43, 3.49)	HIGH
LR-			No serious	No serious	No serious	No serious		0.31 (0.26, 0.37)	HIGH

[a]*van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

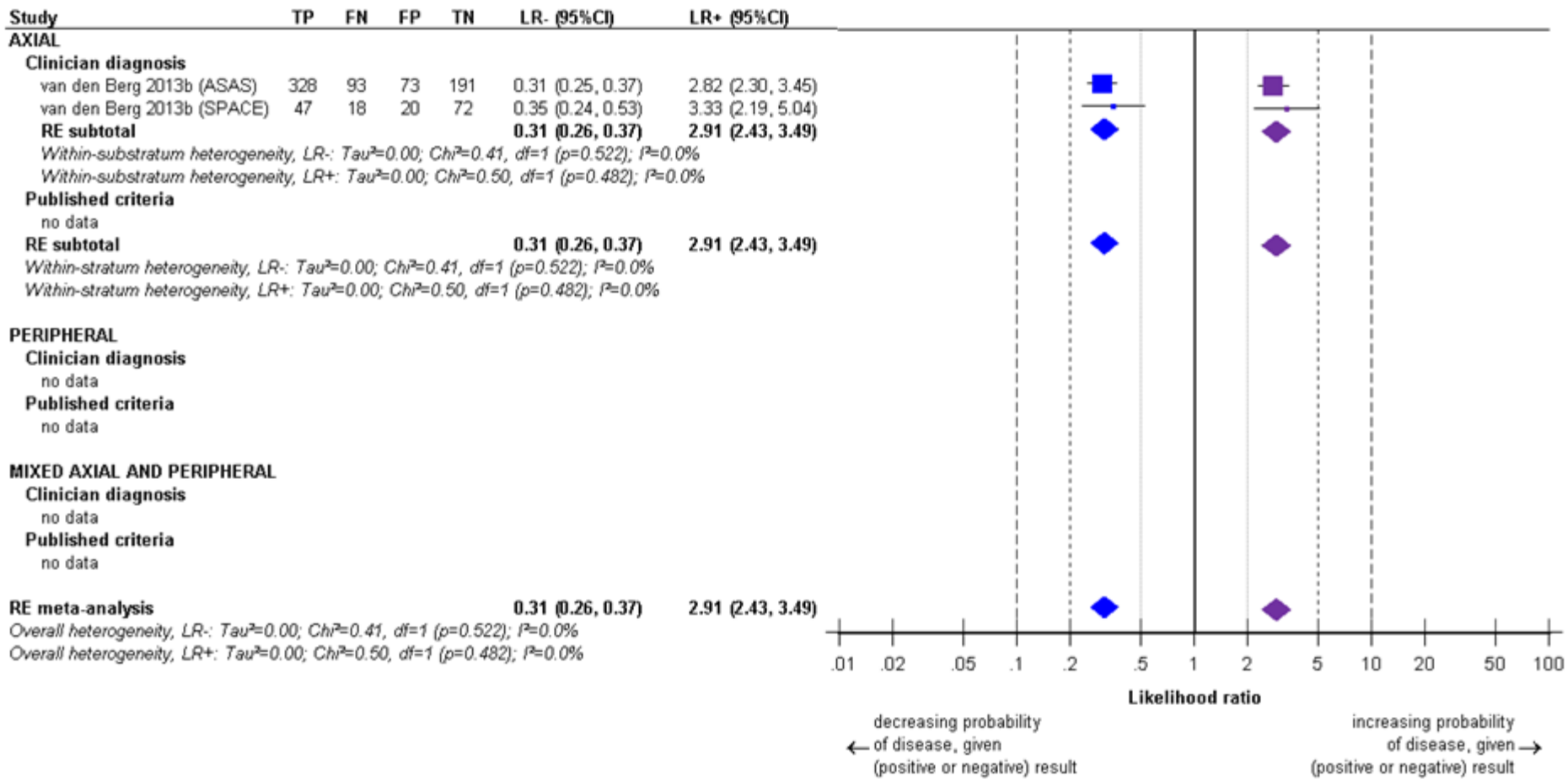


Figure 89: Berlin algorithm -- modification #1 – forest plot: likelihood ratios

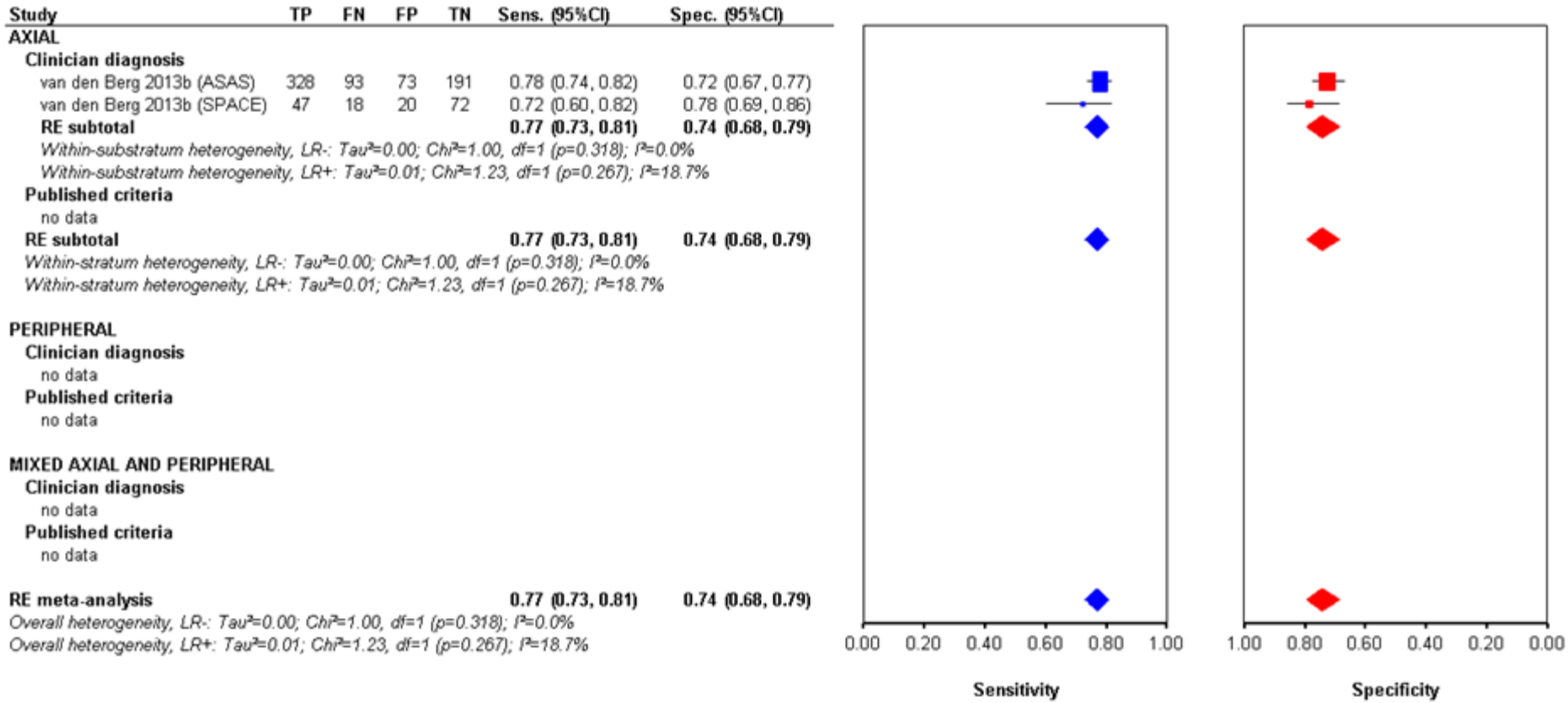


Figure 90: Berlin algorithm -- modification #1 – forest plot: sensitivity and specificity

Berlin algorithm -- modification #2

Table 49: Berlin algorithm -- modification #2 – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	842	3.42 (2.81, 4.16)	HIGH
LR-			No serious	No serious	No serious	No serious		0.27 (0.22, 0.32)	HIGH
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	No serious	No serious	842	3.42 (2.81, 4.16)	HIGH
LR-			No serious	No serious	No serious	No serious		0.27 (0.22, 0.32)	HIGH

[a]*van den Berg 2013b (ASAS); van den Berg 2013b (SPACE)*

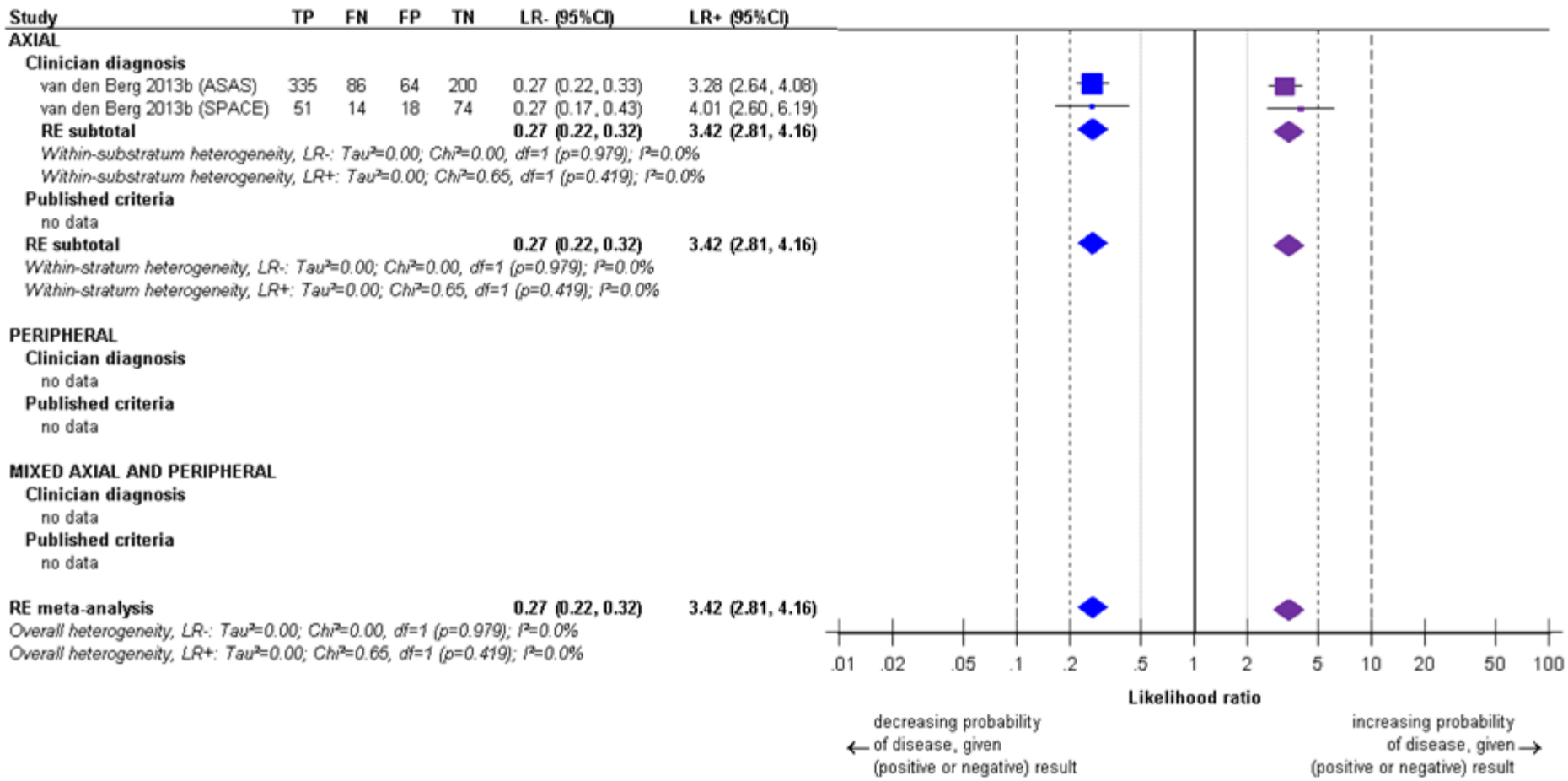


Figure 91: Berlin algorithm -- modification #2 – forest plot: likelihood ratios

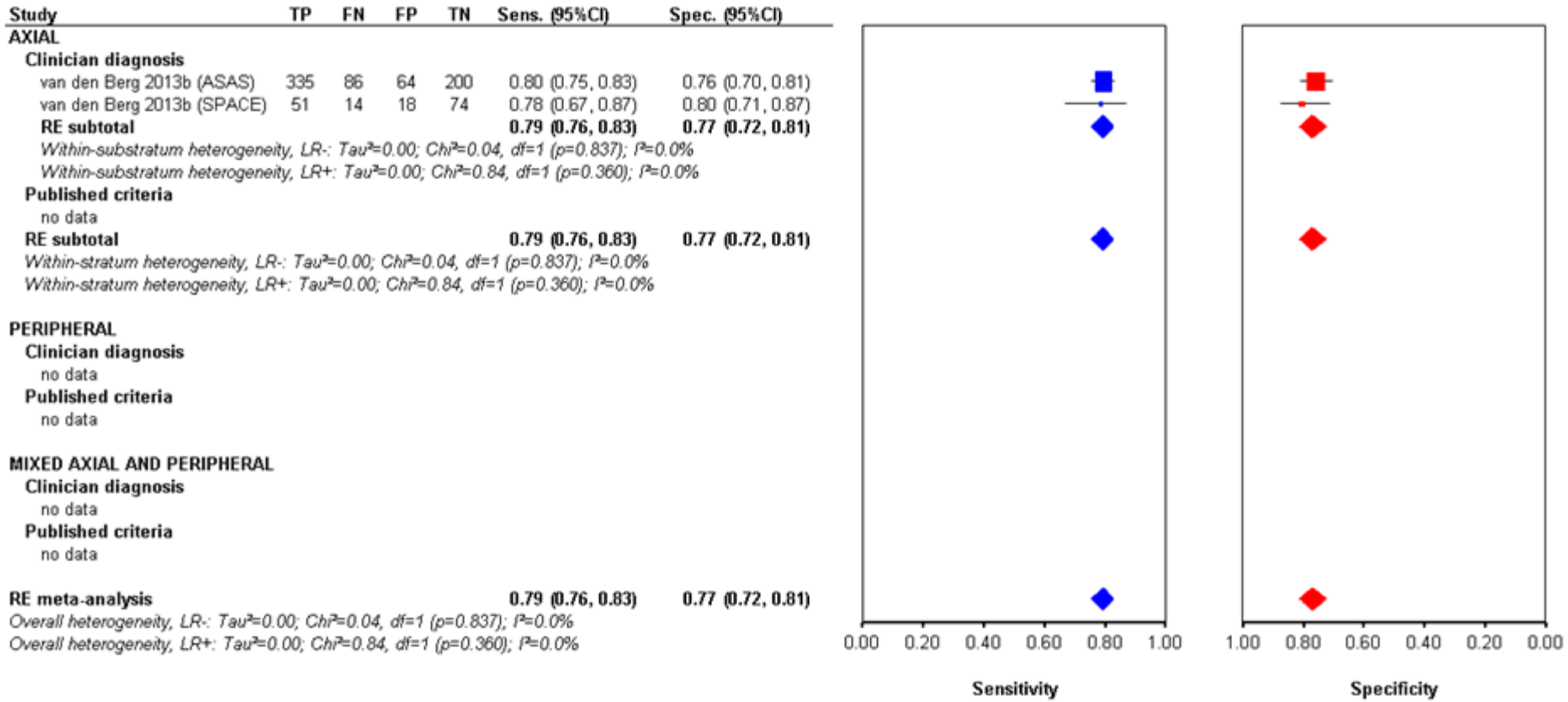


Figure 92: Berlin algorithm -- modification #2 – forest plot: sensitivity and specificity

G.1.8.4 ESSG criteria

Original ESSG criteria

Table 50: Original ESSG criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,357	1.62 (0.95, 2.77)	VERY LOW
LR-			No serious	No serious	No serious	No serious		0.42 (0.36, 0.49)	HIGH
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	No serious	n/a	No serious	Serious[d]	266	2.92 (1.86, 4.57)	MODERATE
LR-			No serious	n/a	No serious	Serious[f]		0.55 (0.46, 0.67)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	3 studies[g]	Cross-sectional	No serious	Serious[c]	Serious[h]	Serious[d]	907	2.68 (1.26, 5.72)	VERY LOW
LR-			No serious	No serious	No serious	Serious[f]		0.44 (0.34, 0.57)	MODERATE
ALL EVIDENCE POOLED									
LR+	6 studies[i]	Cross-sectional	No serious	Serious[c]	No serious	Serious[d]	2,530	2.27 (1.48, 3.46)	LOW
LR-			No serious	No serious	No serious	Serious[f]		0.46 (0.41, 0.52)	MODERATE

[a] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS)*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *Rudwaleit 2011*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[g] *D'Agostino 2011; Godfrin 2004 ; Tomero 2014*
[h] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*
[i] *Dougados 2015 (DESIR); D'Agostino 2011; Godfrin 2004 ; Rudwaleit 2009 (ASAS); Rudwaleit 2011; Tomero 2014*

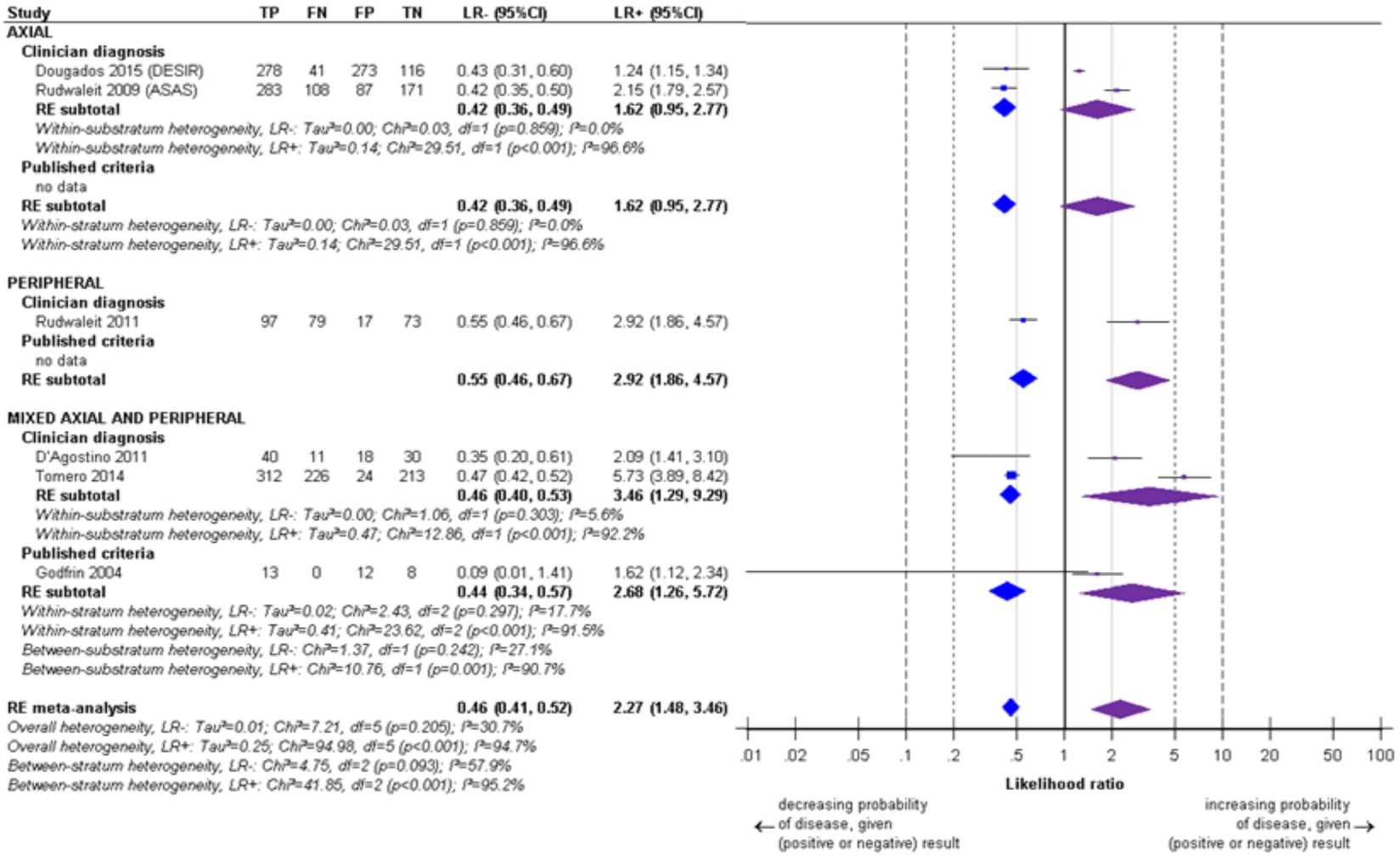


Figure 93: Original ESSG criteria – forest plot: likelihood ratios

Study	TP	FN	FP	TN	Sens. (95%CI)	Spec. (95%CI)
AXIAL						
Clinician diagnosis						
Dougados 2015 (DESIR)	278	41	273	116	0.87 (0.83, 0.90)	0.30 (0.25, 0.35)
Rudwaleit 2009 (ASAS)	283	108	87	171	0.72 (0.68, 0.77)	0.66 (0.60, 0.72)
RE subtotal					**0.81 (0.62, 0.91)**	**0.48 (0.17, 0.80)**
Within-substratum heterogeneity, LR-: Tau^2=0.43; Chi^2=22.17, df=1 (p<0.001); I^2=95.5%						
Within-substratum heterogeneity, LR+: Tau^2=1.16; Chi^2=79.18, df=1 (p<0.001); I^2=98.7%						
Published criteria						
no data						
RE subtotal					**0.81 (0.62, 0.91)**	**0.48 (0.17, 0.80)**
Within-stratum heterogeneity, LR-: Tau^2=0.43; Chi^2=22.17, df=1 (p<0.001); I^2=95.5%						
Within-stratum heterogeneity, LR+: Tau^2=1.16; Chi^2=79.18, df=1 (p<0.001); I^2=98.7%						
PERIPHERAL						
Clinician diagnosis						
Rudwaleit 2011	97	79	17	73	0.55 (0.48, 0.62)	0.81 (0.72, 0.88)
Published criteria						
no data						
RE subtotal					**0.55 (0.48, 0.62)**	**0.81 (0.72, 0.88)**
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
D'Agostino 2011	40	11	18	30	0.78 (0.65, 0.88)	0.63 (0.48, 0.75)
Tomero 2014	312	226	24	213	0.58 (0.54, 0.62)	0.90 (0.85, 0.93)
RE subtotal					**0.68 (0.45, 0.84)**	**0.80 (0.43, 0.95)**
Within-substratum heterogeneity, LR-: Tau^2=0.41; Chi^2=7.59, df=1 (p=0.006); I^2=86.8%						
Within-substratum heterogeneity, LR+: Tau^2=1.33; Chi^2=20.68, df=1 (p<0.001); I^2=95.2%						
Published criteria						
Godfrin 2004	13	0	12	8	0.96 (0.62, 1.00)	0.40 (0.22, 0.62)
RE subtotal					**0.74 (0.50, 0.89)**	**0.69 (0.33, 0.91)**
Within-stratum heterogeneity, LR-: Tau^2=0.57; Chi^2=11.67, df=2 (p=0.003); I^2=82.9%						
Within-stratum heterogeneity, LR+: Tau^2=1.67; Chi^2=38.14, df=2 (p<0.001); I^2=94.8%						
Between-substratum heterogeneity, LR-: Chi^2=4.08, df=1 (p=0.043); I^2=75.5%						
Between-substratum heterogeneity, LR+: Chi^2=17.46, df=1 (p<0.001); I^2=94.3%						
RE meta-analysis					**0.73 (0.60, 0.83)**	**0.65 (0.40, 0.84)**
Overall heterogeneity, LR-: Tau^2=0.43; Chi^2=93.68, df=5 (p<0.001); I^2=94.7%						
Overall heterogeneity, LR+: Tau^2=1.55; Chi^2=213.25, df=5 (p<0.001); I^2=97.7%						
Between-stratum heterogeneity, LR-: Chi^2=59.84, df=2 (p<0.001); I^2=96.7%						
Between-stratum heterogeneity, LR+: Chi^2=95.92, df=2 (p<0.001); I^2=97.9%						

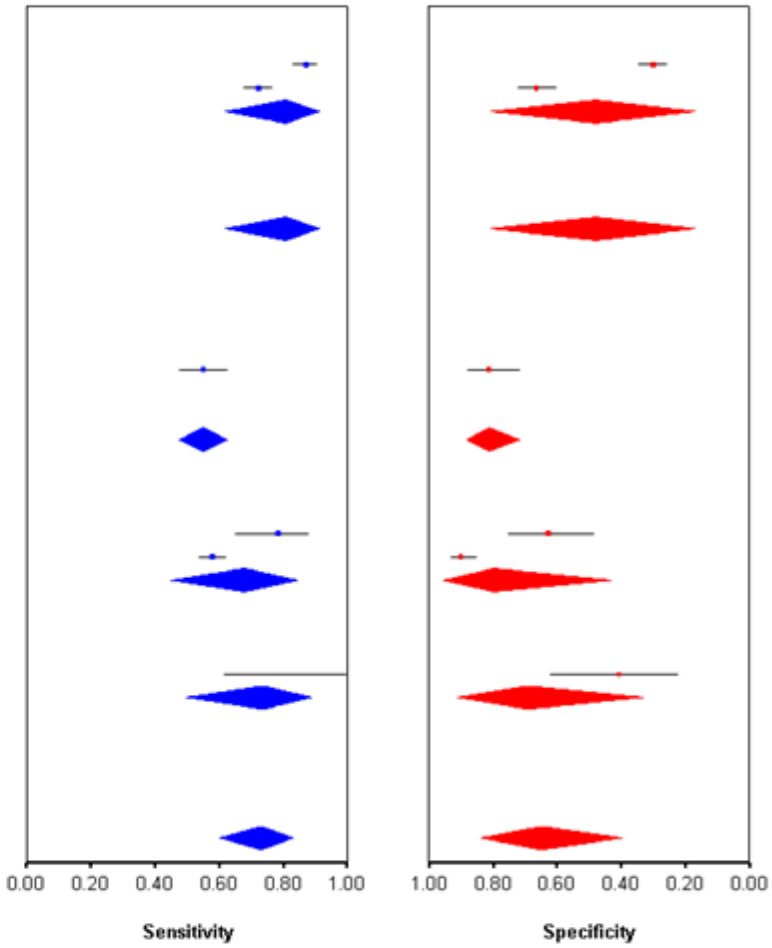


Figure 94: Original ESSG criteria – forest plot: sensitivity and specificity

Modified ESSG criteria

Table 51 Modified ESSG criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,357	1.70 (0.84, 3.42)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	No serious		0.28 (0.18, 0.46)	LOW
PERIPHERAL									
LR+	1 study[e]	Cross-sectional	No serious	n/a	No serious	No serious	266	3.31 (2.12, 5.15)	HIGH
LR-			No serious	n/a	No serious	Serious[f]		0.46 (0.37, 0.57)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	3 studies[g]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	1,623	2.08 (1.12, 3.84)	VERY LOW
LR-			No serious	Serious[c]	No serious	Serious[f]		0.34 (0.21, 0.55)	LOW

[a] *Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS)*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *Rudwaleit 2011*
[f] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*
[g] Dougados 2015 (DESIR); Rudwaleit 2009 (ASAS); Rudwaleit 2011

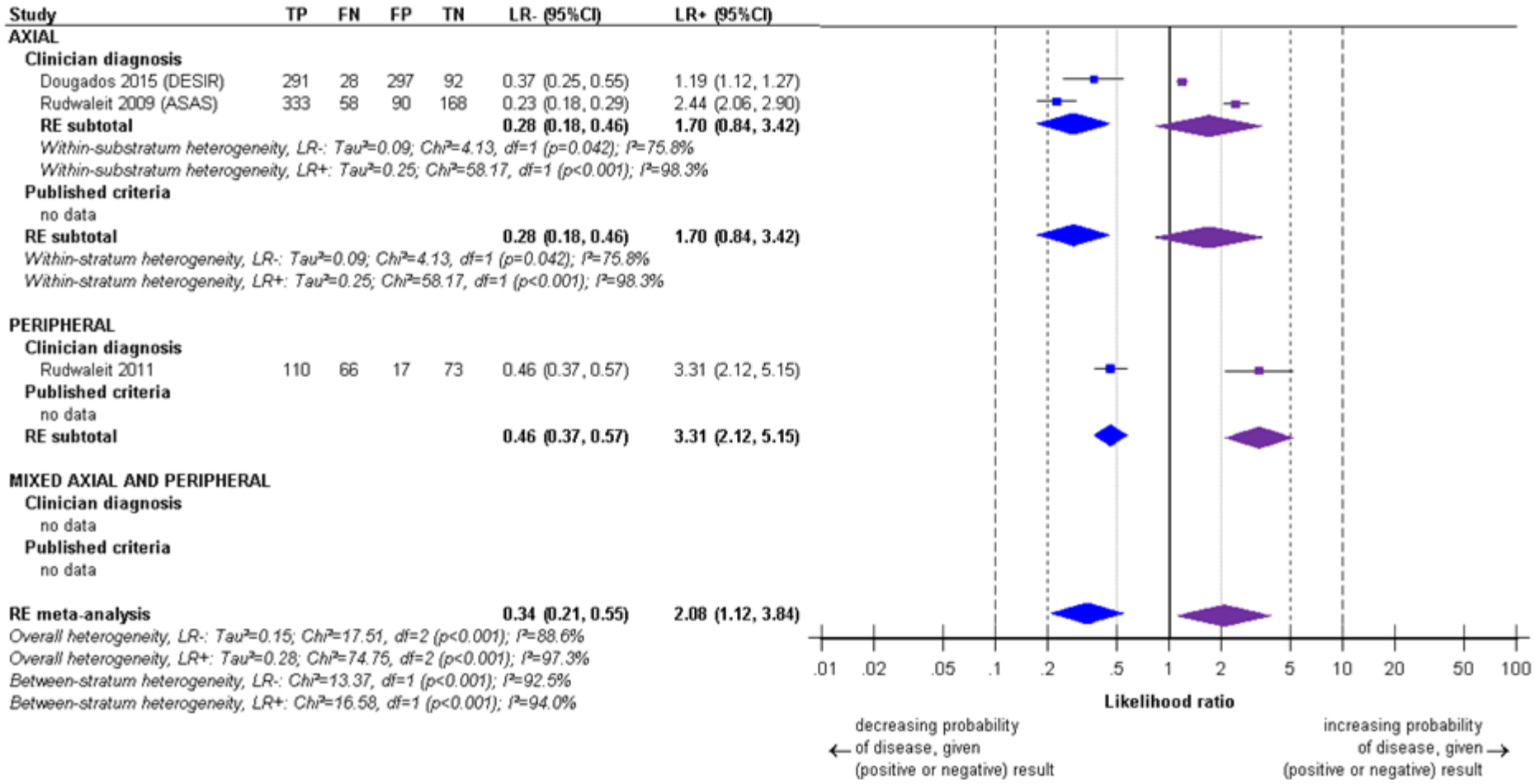


Figure 95: Modified ESSG criteria – forest plot: likelihood ratios

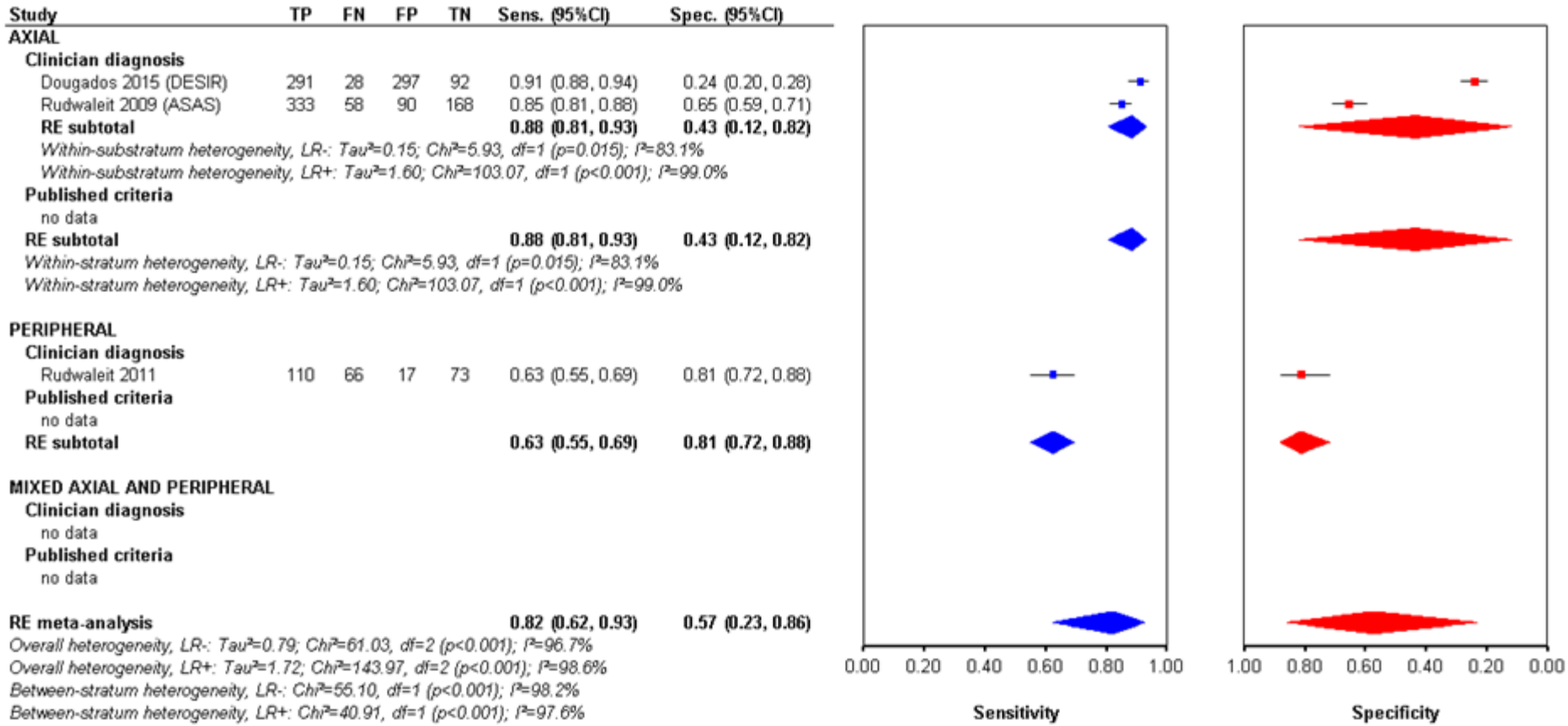


Figure 96: Modified ESSG criteria – forest plot: sensitivity and specificity

G.1.8.5 New York criteria

Original New York criteria

Table 52: Original New York criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	212	16.68 (8.19, 33.97)	MODERATE
LR-			Serious	n/a	No serious	Serious[b]		0.28 (0.15, 0.51)	LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	212	16.68 (8.19, 33.97)	MODERATE
LR-			Serious	n/a	No serious	Serious[b]		0.28 (0.15, 0.51)	LOW

[a] *Rigby 1993*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Rigby 1993	22	8	8	174	0.28 (0.15, 0.51)	16.68 (8.19, 33.97)
Published criteria						
no data						
RE subtotal					**0.28 (0.15, 0.51)**	**16.68 (8.19, 33.97)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.28 (0.15, 0.51)**	**16.68 (8.19, 33.97)**

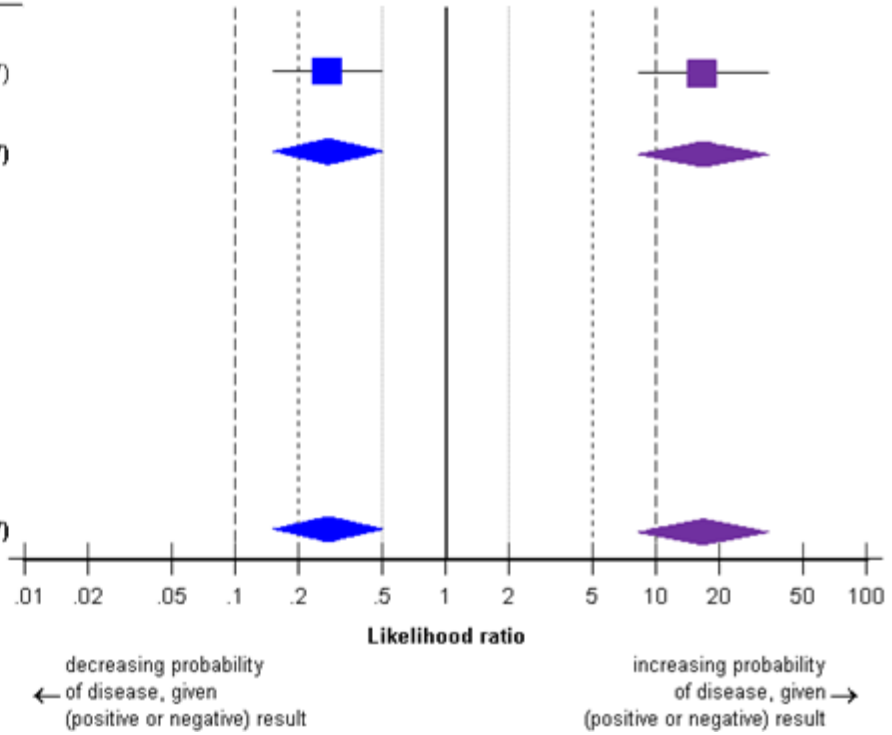


Figure 97: Original New York criteria – forest plot: likelihood ratios

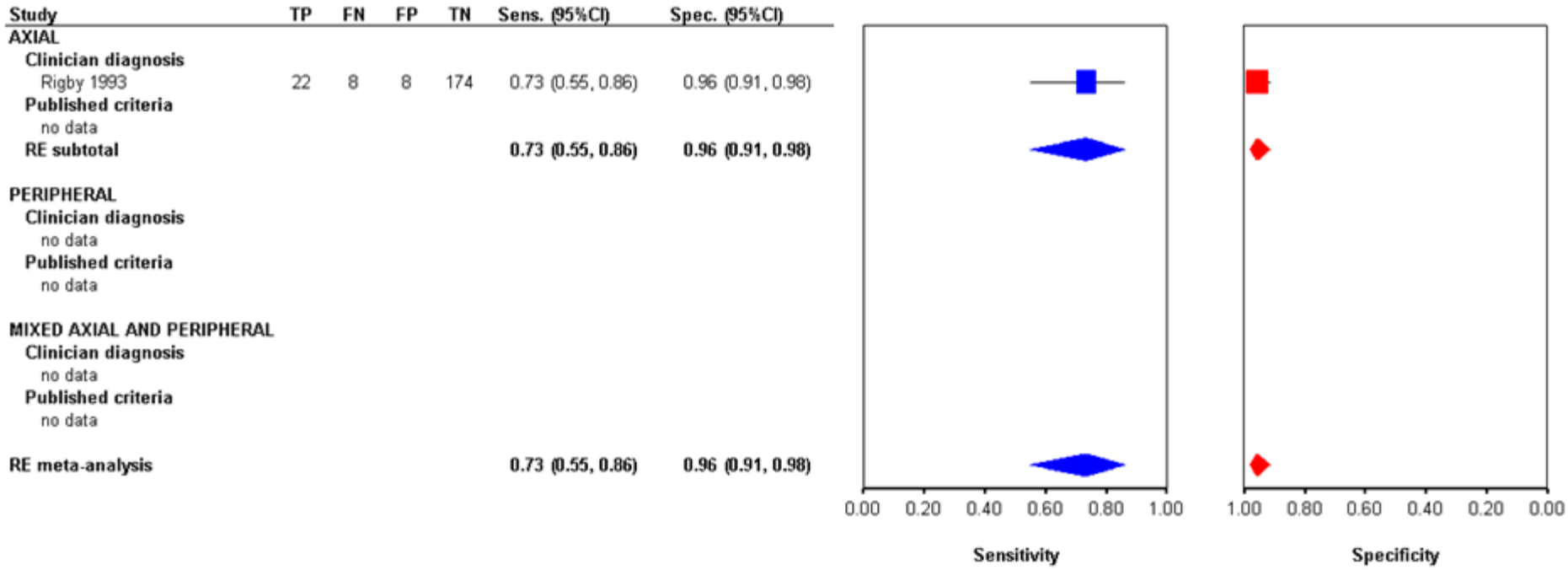


Figure 98: Original New York criteria – forest plot: sensitivity and specificity

Modified New York criteria

Table 53: Modified New York criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	920	7.75 (0.88, 67.89)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[e]		0.40 (0.12, 1.34)	VERY LOW
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	Serious[b]	Serious[c]	No serious	Serious[d]	920	7.75 (0.88, 67.89)	VERY LOW
LR-			Serious[b]	Serious[c]	No serious	Serious[e]		0.40 (0.12, 1.34)	VERY LOW

[a] *Dougados 2015 (DESIR); Rigby 1993*
[b] *>33.3% of weight in meta-analysis comes from studies with serious risk of bias*
[c] *I2 ≥ 50%*
[d] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*
[e] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR- spans 0.5).*

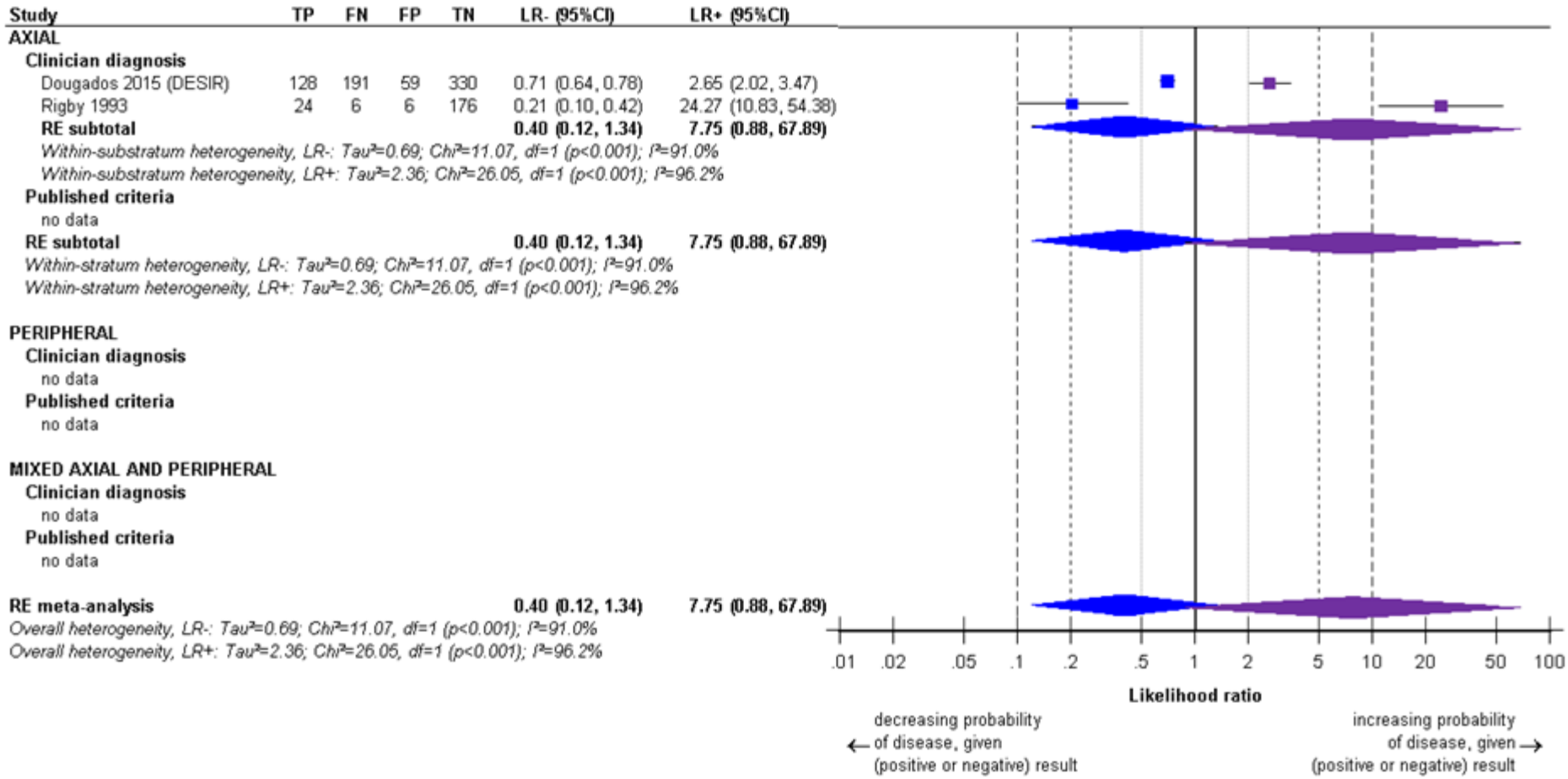


Figure 99: Modified New York criteria – forest plot: likelihood ratios

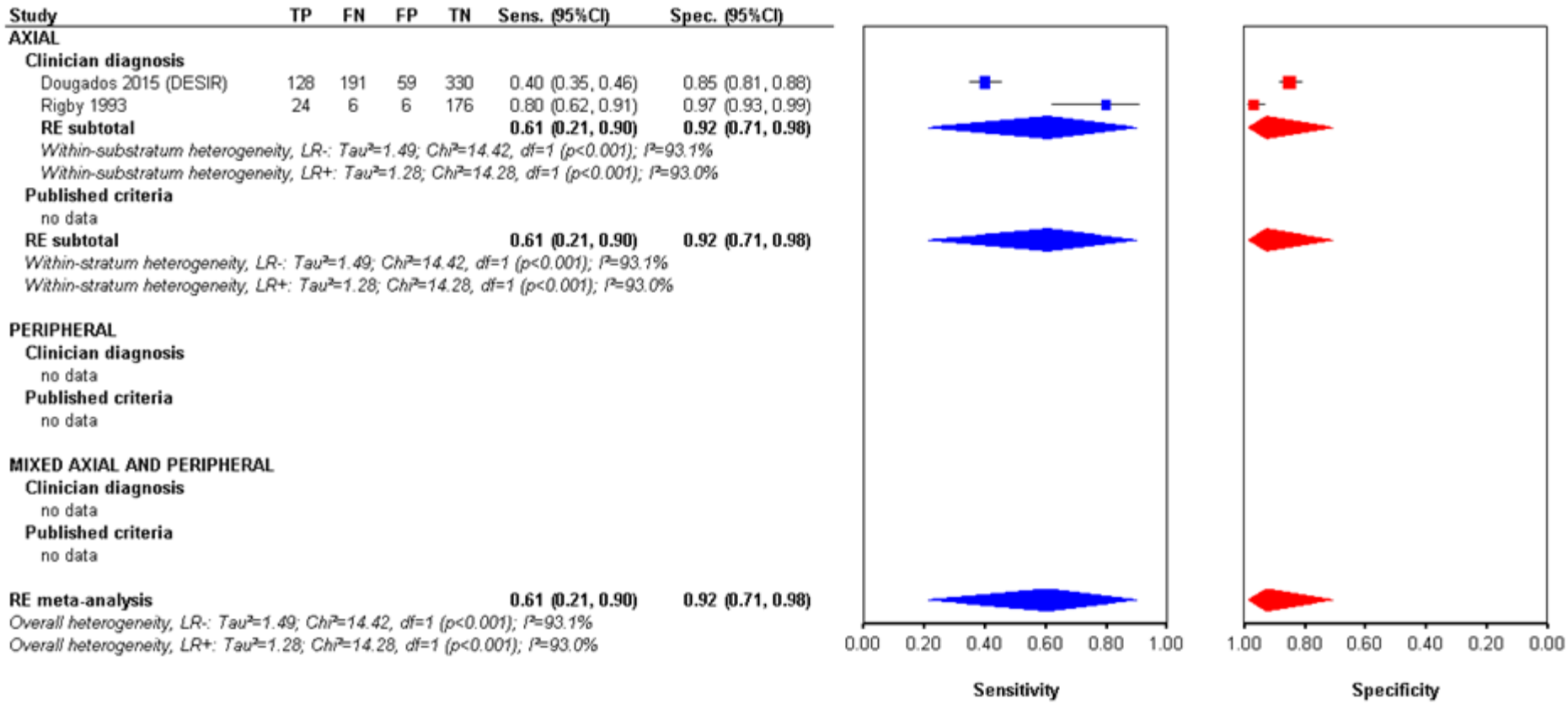


Figure 100: Modified New York criteria – forest plot: sensitivity and specificity

G.1.8.6 Rome criteria

Rome criteria (clinical)

Table 54: Rome criteria (clinical) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	Serious[b]	212	2.21 (1.08, 4.49)	LOW
LR-			Serious	n/a	No serious	No serious		0.83 (0.67, 1.04)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	Serious[b]	212	2.21 (1.08, 4.49)	LOW
LR-			Serious	n/a	No serious	No serious		0.83 (0.67, 1.04)	MODERATE

[a] *Rigby 1993*
[b] *At a 95% confidence level, data are consistent with both meaningful predictive value and no predictive value (i.e. 95% CI for LR+ spans 2).*

Study	TP	FN	FP	TN	LR- (95%CI)	LR+ (95%CI)
AXIAL						
Clinician diagnosis						
Rigby 1993	8	22	22	160	0.83 (0.67, 1.04)	2.21 (1.08, 4.49)
Published criteria						
no data						
RE subtotal					**0.83 (0.67, 1.04)**	**2.21 (1.08, 4.49)**
PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
MIXED AXIAL AND PERIPHERAL						
Clinician diagnosis						
no data						
Published criteria						
no data						
RE meta-analysis					**0.83 (0.67, 1.04)**	**2.21 (1.08, 4.49)**

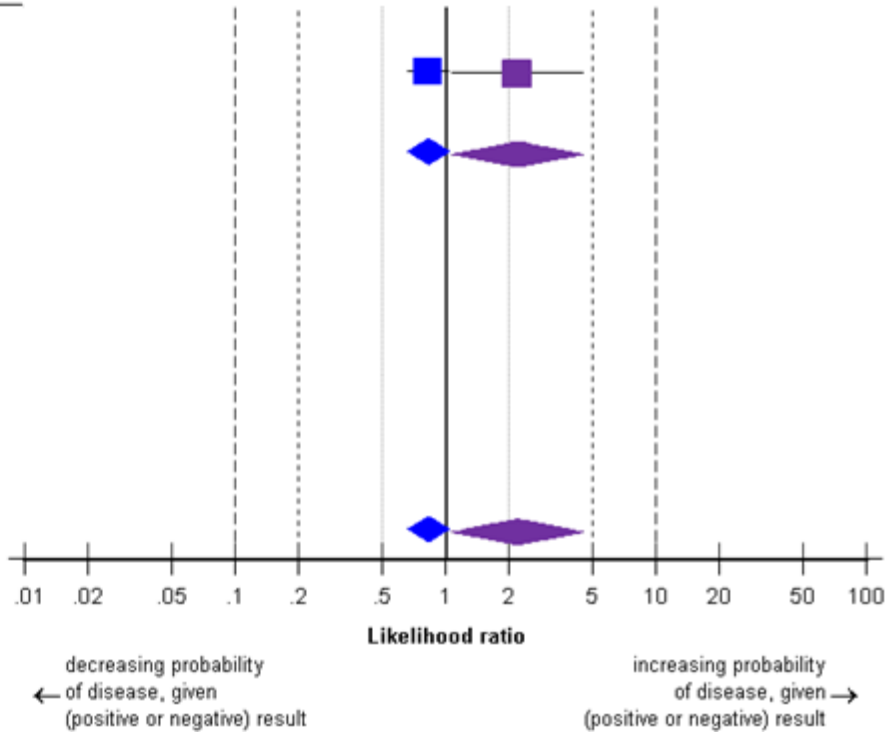


Figure 101: Rome criteria (clinical) – forest plot: likelihood ratios

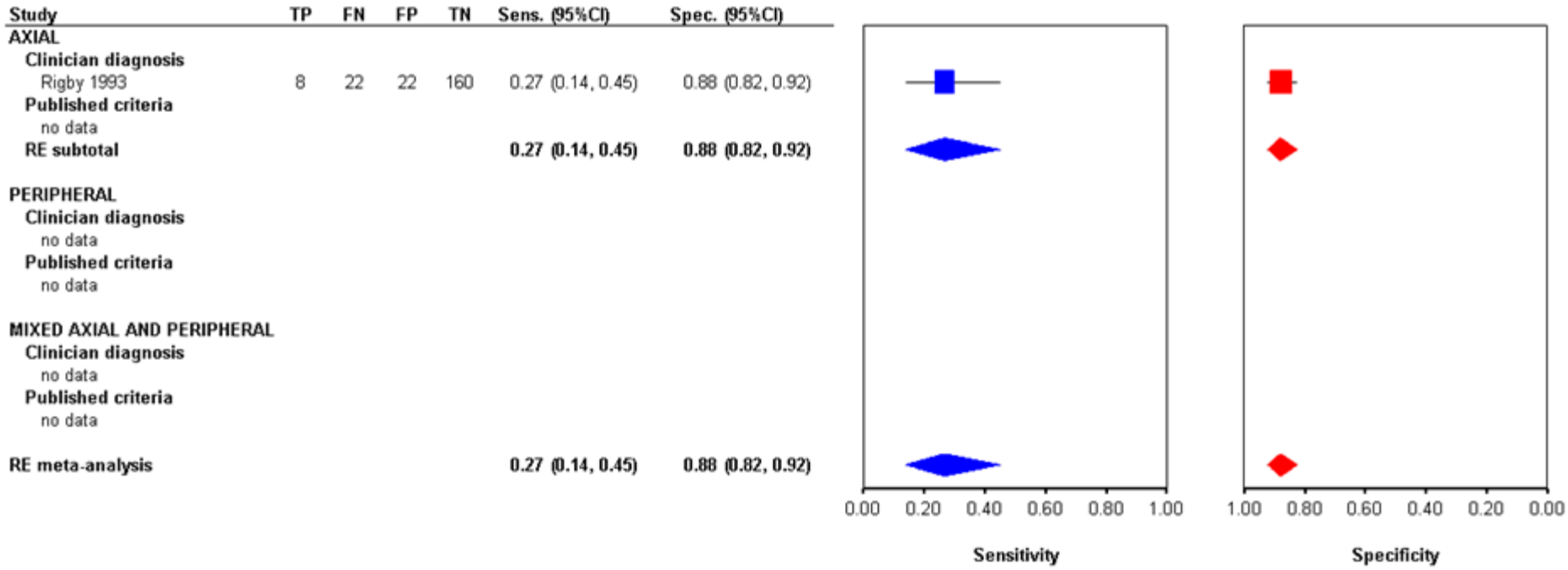


Figure 102: Rome criteria (clinical) – forest plot: sensitivity and specificity

Rome criteria (radiographic)

Table 55: Rome criteria (radiographic) – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	212	39.43 (14.81, 104.99)	MODERATE
LR-			Serious	n/a	No serious	No serious		0.14 (0.05, 0.34)	MODERATE
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	212	39.43 (14.81, 104.99)	MODERATE
LR-			Serious	n/a	No serious	No serious		0.14 (0.05, 0.34)	MODERATE

[a] *Rigby 1993*

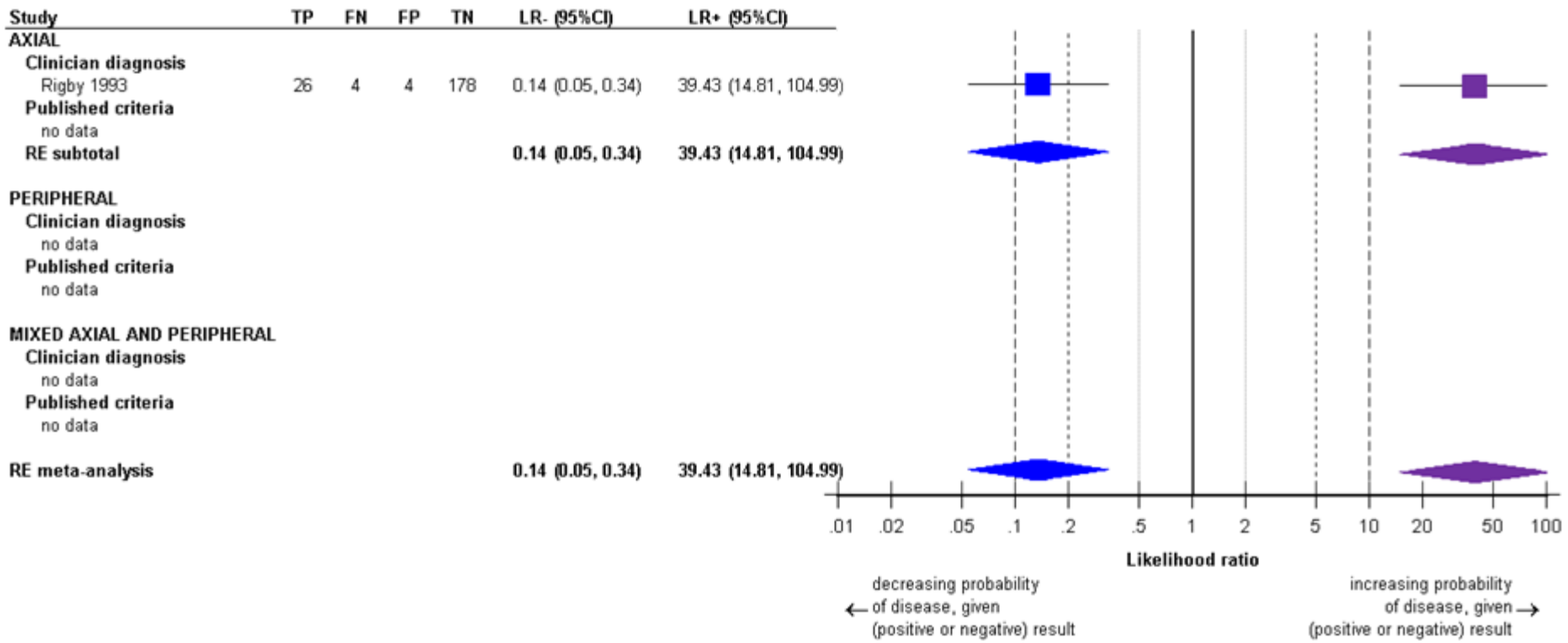


Figure 103: Rome criteria (radiographic) – forest plot: likelihood ratios

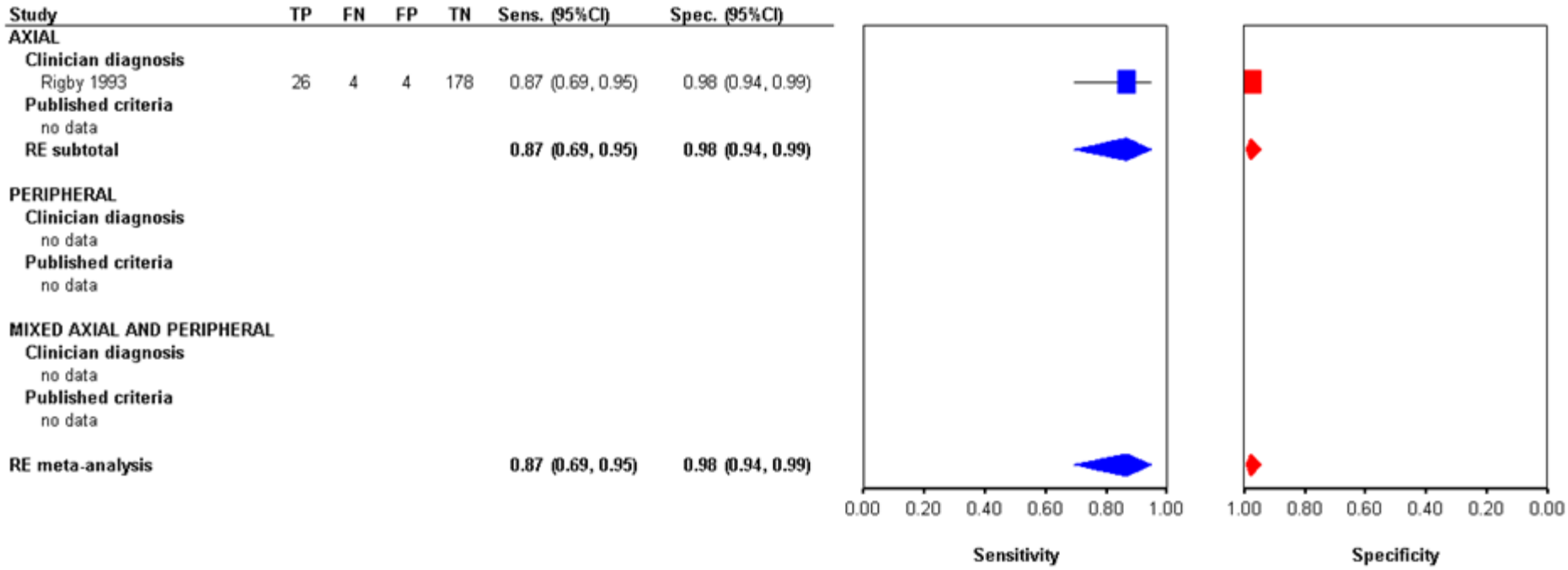


Figure 104: Rome criteria (radiographic) – forest plot: sensitivity and specificity

G.1.8.7 ASAS peripheral criteria

Table 56: ASAS peripheral criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	266	4.38 (2.79, 6.88)	HIGH
LR-			No serious	n/a	No serious	No serious		0.27 (0.20, 0.36)	HIGH
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	No serious	n/a	No serious	No serious	266	4.38 (2.79, 6.88)	HIGH
LR-			No serious	n/a	No serious	No serious		0.27 (0.20, 0.36)	HIGH
LR-			No serious	n/a	No serious	No serious		0.27 (0.20, 0.36)	HIGH

[a]Rudwaleit 2011

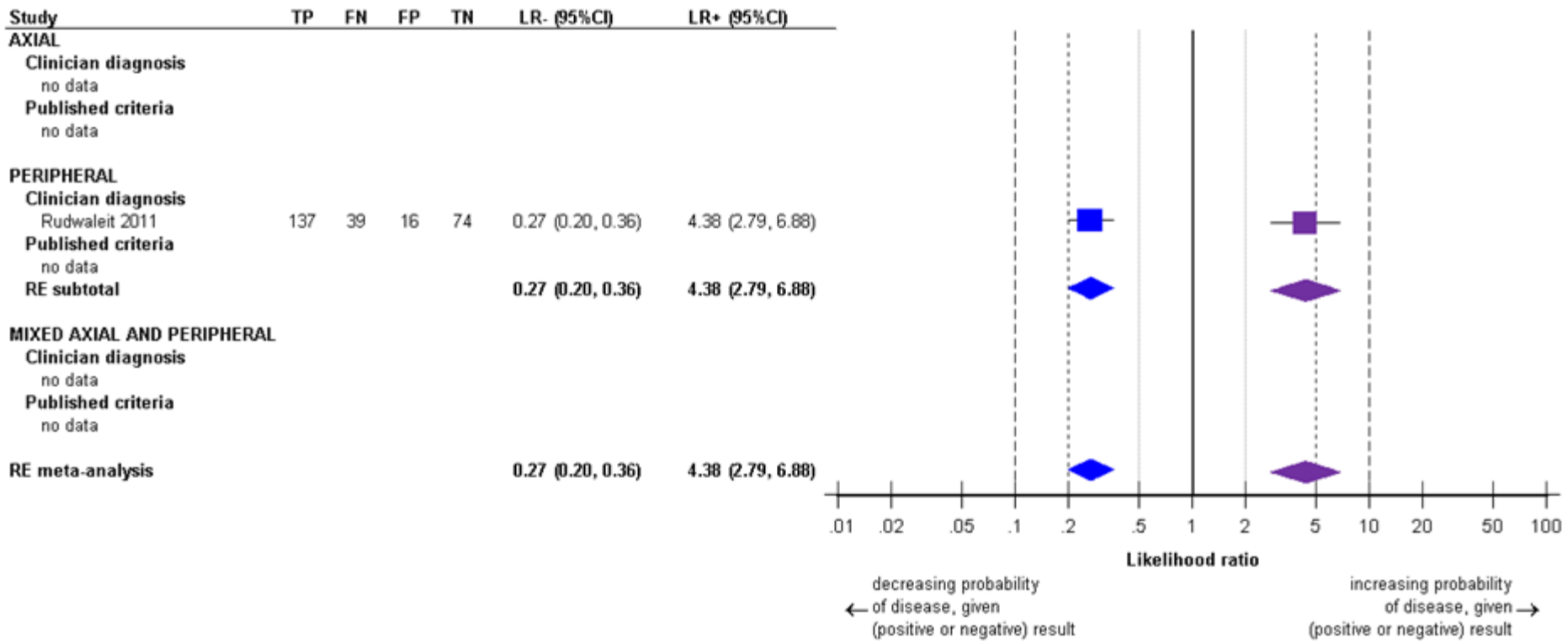


Figure 105: ASAS peripheral criteria – forest plot: likelihood ratios

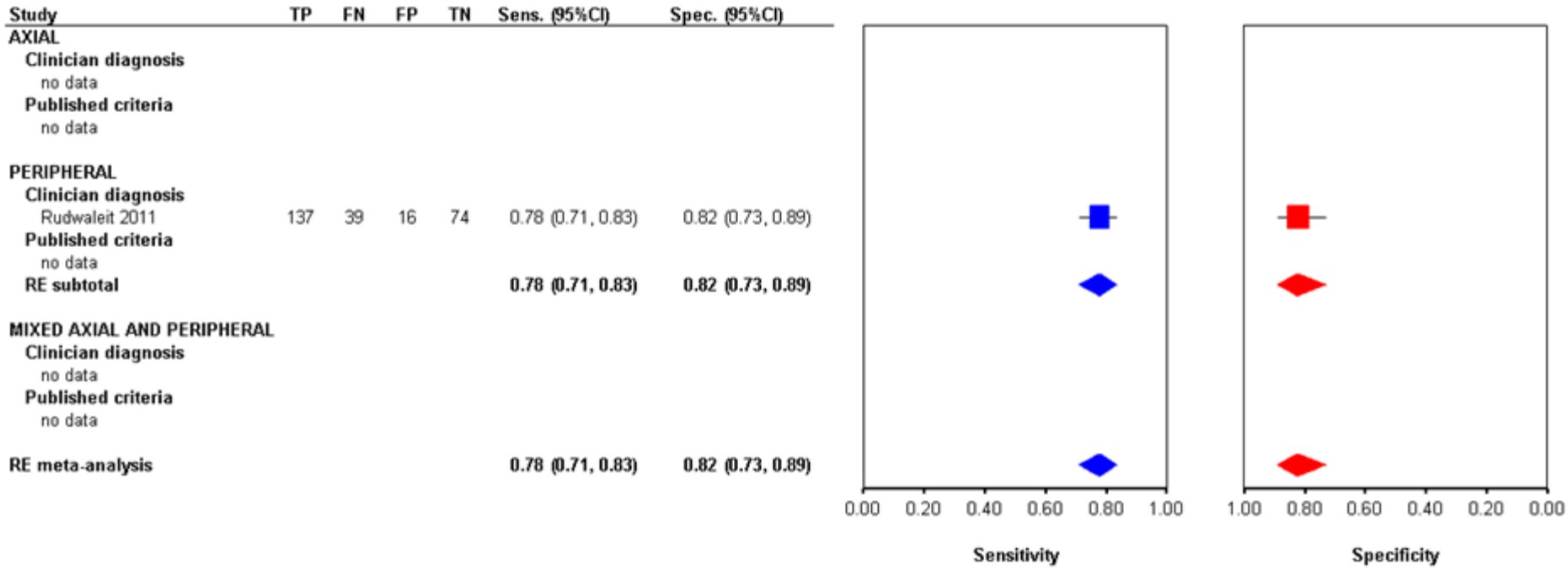


Figure 106: ASAS peripheral criteria – forest plot: sensitivity and specificity

G.1.8.8 French Society for Rheumatology criteria for reactive arthritis

Table 57 French Society for Rheumatology criteria – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95%CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	217	10.19 (6.01, 17.26)	MODERATE
LR-			Serious	n/a	No serious	No serious		0.23 (0.10, 0.49)	MODERATE
MIXED AXIAL AND PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
ALL EVIDENCE POOLED									
LR+	1 study[a]	Cross-sectional	Serious	n/a	No serious	No serious	217	10.19 (6.01, 17.26)	MODERATE
LR-			Serious	n/a	No serious	No serious		0.23 (0.10, 0.49)	MODERATE

[a] *Hulsemann 1999*

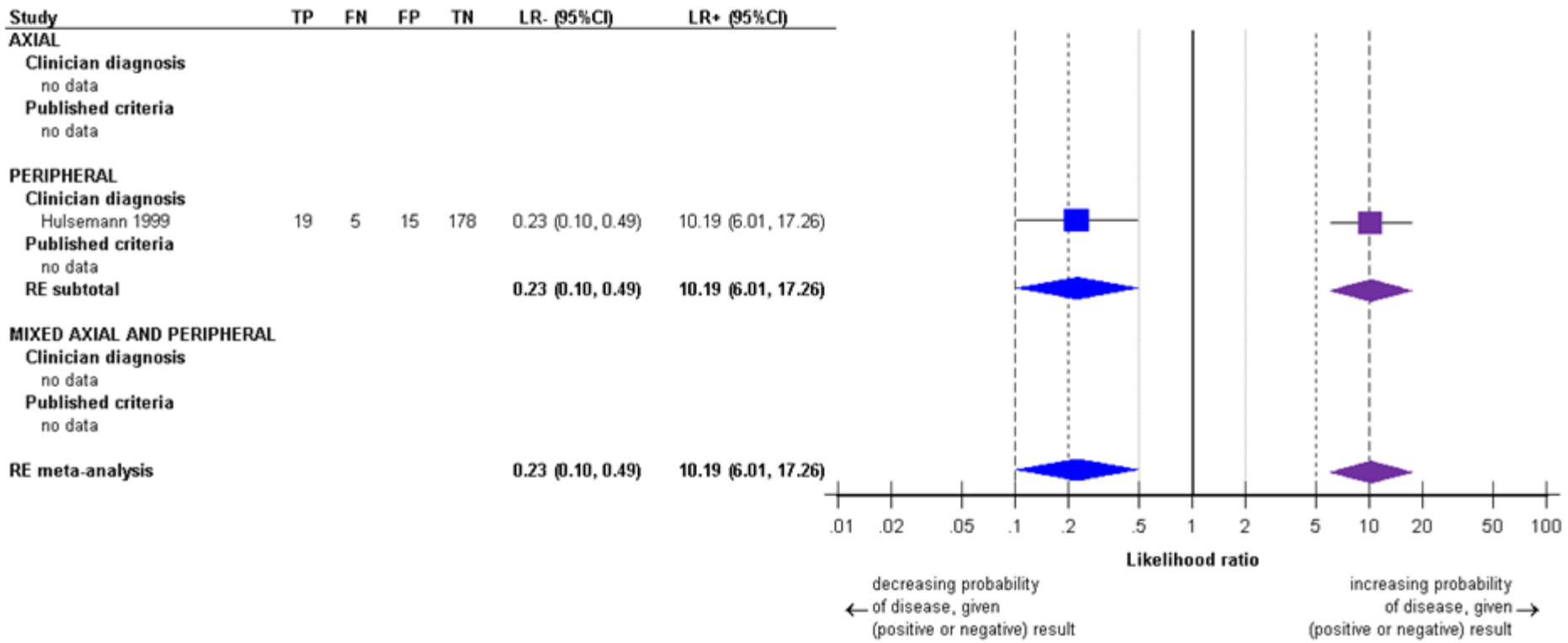


Figure 107: French Society for Rheumatology criteria for reactive arthritis – forest plot: likelihood ratios

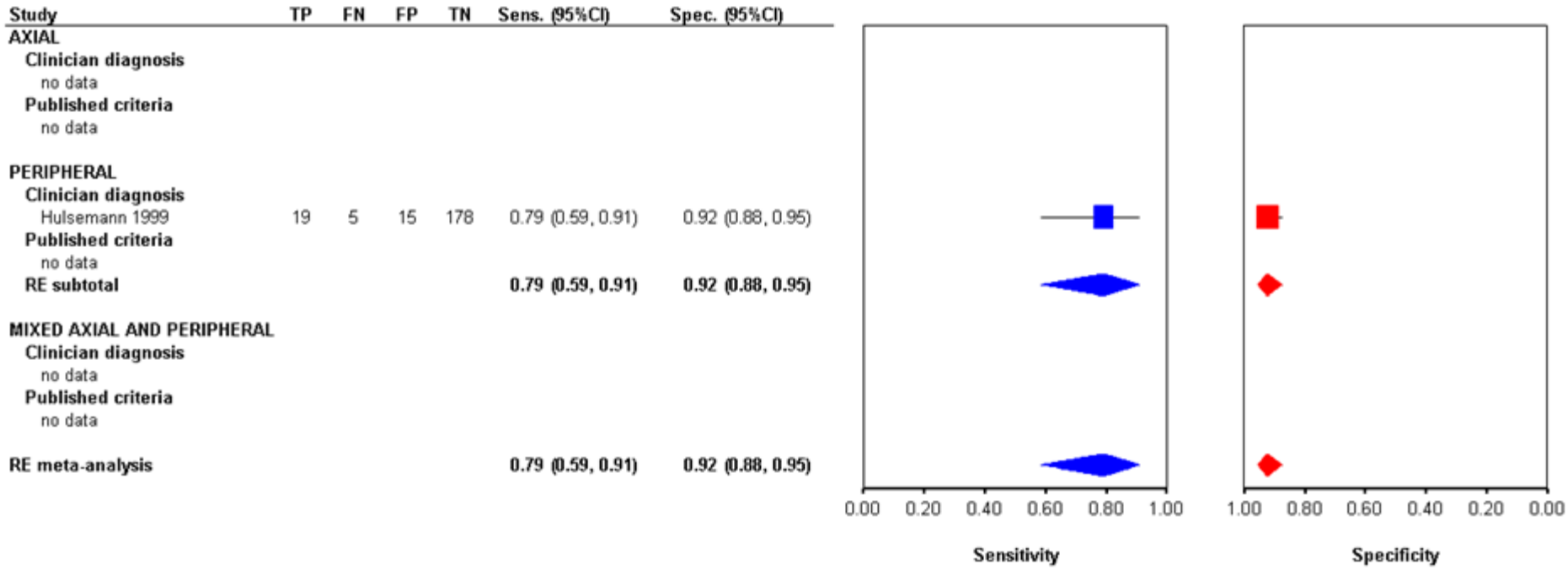


Figure 108: French Society for Rheumatology criteria for reactive arthritis – forest plot: sensitivity and specificity

G.1.8.9 Diagnosis of spondyloarthritis in people presenting with acute anterior uveitis

DUET algorithm for acute anterior uveitis

Table 58: DUET algorithm for acute anterior uveitis – GRADE table

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Total N	Summary of findings (95% CI)	Quality
AXIAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
PERIPHERAL									
LR+	0 studies	-	-	-	-	-	-	-	-
LR-			-	-	-	-		-	-
MIXED AXIAL AND PERIPHERAL									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	173	48.28 (12.23, 190.51)	MODERATE
LR-			No serious	No serious	Serious[b]	No serious		0.04 (0.01, 0.13)	MODERATE
ALL EVIDENCE POOLED									
LR+	2 studies[a]	Cross-sectional	No serious	No serious	Serious[b]	No serious	173	48.28 (12.23, 190.51)	MODERATE
LR-			No serious	No serious	Serious[b]	No serious		0.04 (0.01, 0.13)	MODERATE

[a] *Haroon 2015; Haroon 2015*

[b] *>33.3% of weight in meta-analysis comes from studies with suboptimal reference standard (published classification criteria, rather than expert diagnosis)*

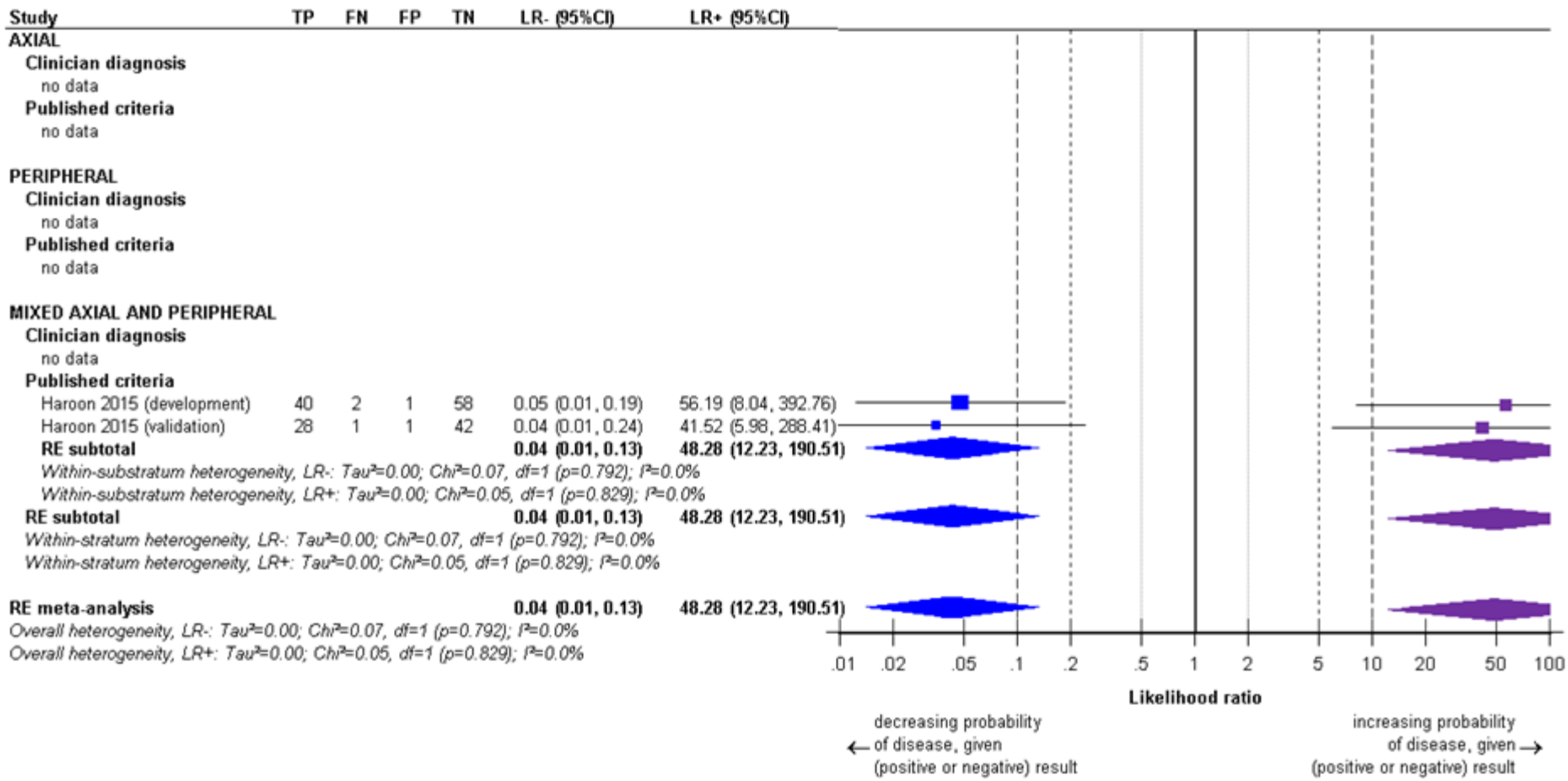


Figure 109: DUET algorithm for acute anterior uveitis – forest plot: likelihood ratios

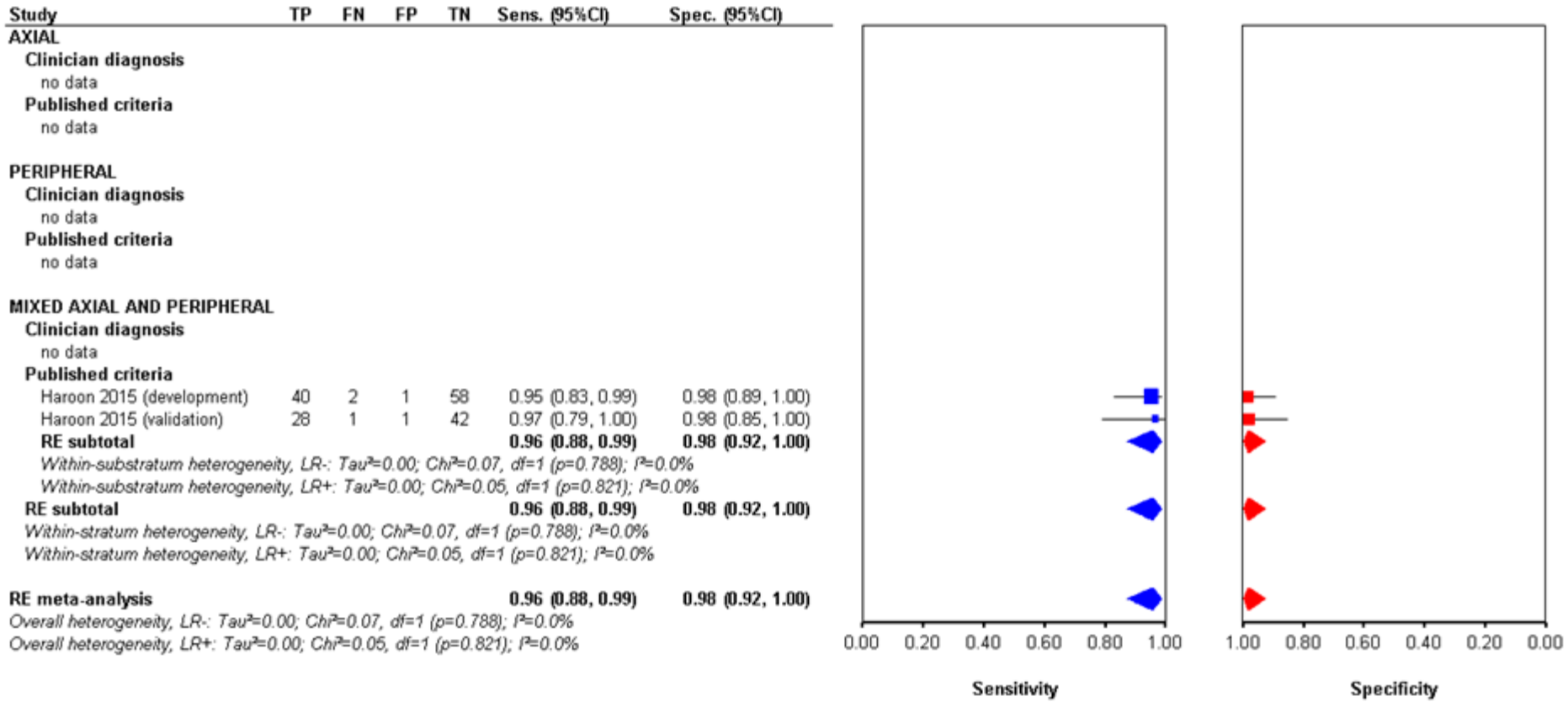


Figure 110: DUET algorithm for acute anterior uveitis – forest plot: sensitivity and specificity

G.1.9 Microbiology testing in Reactive Arthritis

Review Question 11

- What is the diagnostic utility of testing for infection such as salmonella, shigella, yersinia, campylobacter and chlamydia in cases of suspected reactive arthritis?

Table 59 GRADE table for microbiology testing in reactive arthritis

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Total N	Summary of findings (95%CI)	Quality
Salmonella – stool culture – post-outbreak										
LR+	1 (Locht)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	29	1.25 (0.89, 1.78)	LOW
LR-				n/a		Serious[2]			0.24 (0.03, 2.00)	LOW
Salmonella - any antibodies – post-outbreak										
LR+	1 (Mattila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	1.57 (0.89, 1.78)	LOW
LR-				n/a		Serious[2]			0.38 (0.14, 1.02)	LOW
Salmonella – IgA antibodies – post-outbreak										
LR+	1 (Mattila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	2.61 (0.56, 12.10)	LOW
LR-				n/a		Serious[2]			0.85 (0.65, 1.10)	LOW
Salmonella – IgM antibodies – post-outbreak										
LR+	1 (Mattila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	1.48 (0.94, 2.33)	LOW
LR-				n/a		Serious[2]			0.48 (0.20, 1.15)	LOW
Salmonella – IgG antibodies – post-outbreak										
LR+	1 (Mattila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	1.57 (1.01, 2.43)	LOW
LR-				n/a		Serious[2]			0.38 (0.14, 1.02)	LOW
Campylobacter, Salmonella and Yersinia – antibodies – post-outbreak										
LR+	1 (Uotila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	1.71 (0.56, 5.26)	LOW
LR-				n/a		Serious[2]			0.86 (0.62, 1.19)	LOW
Campylobacter, Salmonella and Yersinia – faecal culture – post-outbreak										

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Total N	Summary of findings (95%CI)	Quality
LR+	1 (Uotila)	Cross-sectional	Not serious	n/a	Serious[1]	Serious[2]	-	45	0.76 (0.14, 4.13)	LOW
LR-				n/a		Serious[2]			1.03 (0.84, 1.27)	LOW
Yersinia – IgA – 1-2 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	138	1.07 (0.98, 1.16)	LOW
LR-				Not serious		Serious[2]			0.15 (0.01, 1.52)	LOW
Yersinia – IgM – 1-2 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	138	0.97 (0.83, 1.13)	LOW
LR-				Not serious		Serious[2]			1.17 (0.53, 2.57)	LOW
Yersinia – IgG – 1-2 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	138	0.99 (0.91, 1.07)	LOW
LR-				Not serious		Serious[2]			1.43 (0.33, 6.22)	LOW
Yersinia – IgA – 6-8 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Serious	Serious[1]	Serious[2]	-	129	3.46 (0.81, 14.84)	VERY LOW
LR-				Not serious		Not serious			0.26 (0.15, 0.46)	MODERATE
Yersinia – IgM – 6-8 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	129	0.93 (0.57, 1.54)	LOW
LR-				Not serious		Serious[2]			1.02 (0.91, 1.15)	LOW
Yersinia – IgG – 6-8 months										

Measure	Studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Total N	Summary of findings (95%CI)	Quality
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	129	1.27 (1.02, 1.59)	LOW
LR-				Not serious		Serious[2]			0.39 (0.16, 0.96)	LOW
Yersinia – IgA – 12-16 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	85	3.35 (1.36, 8.27)	LOW
LR-				Not serious		Not serious			0.19 (0.08, 0.42)	MODERATE
Yersinia – IgM – 12-16 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	85	1.09 (0.44, 2.71)	LOW
LR-				Not serious		Serious[2]			1.01 (0.87, 1.16)	LOW
Yersinia – IgG – 12-16 months										
LR+	2 (Granfors, Toivanen)	Prospective cohort	Not serious	Not serious	Serious[1]	Serious[2]	-	85	1.76 (0.87, 3.53)	LOW
LR-				Not serious		Serious[2]			0.48 (0.28, 0.81)	LOW

[1]Does not cover full population of interest. [2]Confidence intervals for likelihood ratio contain multiple clinically distinct scenarios.

G.2 Pharmacological management

G.2.1 Pharmacological interventions for axial symptoms of spondyloarthritis

Review question 20

- What is the comparative effectiveness of non-steroidal anti-inflammatory drugs (NSAIDs) for management of axial symptoms of spondyloarthritis?

Table 60 NSAID therapy – network meta-analyses

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Global pain (VAS)*	24[a]	Serious[1]	Serious[2]	Not serious	Not serious	See NMA graph	Low
Discontinuation due to adverse events	19[b]	Serious[1]	Serious[2]	Not serious	Not serious	See NMA graph	Low
Discontinuation due to lack of efficacy	14[c]	Serious[1]	Not serious	Not serious	Not serious	See NMA graph	Moderate

*All outcomes from studies were converted to a 0-100 scale before running the analysis. Missing standard deviations were imputed based on the distribution of standard deviations from those studies were they were reported.

[a]Astorga 1987; Barkhuizen 2006; Batlle-Gualda 1996; Bird 1986; Burry 1980; Dougados 1999, Dougados 2001; Gibson 1980; Good 1977; Johnsen 1992; Juvakoski 1982, Khan 1987, Lomen 1986; Nahir 1980; Pasero 1994; Rejholec 1980; Schwarzer 1990; Shipley 1980; Sieper 2008; Sturrock 2008; Tannenbaum 1984; van der Heijde 2005; Villa Alcazar 1996; Walker 2016

[b]Barkhuizen 20066; Batlle Gualda 1996; Bird 1986; Burry 1980; Dougados 1999; Dougados 2001; Good 1977; Juvakoski 1982; Khan 1987; Lomen 1986; Mayrhofer 1990; Shipley 1980; Sieper 2008; Sturrock 1974; Sydnes 1984; Tannenbaum 1984; Van der Heijde 2005; Villa Alacazar 1996; Walker 2016

[c]Barkhuizen 2006; Batlle Gualda 1996; Dougados 1999; Dougados 2001; Juvakoski 1982; Khan 1987; Lomen 1986; Mayrhofer 1990; Schwarzer 1990; Shipley 1980; Sieper 2008; Tannenbaum 1984; Van der Heijde 2005; Villa Alacazar 1996

[1]Many included studies have poorly reported methods, which makes it difficult to rule out the possibility of high levels of bias in the studies

[2]Random effects model selected using Deviance Information Criterion

Meta-analysis - Pain

Table 61 Model fit

Model	Number of data points	Residual Deviance over studies with complete data	Residual Deviance over all studies	DIC

Model	Number of data points	Residual Deviance over studies with complete data	Residual Deviance over all studies	DIC
RE consistency	53	30.96	52.26	847.67
RE inconsistency	53	32.5	54.16	855.21
FE consistency	53	41.03	84.64	872.49

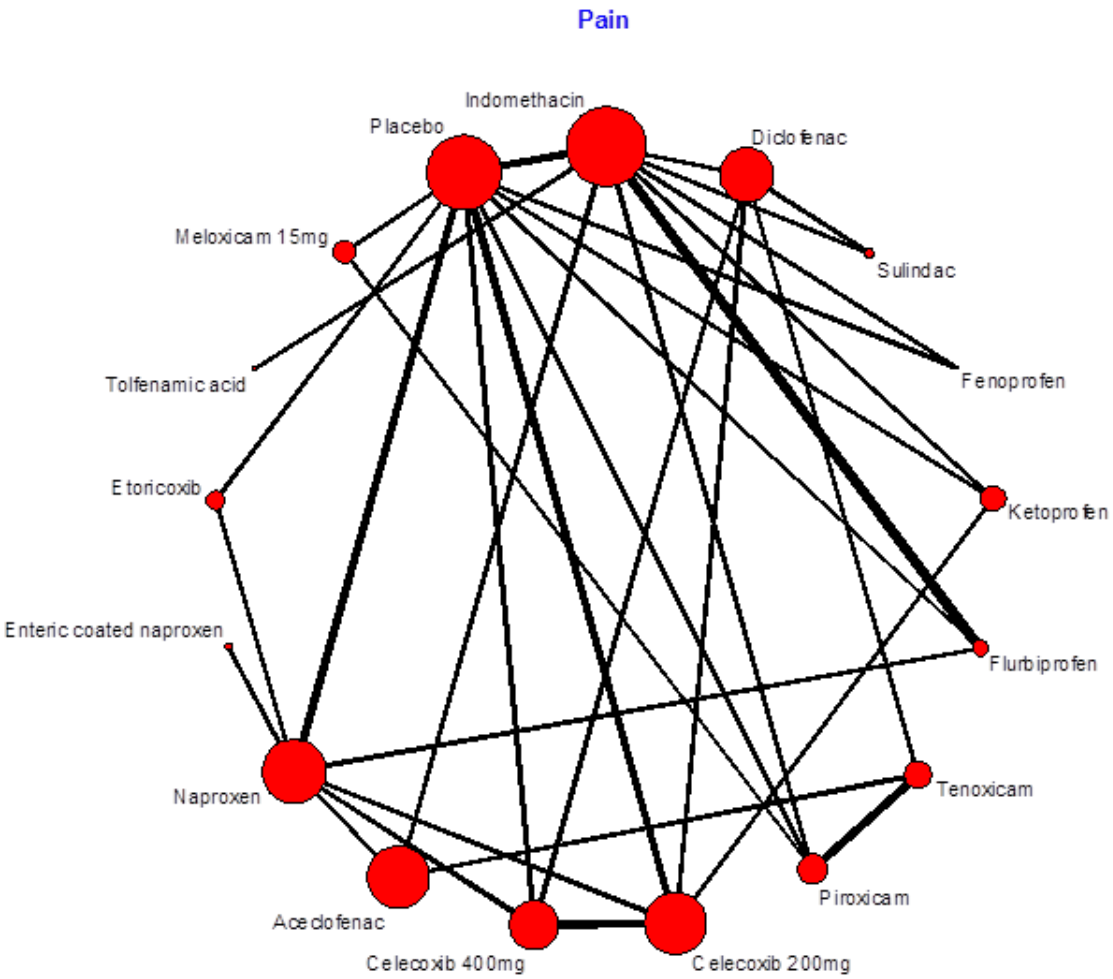


Figure 111 Network diagram

Table 62 Random effects consistency model: mean difference (95% credible interval) – positive value indicates worse outcome for row

Indomethacin																
-5.0 (-13.9, 3.9)	**Diclofenac**															
6.0 (-6.3, 18.3)	11.0 (-1.0, 23.0)	**Sulindac**														
3.1 (-14.2, 20.7)	8.1 (-10.9, 27.4)	-2.9 (-23.7, 18.3)	**Fenoprofen**													
4.9 (-5.6, 15.8)	9.9 (-2.5, 22.8)	-1.1 (-16.6, 14.7)	1.8 (-18.0, 21.6)	**Ketoprofen**												
1.7 (-6.4, 9.9)	6.7 (-4.6, 18.1)	-4.3 (-18.7, 10.2)	-1.4 (-20.2, 17.1)	-3.2 (-15.9, 9.1)	**Flurbiprofen**											
0.6 (-10.7, 12.5)	5.6 (-6.7, 18.5)	-5.4 (-21.1, 10.9)	-2.5 (-22.8, 17.9)	-4.3 (-18.9, 10.6)	-1.2 (-14.1, 12.7)	**Tenoxicam**										
-1.4 (-11.9, 9.9)	3.7 (-8.7, 16.9)	-7.4 (-22.8, 8.8)	-4.5 (-24.2, 15.5)	-6.3 (-20.1, 8.1)	-3.2 (-15.3, 10.0)	-2.0 (-13.5, 9.7)	**Piroxicam**									
-0.9 (-10.0, 8.9)	4.2 (-5.4, 14.3)	-6.9 (-20.7, 7.6)	-4.0 (-22.8, 14.8)	-5.8 (-17.0, 5.5)	-2.7 (-13.4, 8.8)	-1.4 (-14.8, 11.6)	0.6 (-12.3, 12.9)	**Celecoxib 200mg**								
-0.7 (-10.7, 10.1)	4.4 (-5.7, 15.0)	-6.7 (-21.0, 8.2)	-3.8 (-23.2, 15.7)	-5.6 (-18.4, 7.4)	-2.4 (-14.0, 9.8)	-1.2 (-15.1, 12.4)	0.8 (-12.9, 13.8)	0.2 (-8.9, 9.2)	**Celecoxib 400mg**							

-0.5 (-9.8, 9.3)	4.6 (-7.0, 16.4)	-6.5 (-21.3, 8.6)	-3.6 (-23.1, 15.8)	-5.4 (-18.9, 8.1)	-2.2 (-13.7, 9.56)	-1.0 (-12.1, 9.5)	1.0 (-11.8, 13.0)	0.5 (-11.7, 12.1)	0.2 (-12.5, 12.6)	**Aceclofenac**						
-4.2 (-12.9, 5.3)	0.8 (-9.7, 12.1)	-10.3 (-24.3, 4.7)	-7.3 (-25.8, 11.4)	-9.1 (-21.1, 3.4)	-6.1 (-15.5, 4.4)	-4.8 (-17.3, 7.7)	-2.8 (-14.9, 9.2)	-3.4 (-12.8, 6.5)	-3.6 (-13.6, 7.0)	-3.8 (-13.7, 6.9)	**Naproxen**					
-4.2 (-22.8, 15.3)	0.9 (-18.8, 21.2)	-10.2 (-31.8, 12.2)	-7.3 (-32.2, 17.8)	-9.1 (-29.5, 11.8)	-6.0 (-25.1, 13.9)	-4.8 (-25.5, 16.1)	-2.8 (-23.4, 17.8)	-3.4 (-22.4, 16.2)	-3.6 (-23.0, 16.12)	-3.8 (-23.1, 16.2)	0.0 (-16.8, 16.8)	**Enteric coated Naproxen**				
-13.4 (-26.5, 0.8)	-8.4 (-22.9, 7.1)	-19.4 (-36.8, -1.2)	-16.5 (-37.4, 4.5)	-18.3 (-33.9, -2.3)	-15.3 (-29.0, -0.2)	-14.0 (-30.3, 2.3)	-12.0 (-27.7, 3.6)	-12.6 (-26.6, 1.8)	-12.9 (-27.4, 2.2)	-13.1 (-27.6, 2.4)	-9.2 (-21.4, 3.0)	-9.3 (-30.0, 11.3)	**Etoricoxib**			
-16.5 (-34.5, 1.6)	-11.5 (-31.6, 8.7)	-22.5 (-44.2, -0.6)	-19.6 (-44.9, 5.5)	-21.4 (-42.6, -0.5)	-18.2 (-38.1, 1.7)	-17.0 (-38.7, 4.2)	-15.1 (-36.5, 5.7)	-15.6 (-36.1, 4.6)	-15.8 (-36.8, 4.8)	-16.0 (-36.5, 4.4)	-12.2 (-32.8, 7.6)	-12.2 (-39.0, 13.7)	-2.9 (-26.0, 19.2)	**Tolfenamic acid**		
-4.1 (-17.7, 10.4)	0.9 (-14.2, 16.8)	-10.2 (-27.8, 8.4)	-7.3 (-28.4, 14.4)	-9.1 (-25.1, 7.5)	-6.0 (-20.6, 9.7)	-4.8 (-20.4, 11.0)	-2.7 (-15.8, 10.0)	-3.3 (-18.2, 11.9)	-3.5 (-18.9, 12.3)	-3.8 (-18.8, 12.2)	0.1 (-14.5, 14.7)	0.1 (-22.2, 22.2)	9.3 (-8.1, 26.7)	12.3 (-10.2, 35.4)	**Meloxicam 15mg**	
15.5 (8.2, 23.0)	20.2 (10.8, 30.5)	9.2 (-4.1, 23.1)	12.1 (-5.3, 29.5)	10.2 (-0.2, 21.1)	13.3 (4.9, 22.9)	14.6 (2.8, 26.4)	16.6 (6.1, 26.9)	16.0 (7.7, 24.7)	15.8 (6.4, 25.6)	15.6 (5.6, 26.3)	19.4 (11.5, 27.1)	19.3 (0.9, 37.8)	28.6 (16.4, 40.7)	31.6 (12.4, 51.4)	19.3 (6.5, 32.1)	**Placebo**

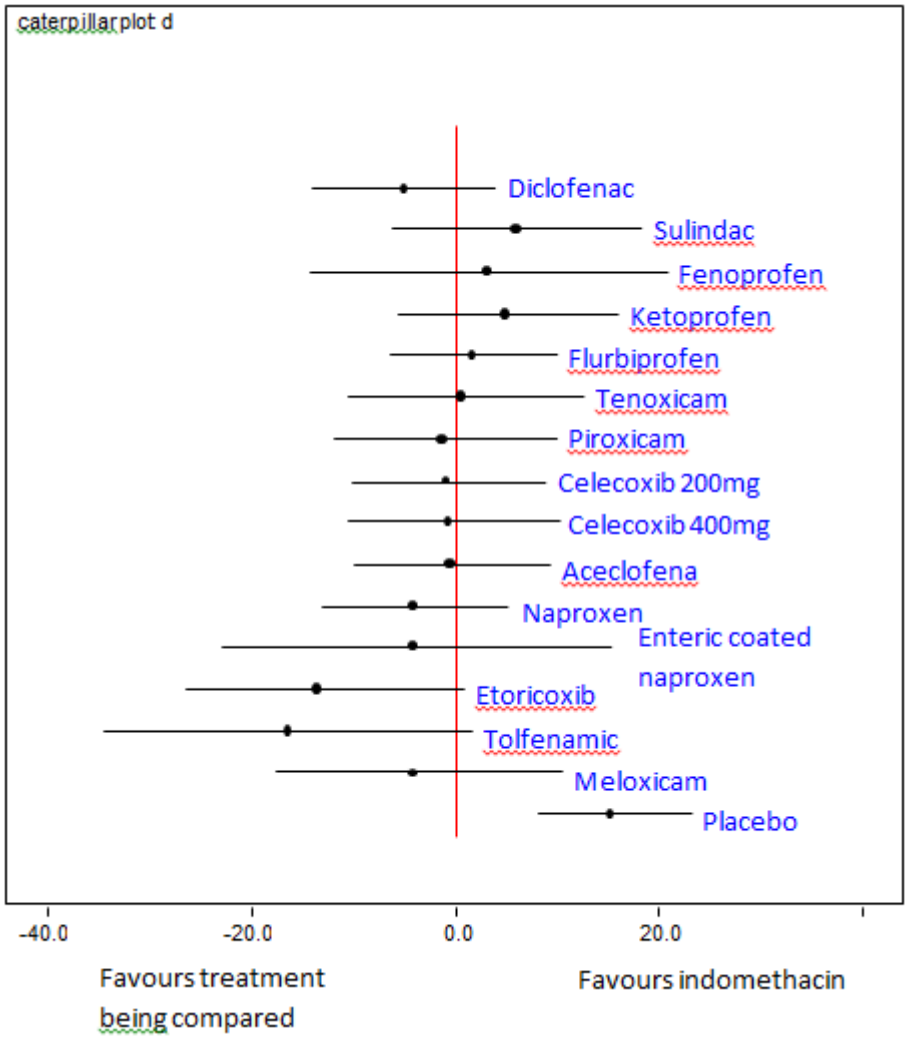
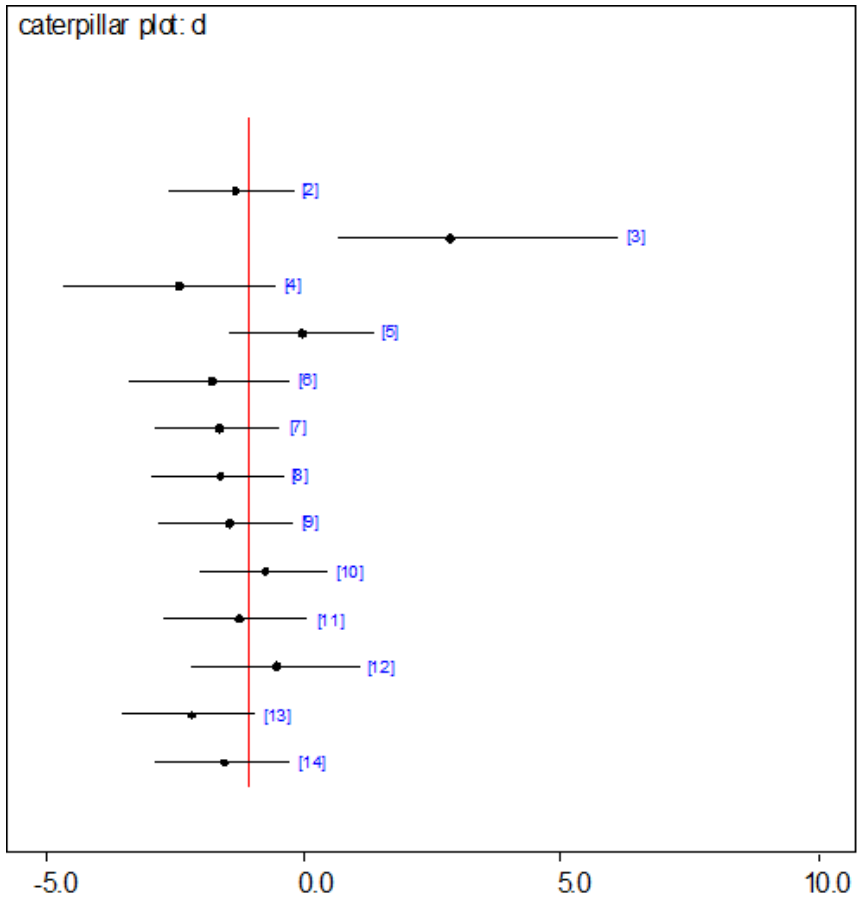


Figure 112 Pain NMA results

Meta-analysis – discontinuation due to adverse events

Table 63 Model fit

Model	Number of data points	Residual Deviance over all studies	DIC
Random Effects	43	58.4	197.55
Fixed Effects	43	63.24	200.86



All ORs compared to indomethacin

2: Diclofenac

3: Sulindac

4: Ketoprofen

5: Flurbiprofen

6: Tenoxicam

7: Pirpxicam

8: Celecoxib 200mg

9: Celecoxib 400mg

10: Acelofenac

11: Naproxen

12: Etorcoxib

13: Placebo

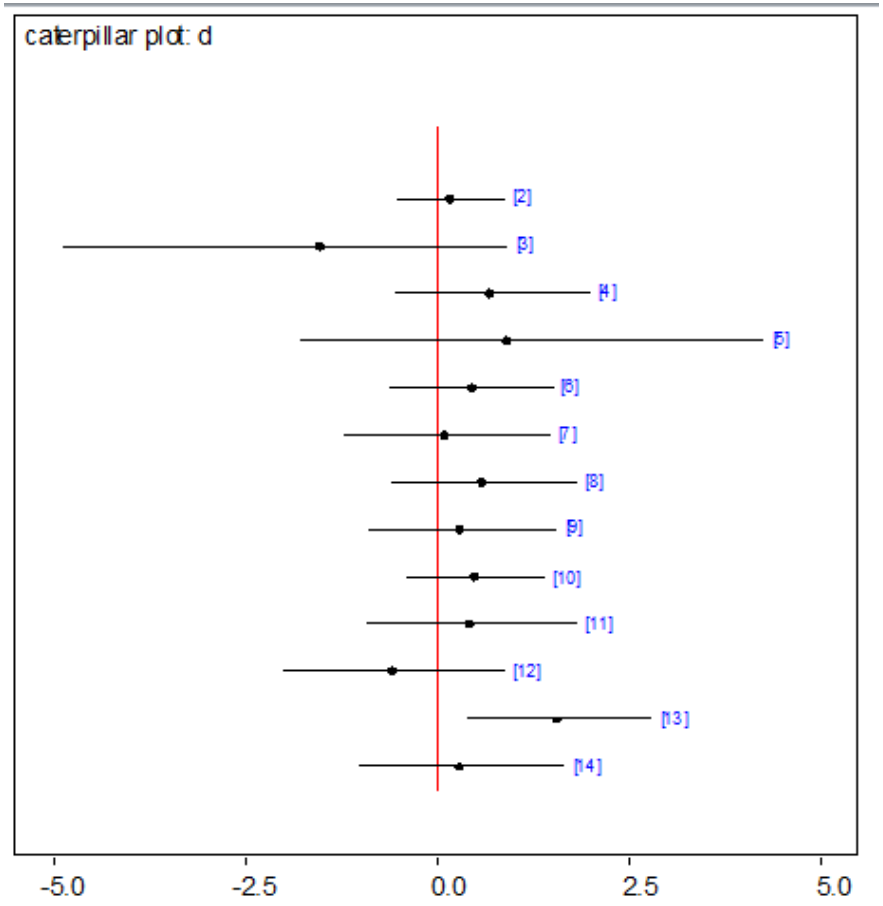
14: Meloxicam

Figure 113 Random effects model: mean difference (95% credible interval) – an odds ratio greater than 1 indicates higher discontinuations

Meta-analysis – discontinuation due to lack of efficacy

Table 64 Model fit

Model	Number of data points	Residual Deviance over all studies	DIC
Random Effects	33	28.52	171.89
Fixed Effects	33	28.11	171.07



All ORs compared to indomethacin

2: Diclofenac

3: Sulindac

4: Ketoprofen

5: Flurbiprofen

6: Tenoxicam

7: Pirpxicam

8: Celecoxib 200mg

9: Celecoxib 400mg

10: Acelofenac

11: Naproxen

12: Etorcoxib

13: Placebo

14: Meloxicam

Figure 114 Fixed effects model: mean difference (95% credible interval) – an odds ratio greater than 1 indicates higher

discontinuations

Table 65 Continuous versus on-demand NSAID therapy

Outcome	No. of studies	Risk of bias	Inconsistency	Indirectness	Imprecision	Estimate (CI)	Overall quality
Global pain (VAS)	1[a]	Not serious	N/A	Not serious	Serious[1]	MD -3.00 (-9.02, 3.02)	Moderate
Radiographic progression (m-SASSS)	1[a]	Not serious	N/A	Not serious	Not serious	MD -1.10 (-1.68, -0.52)	High
Serious adverse events	1[a]	Not serious	N/A	Not serious	Serious[1]	RR 0.78 (0.41, 1.49)	Moderate
Depression	1[a]	Not serious	N/A	Not serious	Not serious	RR 3.91 (1.25, 12.19)	High

[a]Guellec 2014

[1]Non-significant result

G.2.2 Pharmacological management of peripheral spondyloarthritis

Review Question 21

- What is the comparative effectiveness of the following pharmacological interventions for the management of peripheral spondyloarthritis:
  - corticosteroids
  - non-steroidal anti-inflammatory drugs (NSAIDs)
  - standard disease-modifying anti-rheumatic drugs (DMARDs)?

GRADE profiles, DMARD vs DMARD

Table 66 Pain related outcomes

Quality assessment							No of patients		Effect	Quality
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	
Pain, 24 weeks (pain score via VAS, 100mm)										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Mean difference (95%CI) Ciclosporin vs sulfasalazine, -9.90 (-22.04 to 2.24) Ciclosporin vs symptomatic therapy, -14.7 (-27.85 to -1.55) Sulfasalazine vs symptomatic therapy, -4.80 (-14.96 to 5.36),	VERY LOW
Tender joint counts, 24 weeks										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin)	31	Mean difference (95%CI) Ciclosporin vs sulfasalazine, -1.90 (-	VERY LOW

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
							32 (sulfasalazine)		6.05 to 2.25) Ciclosporin vs symptomatic therapy, -4.10 (-8.54 to 0.34) Sulfasalazine vs symptomatic therapy, -2.20 (-5.92 to 1.52),	
Painful joints, 12 months										
Sparda ro 1995	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	10	13	Mean difference (95%CI), -2.00 (-4.94 to 0.94)	VERY LOW

[1]Open label, allocation concealment unclear
[2]Differences not statistically significant

Table 67 Swollen joints

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
Swollen joint counts, 24 weeks										
Salvar ani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasala	31	Mean difference (95%CI) Ciclosporin vs sulfasalazine, 0.40 (-3.57 to 2.77) Ciclosporin vs symptomatic	VERY LOW

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
							zine)		therapy,3.00 (-6.12 to 0.12) Sulfasalazine vs symptomatic therapy, --2.60 (-5.39 to 0.19),	
Swollen joints, 12 months										
Spardaro 1995	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	10	13	Mean difference (95%CI), -0.90 (-2.92 to 1.12)	VERY LOW

[1]Open label, allocation concealment unclear
[2]Differences not statistically significant

Table 68 Global assessment outcomes

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
Patient global disease assessment, 24 weeks										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Decrease by ≥1 point ciclosporin 61% vs symptomatic therapy 33%	VERY LOW
Patient assessment of disease, 12 months (mm)										

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
Sparda ro 1995	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	10	13	Mean difference (95%CI), 7.30 (-14.82 to 29.42)	VERY LOW
Physician global disease assessment, 24 weeks										
Salvar ani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Decrease by ≥1 point ciclosporin 66% vs symptomatic therapy 32% Decrease by ≥2 points ciclosporin 24% vs symptomatic therapy 0%; ciclosporin 24% vs sulfasalazine 3%	VERY LOW
Physician assessment of disease, 12 months (mm)										
Sparda ro 1995	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	10	13	Mean difference (95%CI) -14.80 (-27.20 to -2.40)	VERY LOW

[1]Open label, allocation concealment unclear
[2]Differences not statistically significant

Table 69 CRP

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
CRP, 24 weeks (mg/dl)										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Mean difference (95%CI) Ciclosporin vs sulfasalazine, -1.90 (-6.05 to 2.25) Ciclosporin vs symptomatic therapy, -4.10 (-8.54 to 0.34) Sulfasalazine vs symptomatic therapy, -2.20 (-5.92 to 1.52),	VERY LOW
CRP, 12 months (mg/dl)										
Spardaro 1995	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	17	18	Mean difference (95%CI) 4.20 (-11.87 to 20.27)	VERY LOW

[1]Open label, allocation concealment unclear
[2]Differences not statistically significant

Table 70 ACR criteria

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
ACR20 response rate, 24 weeks										
Salvarani	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclospori	31	Difference 6.9% (ciclosporin vs	VERY LOW

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
2001							n) 32 (sulfasalazine)		sulfasalazine), 12.1% (ciclosporin vs symptomatic therapy), 5.2% (sulfasalazine vs symptomatic therapy)	
ACR50 response rate, 24 weeks										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Difference 15.2% (ciclosporin vs sulfasalazine), 24.5% (ciclosporin vs symptomatic therapy), 9.3% (sulfasalazine vs symptomatic therapy)	VERY LOW
ACR70 response rate, 24 weeks										
Salvarani 2001	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	36 (ciclosporin) 32 (sulfasalazine)	31	Difference 13.8% (ciclosporin vs sulfasalazine and vs symptomatic therapy), 0% (sulfasalazine vs symptomatic therapy)	VERY LOW

[1]Open label, allocation concealment unclear
[2]Lack of appropriate measures of uncertainty

GRADE profiles, NSAID vs NSAID

Table 71 Pain related outcomes

Quality assessment							No of patients		Effect	
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect	Quality
Pain scores, 17 weeks (scale not reported)										
Juvakoski & Lassus, 1982	Reactive arthritis	Serious[1]	N/A	None	Serious[2]	None	50	50 (crossover)	No significant difference between the groups	LOW

[1]Allocation concealment unclear
[2]Differences not statistically significant

G.2.3 Switching or augmenting pharmacological interventions for spondyloarthritis

Review Question 23

- When a first-line treatment has failed, what is the effectiveness of the following for managing spondyloarthritis:
  - switching to a different pharmacological intervention?
  - augmenting with a second pharmacological intervention?

Table 72 GRADE profiles, ciclosporin and methotrexate vs placebo and methotrexate

Quality assessment							No of patients		Effect		Quality
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect		
Patient global pain (via VAS, cm), 12months											
Fraser 2005	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	38	34		Mean difference, -1.00 (95%CI -3.97 to 1.97)	VERY LOW
Tender joint counts, 12 months											
Fraser 2005	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	38	34		Mean difference, 4.40 (95%CI -3.58 to 12.38)	VERY LOW
Swollen joint counts, 12 months											
Fraser 2005	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	38	34		Mean difference, -1.20 (95%CI -	VERY LOW

Quality assessment							No of patients		Effect		
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect		Quality
										3.90 to 1.50)	
Patient global assessment of disease activity (via VAS, cm), 12 months											
Fraser 2005	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	38	34		Mean difference, -0.80 (95%CI -2.07 to 0.47)	VERY LOW
HAQ score, 12 months											
Fraser 2005	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[2]	None	38	34		Mean difference, 0.00 (95%CI -0.26 to 0.26)	VERY LOW

[1]No details on randomisation, allocation concealment unclear
[2]Differences not statistically significant

Table 73 GRADE profiles, tight control in early psoriatic arthritis

Quality assessment							No of patients		Effect		
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect (95%CI)		Quality
ACR20, 48weeks											
Coates	Psoriatic	Very	N/A	Serious[2]	Serious[3]	None	101	105	OR 1.91	p=0.039	VERY

Quality assessment							No of patients		Effect		Quality
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect (95%CI)		
2015	arthritis	serious[1]							(1.03 to 3.55)		LOW
ACR50, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	Serious[2]	None	None	101	105	OR 2.36 (1.25 to 4.47)	p=0.0081	VERY LOW
ACR70, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	Serious[2]	None	None	101	105	OR 2.64 (1.32 to 5.26)	p=0.0058	VERY LOW
BASDAI MCID*, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	Serious[4]	Serious[3]	None	81	79	RR 1.26 (1.00 to 1.61)		VERY LOW
BASFI MCID*, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	Serious[4]	Serious[3]	None	81	80	RR 1.51 (1.10 to 2.09)		VERY LOW
HAQ MCID*, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[3]	None	91	90	RR 1.42 (1.05 to 1.92)		VERY LOW
ASAS20, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[3]	None	80	79	RR 1.47 (1.07 to 2.01)		VERY LOW

Quality assessment							No of patients		Effect		Quality
No of studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Comparator	Relative Effect (95%CI)		
ASA40, 48weeks											
Coates 2015	Psoriatic arthritis	Very serious[1]	N/A	None	Serious[3]	None	80	81	RR 1.50 (1.00 to 2.24)		VERY LOW

[1]Open-label, standard care not defined
[2]Measure not in clinical use, rheumatoid arthritis tool
[3]Using GRADE default MID interval for dichotomous outcomes of (0.8, 1.25)
[4]Measure for ankylosing spondylitis

G.2.4 Biological DMARDs for spondyloarthritis

Review questions 24, 25, and 26

- What is the effectiveness of systemic biological disease-modifying anti-rheumatic drugs for managing symptoms of enteropathic arthritis?
- What is the effectiveness of systemic biological disease-modifying anti-rheumatic drugs for managing symptoms of reactive arthritis?
- What is the effectiveness of systemic biological disease-modifying anti-rheumatic drugs for managing symptoms of undifferentiated spondyloarthritis, excluding non-radiographic ankylosing spondylitis?

Comparison of adalimumab vs placebo over a 12 week period in people with peripheral spondyloarthritis (excluding ankylosing spondylitis and psoriatic arthritis)

Table 74 GRADE tables

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Control	Relative (95% CI)	Absolute	Quality
Swollen joint count (Better indicated by lower values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	19	19	-	MD 2.1 lower (4.07 to 0.13 lower)	VERY LOW
Tender joint count (Better indicated by lower values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	19	19	-	MD 3.5 lower (8.57 lower to 1.57 higher)	VERY LOW
BASDAI (Better indicated by lower values)											
1 (Param	RCT	serious[1]	N/A	serious[2]	no serious imprecision	none	19	19	-	MD 1.5 lower	LOW

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Control	Relative (95% CI)	Absolute	Quality
arta 2013)										(2.85 to 0.15 lower)	
ESR (Better indicated by lower values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	no serious imprecision	none	19	19	-	MD 7.7 lower (14.71 to 0.69 lower)	LOW
CRP (Better indicated by lower values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	19	19	-	MD 9.7 lower (21.41 lower to 2.01 higher)	VERY LOW
QoL: HAQ-DI (Better indicated by lower values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	19	19	-	MD 0.1 lower (0.55 lower to 0.35 higher)	VERY LOW
QoL: HUI-3 (Better indicated by higher values)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	19	19	-	MD 0.04 higher (0.21 lower to 0.29	VERY LOW

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Control	Relative (95% CI)	Absolute	Quality
										higher)	
Adverse events (n people with AEs)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	10/19 (52.6%)	10/19 (52.6%)	OR 1 (0.28 to 3.57)	0 fewer per 1000 (from 289 fewer to 272 more)	VERY LOW
								52.6%		0 fewer per 1000 (from 289 fewer to 272 more)	
Adverse events: serious (n events)											
1 (Paramarta 2013)	RCT	serious[1]	N/A	serious[2]	serious imprecision[3]	none	1/19 (5.3%)	1/19 (5.3%)	OR 1 (0.06 to 17.25)	0 fewer per 1000 (from 49 fewer to 437 more)	VERY LOW
								5.3%		0 fewer per 1000 (from 50 fewer to 438 more)	

[1]Some risk of bias due to lack of detail in reporting of trial methodology (i.e. allocation methods and concealment)
[2]Study did not directly address any of the pre-specified review questions as it involved a mixed population of people with either reactive, enteropathic or undifferentiated spondyloarthropathy
[3]Not a statistically significant difference

G.2.5 Long-term antibiotics for reactive arthritis

Review Question 19

- What is the effectiveness of long-term (4 weeks or longer) treatment with antibiotics for first-line management of reactive arthritis compared with standard treatment?

GRADE profiles

Table 75 All interventions and eligible triggers of reactive arthritis

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	All interventions and triggers of ReA	Control	Absolute (95% CI)	Quality
Painful/tender joints/arthralgia										
Carter (2010), Hoogkamp-Kostanje (2000), Kvien (2004), Putschsky (2006), Sieper (1999), Toivanen (1993), Wakefield (1999), Whaley (1969)	RCTs	very serious[1]	very serious[2]	no serious indirectness[3]	serious imprecision[4]	none	214	192	SMD 0.2 lower (0.83 lower to 0.44 higher)	VERY LOW
Swollen joints										
Carter (2010), Kvien (2004), Putschsky (2006), Toivanen (1993), Yli-Kerttula (2000)	RCTs	very serious[1]	serious[5]	no serious indirectness[3]	serious imprecision[4]	none	172	152	SMD 0.02 higher (0.28 lower to 0.32 higher)	VERY LOW
Pain intensity										
Putschsky (2006)	RCTs	serio	N/A	no serious	serious	none	17	15	MD 1.4	

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	All interventions and triggers of ReA	Control	Absolute (95% CI)	Quality
		us[6]		indirectness[3]	imprecision[4]				higher (0.23 lower to 3.03 higher)	LOW
Pain at movement										
Toivanen (1993)	RCTs	serious[6]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	19	MD 0.39 lower (2.35 lower to 1.57 higher)	LOW
Morning stiffness (0-10 scale)										
Toivanen (1993)	RCTs	serious[6]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	19	MD 1.65 lower (3.74 lower to 0.44 higher)	LOW
Morning stiffness (mins)										
Putschsky (2006)	RCTs	serious[6]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	15	MD 16 higher (26.95 lower to 58.95 higher)	LOW
ESR (Erythrocyte Sedimentation Rate)										
Carter (2010), Putschsky (2006), Toivanen (1993), Whaley (1969), Yli-Kerttula (2000)	RCTs	very serious[1]	serious[5]	no serious indirectness[3]	serious imprecision[4]	none	102	92	SMD 0 higher (0.39 lower to 0.0.39 higher)	VERY LOW
CRP (C-reactive protein)										
Carter (2010), Kvien (2004), Putschsky (2006), Toivanen (1993)	RCTs	very serio	no serious inconsisten	no serious indirectnes	serious imprecisi	none	142	120	SMD 0.08 higher (0.19 lower to 0.34	LOW

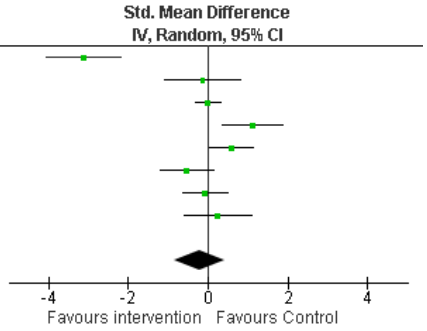
Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	All interventions and triggers of ReA	Control	Absolute (95% CI)	Quality
		us[8]	cy[7]	s[3]	on[4]				higher)	
Fatigue										
Putschsky (2006)	RCTs	serious[6]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	15	MD 40 higher (94.3 lower to 174.3 higher)	LOW

[1] Very serious risk of bias due to lack of reporting on randomisation, allocation concealment and blinding, selective outcome reporting and missing data handling. Some studies only presented data in graphs from which values had to be estimated.
[2] Very serious inconsistency (I^2 > 66%)
[3] Study/studies complied with review protocol requirements
[4] Not a statistically significant difference
[5] Serious inconsistency (33% < I^2 <= 66%)
[6] Some risk of bias due to lack of reporting on randomisation, allocation concealment and blinding, and missing data handling.
[7] No/Low inconsistency (I^2 <=33%)
[8] Very serious risk of bias due to lack of reporting on randomisation, allocation concealment and blinding, and missing data handling. One study only presented data in graphs from which values had to be estimated

Painful or tender joints/arthralgia (assorted scales)

	Intervention			Control				Std. Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Carter 2010	-5.3	2.2	27	1.7	2.2	15	11.2%	-3.12 [-4.07, -2.18]
Hoogkamp-Kostanje 2000	-2.9	2.2	7	-2.6	2.2	11	11.1%	-0.13 [-1.08, 0.82]
Kvien 2004	-1.79	3.97	81	-1.76	7.57	71	14.3%	-0.01 [-0.32, 0.31]
Putschky 2006	0.1	1.8	17	-2.2	2.2	15	12.3%	1.12 [0.37, 1.88]
Sieper 1999	-4.6	4.2	27	-7.3	4.9	28	13.4%	0.58 [0.04, 1.12]
Toivanen 1993	-1.76	2.95	17	-0.16	2.95	19	12.7%	-0.53 [-1.20, 0.14]
Wakefield 1999	-7.74	12.95	27	-6.68	11.68	22	13.3%	-0.08 [-0.65, 0.48]
Whaley 1969	-3.2	4.7	11	-4.39	4.51	11	11.8%	0.25 [-0.59, 1.09]
Total (95% CI)			**214**			**192**	**100.0%**	**-0.20 [-0.83, 0.44]**

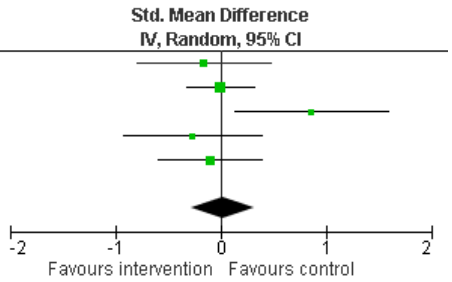
Heterogeneity: Tau² = 0.71; Chi² = 57.85, df = 7 (P < 0.00001); I² = 88%
Test for overall effect: Z = 0.61 (P = 0.54)



Swollen joints (assorted scales, lower values indicating favourable outcomes)

Study or Subgroup	Intervention Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Carter 2010	-3.8	4.9	27	-3	4.9	15	16.0%	-0.16 [-0.79, 0.47]
Kvien 2004	-1.44	2.27	81	-1.44	1.83	71	34.0%	0.00 [-0.32, 0.32]
Putschky 2006	-0.6	1.7	17	-2.1	1.7	15	13.0%	0.86 [0.13, 1.59]
Toivanen 1993	-1.24	4.15	17	-0.1	4.15	19	15.1%	-0.27 [-0.93, 0.39]
Yli-Kerttula 2000	-3.15	2.46	30	-2.9	2.55	32	21.9%	-0.10 [-0.60, 0.40]
Total (95% CI)			**172**			**152**	**100.0%**	**0.02 [-0.28, 0.32]**

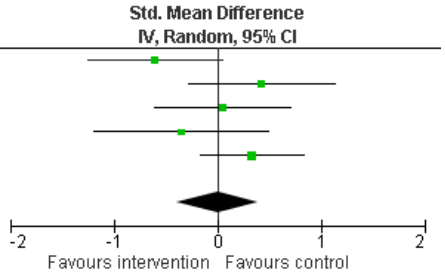
Heterogeneity: Tau² = 0.04; Chi² = 6.36, df = 4 (P = 0.17); I² = 37%
Test for overall effect: Z = 0.15 (P = 0.88)



ESR (mm/hr / mm at end of first hr/ no units)

Study or Subgroup	Intervention Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Carter 2010	-11.1	17.33	27	-0.5	17.33	15	20.4%	-0.60 [-1.25, 0.05]
Putschky 2006	-7	14	17	-14	18	15	18.5%	0.43 [-0.28, 1.13]
Toivanen 1993	-2.11	17.33	17	-2.99	17.33	19	20.1%	0.05 [-0.60, 0.70]
Whaley 1969	-6.95	5.26	11	-5.3	3.73	11	14.7%	-0.35 [-1.19, 0.50]
Yli-Kerttula 2000	-45	31.77	30	-56	32.8	32	26.3%	0.34 [-0.17, 0.84]
Total (95% CI)			**102**			**92**	**100.0%**	**0.00 [-0.39, 0.39]**

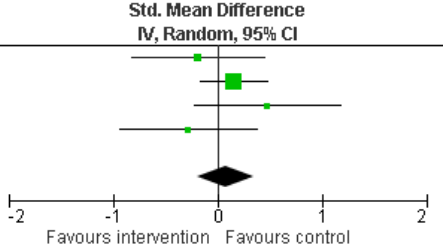
Heterogeneity: Tau² = 0.09; Chi² = 7.09, df = 4 (P = 0.13); I² = 44%
Test for overall effect: Z = 0.02 (P = 0.99)



CRP (hsCRP/CRP (mg/l))

Study or Subgroup	Intervention Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Carter 2010	-0.66	3	27	-0.08	3	15	16.6%	-0.19 [-0.82, 0.44]
Kvien 2004	-25	61.99	81	-35	64.5	71	54.3%	0.16 [-0.16, 0.48]
Putschky 2006	-0.8	2.3	17	-2.1	3	15	13.6%	0.48 [-0.23, 1.18]
Toivanen 1993	-1.3	15.6	17	3.16	15.6	19	15.5%	-0.28 [-0.94, 0.38]
Total (95% CI)			**142**			**120**	**100.0%**	**0.08 [-0.19, 0.34]**

Heterogeneity: Tau² = 0.01; Chi² = 3.30, df = 3 (P = 0.35); I² = 9%
Test for overall effect: Z = 0.55 (P = 0.58)



Adverse events (all)

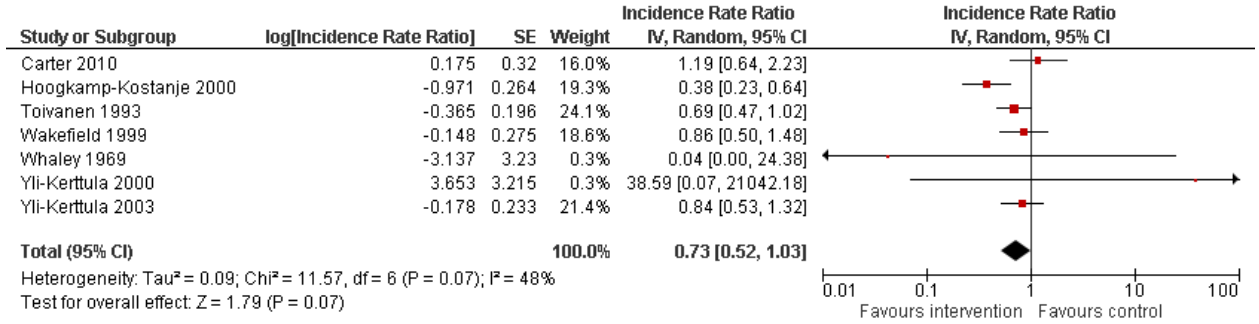


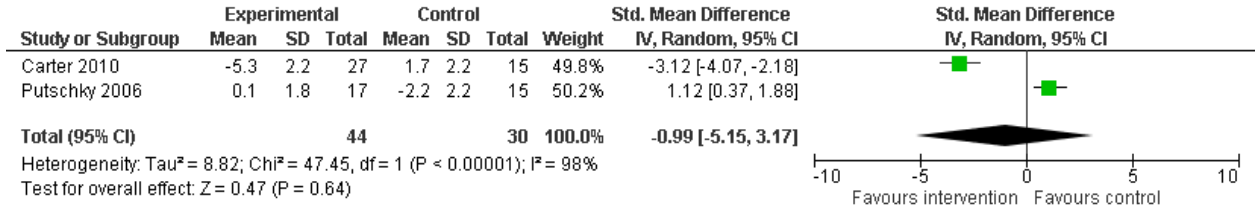
Table 76 Urogenital triggers only

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	UG triggers only	Control	Absolute (95% CI)	Quality
UG_painful/tender joints/arthralgia										
Carter (2010); Putschky (2006)	RCTs	serious[1]	very serious[2]	no serious indirectness[3]	serious imprecision[4]	none	62	30	SMD 0.99 lower (5.15 lower to 3.17 higher)	VERY LOW
UG_swollen joints										
Carter (2010); Putschky (2006)	RCTs	serious[1]	very serious[2]	no serious indirectness[3]	serious imprecision[4]	none	44	30	SMD 0.33 higher (0.67 lower to 1.33 higher)	VERY LOW
UG_ Pain intensity										
Putschky (2006)	RCTs	serious[1]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	15	MD 1.4 higher (0.23 lower to 3.03 higher)	LOW
UG_morning stiffness (mins)										
Putschky (2006)	RCTs	serious[1]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	15	MD 16 higher (26.95 lower to 58.95 higher)	LOW
UG_ESR										

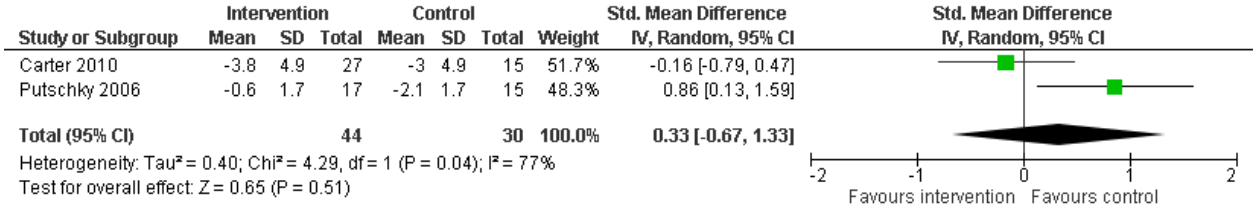
Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	UG triggers only	Control	Absolute (95% CI)	Quality
Carter (2010); Putschky (2006)	RCTs	serious[1]	very serious[2]	no serious indirectness[3]	serious imprecision[4]	none	44	30	SMD 0.1 lower (-1.10 lower to 0.91 higher)	VERY LOW
UG_CRP										
Carter (2010); Putschky (2006)	RCTs	serious[1]	serious[5]	no serious indirectness[3]	serious imprecision[4]	none	44	30	SMD 0.13 higher (0.53 lower to 0.78 higher)	VERY LOW
UG_Fatigue										
Putschky (2006)	RCTs	serious[1]	N/A	no serious indirectness[3]	serious imprecision[4]	none	17	15	MD 40 higher (94.3 lower to 174.3 higher)	LOW

[1] Some risk of bias due to lack of reporting on randomisation, allocation concealment and blinding, and missing data handling.
[2] Very serious inconsistency (I^2 > 66%)
[3] Study/studies complied with review protocol requirements
[4] Not a statistically significant difference
[5] Serious inconsistency (33% < I^2 <= 66%)

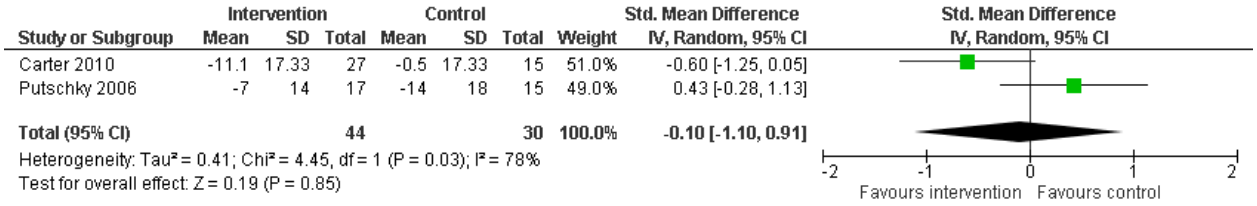
Painful or tender joints/arthralgia



Swollen joints



ESR (mm/hr / mm at end of first hr)



CRP (hsCRP/CRP (mg/l))

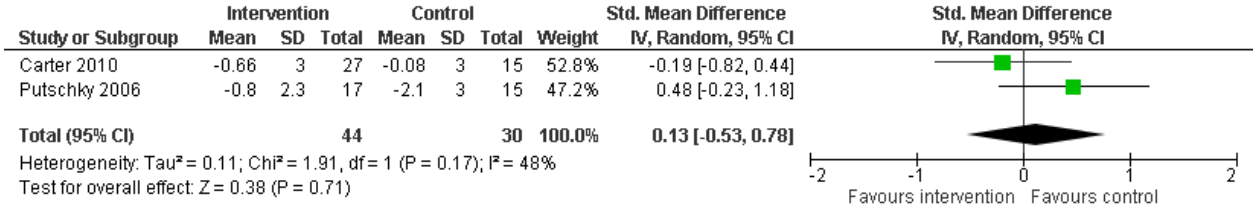


Table 77 Gastrointestinal triggers only

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	GI triggers only	Control	Absolute (95% CI)	Quality

GI_painful/tender joints/arthralgia										
Hoogkamp-Korstanje (2010), Sieper (1999)	RCTs	very serious[1]	very serious[2]	no serious indirectness[3]	serious imprecision[4]	none	21	36	SMD 0.53 higher (0.68 lower to 1.75 higher)	VERY LOW

[1] Serious risk of bias due to a number of issues with study reporting, including issues around missing data handling, potential selective outcome reporting, and need to estimate outcome values from graphs.
[2] Very serious inconsistency (I^2=77%)
[3] Both studies met with review protocol requirements.
[4] Not a statistically significant differnece

Painful or tender joints/arthralgia

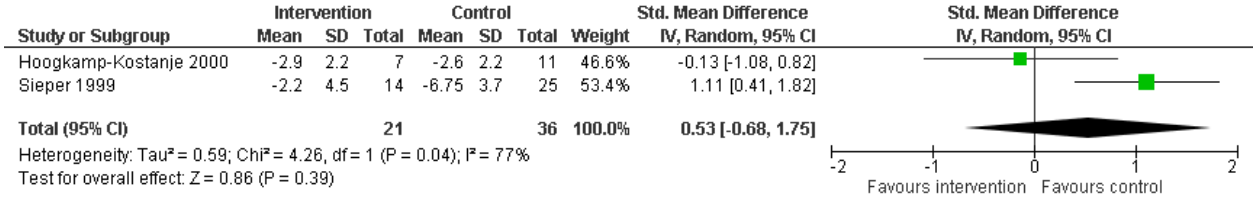


Table 78 Long-term secondary follow up

Quality assessment							No of patients		Effect		
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Long term secondary follow up	Control	Relative (95% CI)	Absolute	Quality
Long term_ESR											
Yli-Kerttula (2003)	observational studies	very serious[1]	N/A	serious[2]	serious imprecision[3]	none	26	27	-	MD 10.2 higher (4.39 lower to 24.79 higher)	VERY LOW
Long term_MRI findings											
Yli-Kerttul	observational studies	very seriou	N/A	serious[2]	serious imprecision	none	0/3 (0%)	3/3 (100	RR 0.14 (0.01 to	860 fewer per 1000 (from 990 fewer to	VERY LOW

Quality assessment							No of patients		Effect		Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Long term secondary follow up	Control	Relative (95% CI)	Absolute	
a (2003)		s[1]			[4]			%)	1.96)	960 more)	
								100 %		860 fewer per 1000 (from 990 fewer to 960 more)	
Long term_radiographic findings											
Yli-Kerttula (2003)	observational studies	very serious[1]	N/A	serious[2]	serious imprecision[4]	none	1/5 (20%)	3/6 (50 %)	RR 0.4 (0.06 to 2.75)	300 fewer per 1000 (from 470 fewer to 875 more)	VERY LOW
								50%		300 fewer per 1000 (from 470 fewer to 875 more)	
Long term_clinical findings of SpA											
Yli-Kerttula (2003)	observational studies	very serious[1]	N/A	serious[2]	serious imprecision[4]	none	2/26 (7.7%)	11/27 (40.7%)	RR 0.19 (0.05 to 0.77)	330 fewer per 1000 (from 94 fewer to 387 fewer)	VERY LOW
								40.7 %		330 fewer per 1000 (from 94 fewer to 387 fewer)	

[1] Original study lacked clarity regarding reporting of randomisation, blinding and allocation. This follow up study did not capture all of the original patient population.
[2] Study design does not entirely match protocol.
[3] Not a statistically significant difference
[4] Using GRADE default MID interval for dichotomous outcomes of (0.8, 1.25)

G.3 Non-pharmacological management

G.3.1 Manual therapies for spondyloarthritis

Review question 14

- What is the effectiveness of manual therapies compared with standard care for managing spondyloarthritis?

Table 79 GRADE tables

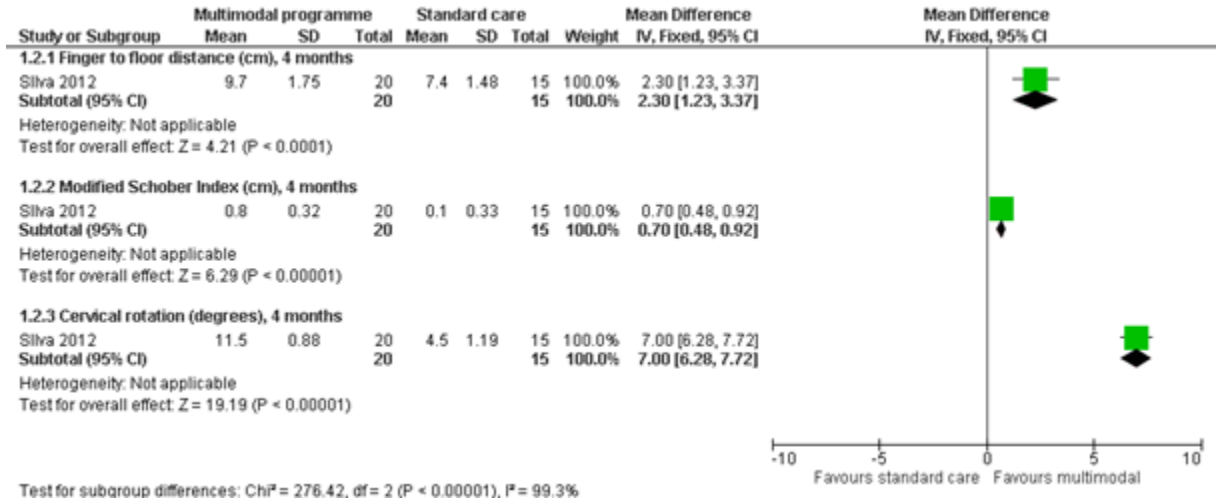
Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Individualised programme	Standard care	Absolute (95% CI)	
Composite measures: BASFI (follow-up 8 weeks; Better indicated by lower values)										
1 (Widberg 2009)	RCT	serious[1]	N/A	serious[2]	serious[3]	none	16	16	MD -0.3 (-1.63 to +1.03)	VERY LOW
Composite measures: BASDAI (follow-up 8 weeks; Better indicated by lower values)										
1 (Widberg 2009)	RCT	serious[1]	N/A	serious[2]	serious[3]	none	16	16	MD 0 (-1.27 to +1.27)	VERY LOW
Composite measures: BASMI (follow-up 8 weeks; Better indicated by lower values)										
1 (Widberg 2009)	RCT	serious[1]	N/A	serious[2]	not serious	none	16	16	MD -1.2 (-2.27 to -0.13)	LOW
Joint mobility - Finger to floor distance (cm) (follow-up 4 months; Better indicated by lower values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD 2.3 (1.23 to 3.37)	VERY LOW
Joint mobility, Modified Schober Index (cm) (follow-up 4 months; Better indicated by higher values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD 0.7 (0.48 to 0.92)	VERY LOW
Joint mobility, Cervical rotation (degrees) (follow-up 4 months; Better indicated by higher values)										

Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Individualised programme	Standard care	Absolute (95% CI)	
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD 7 (6.28 to 7.72)	VERY LOW
Composite measures: BASDAI (follow-up 4 months; Better indicated by lower values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD -1.4 (-1.62 to -1.18)	VERY LOW
Composite measures: HAQ-S (follow-up 4 months; Better indicated by lower values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD -0.6 (-0.7 to -0.5)	VERY LOW
QoL: SF36 - ECS (follow-up 4 months; Better indicated by higher values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	no serious	none	20	15	MD 3.6 (0.89 to 6.31)	VERY LOW
QoL: SF36 - PCS (follow-up 4 months; Better indicated by higher values)										
1 (Silva 2012)	CCT	very serious[4]	N/A	serious[2]	not serious	none	20	15	MD 15.5 (13.49 to 17.51)	VERY LOW

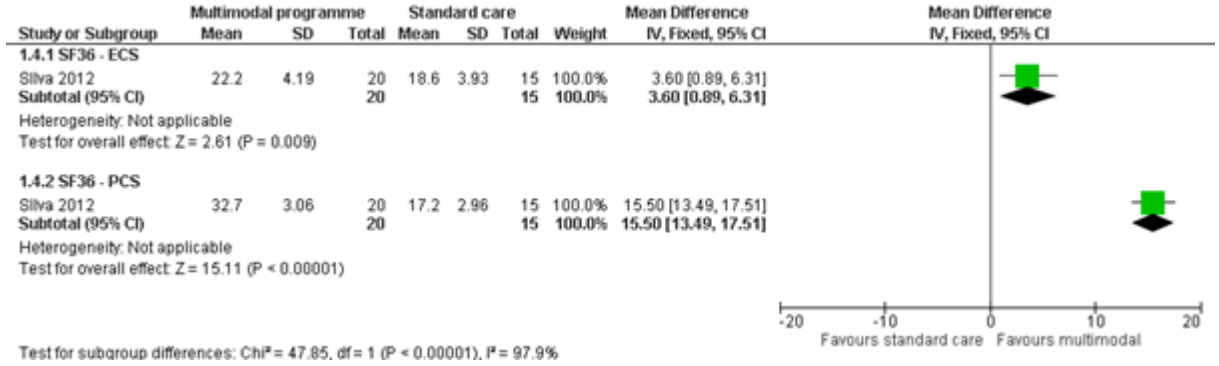
[1] Small study with no details provided of the blinding procedures for the outcome assessors
[2] Intervention comprised combination of exercise and manual therapy
[3] Not a significant difference
[4] Small, non-randomised controlled trial with baseline differences in age and cervical pain; unclear blinding procedures for outcome assessors

Forest plots for individualised programmes of manual therapy and exercise vs. control (no treatment or usual care)

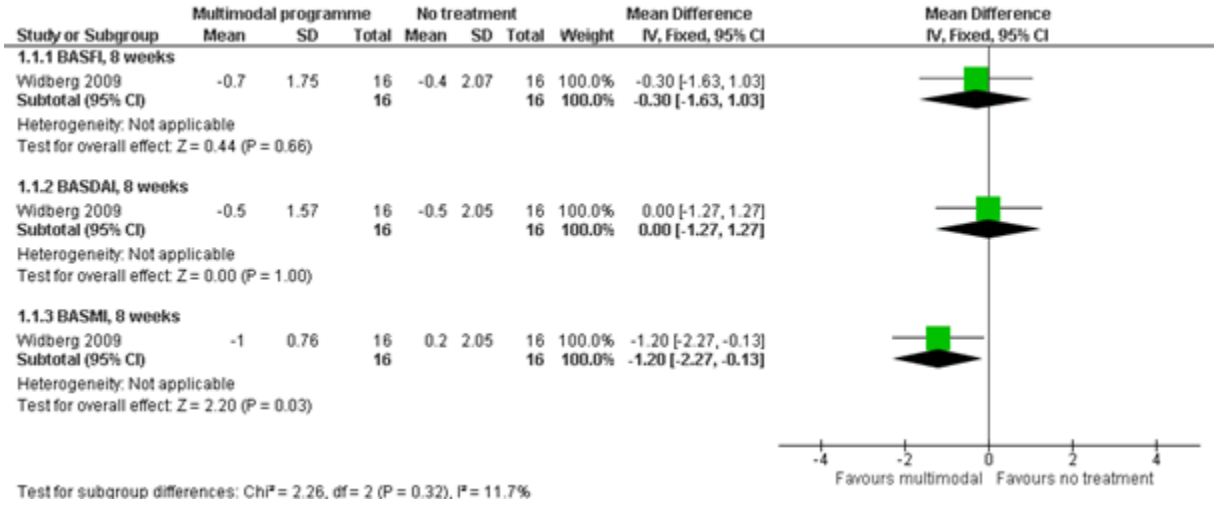
Joint mobility (data from CCT, Silva 2012)



Quality of life (data from CCT, Silva 2012)



Composite measures (data from RCT, Widberg 2009)



Composite measures (data from CCT, Silva 2012)

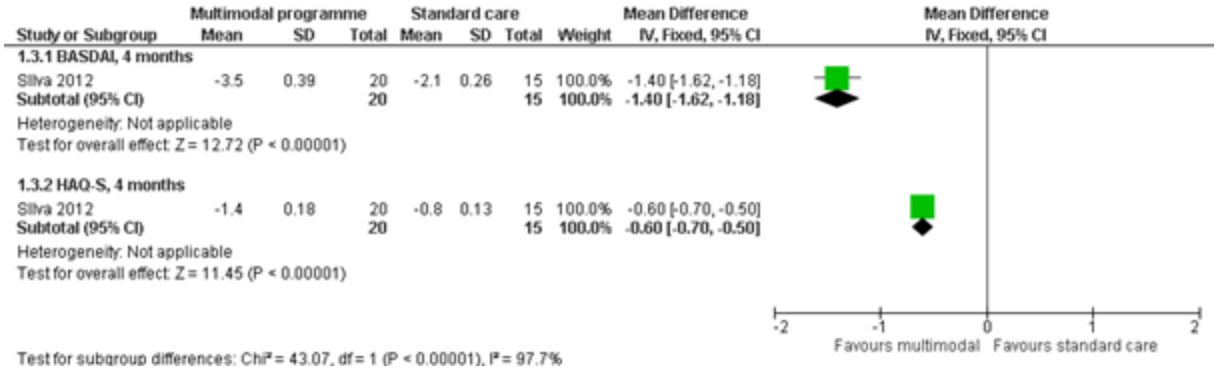


Table 80 GRADE profile for group and individualised multimodal inpatient programme including manual therapy vs. no treatment (data from Lubrano 2006 and 2007)

Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Inpatient programme	No treatment	Absolute (95% CI)	
Pain - Visual analogue scale (follow-up 3 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -25.5 (-28.18 to -22.82)	VERY LOW
Pain - Visual analogue scale (follow-up 6 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -17.8 (-20.14 to -15.46)	VERY LOW
Pain - Visual analogue scale (follow-up 12 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -10.3 (-12.49 to -8.11)	VERY LOW
Joint mobility - Modified Schober's test, cm (follow-up 3 weeks; Better indicated by higher values)										
2 (Lubrano 2006 and 2007)	observational	very serious[1]	N/A	serious[2]	not serious	none	71	71	MD 0.49 (0.29 to 0.69)	VERY LOW
Joint mobility - Modified Schober's test, cm (follow-up 6 weeks; Better indicated by higher values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD 0.4 (0.17 to 0.63)	VERY LOW
Joint mobility - Modified Schober's test, cm (follow-up 12 weeks; Better indicated by higher values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD 0.3 (0.07 to 0.53)	VERY LOW
Joint mobility - Tragus to wall distance, cm (follow-up 3 weeks; Better indicated by higher values)										
2 (Lubrano	observ	very	N/A	serious[2]	not serious	none	71	71	MD 4.09 (1.69 to	VERY

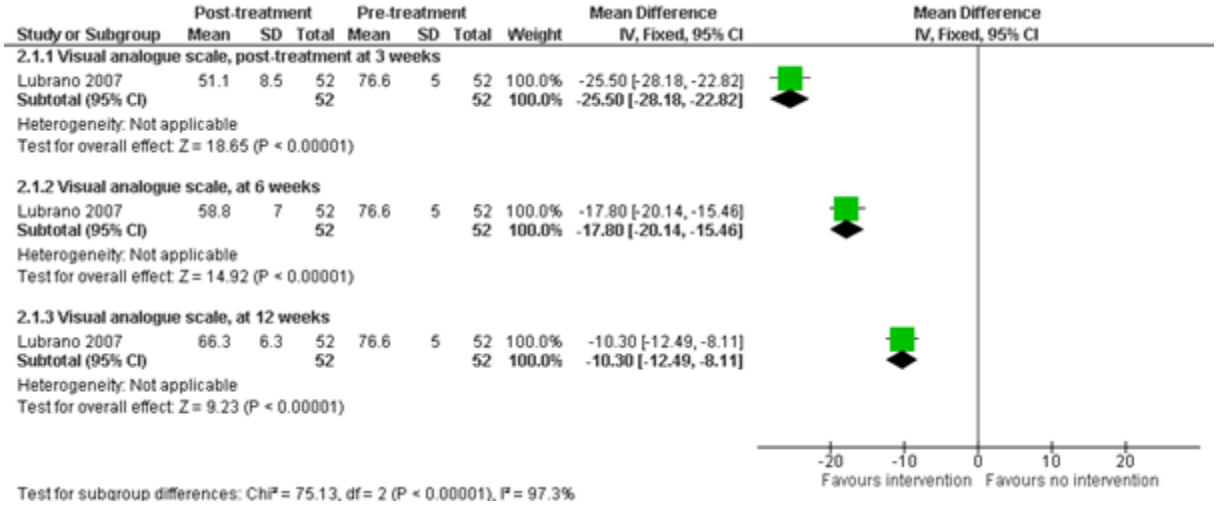
Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Inpatient programme	No treatment	Absolute (95% CI)	
2006 and 2007)	ational	serious[1]							6.49)	LOW
Joint mobility - Tragus to wall distance, cm (follow-up 6 weeks; Better indicated by higher values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD 4.9 (3.46 to 6.34)	VERY LOW
Joint mobility - Tragus to wall distance, cm (follow-up 12 weeks; Better indicated by higher values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD 3.3 (1.49 to 5.11)	VERY LOW
Quality of life - EQ-5D VAS, 0-100 (follow-up 3 weeks; Better indicated by higher values)										
1 (Lubrano 2006)	observ ational	very serious[1]	N/A	serious[2]	not serious	none	19	19	MD 6.6 (1.11 to 12.09)	VERY LOW
Composite measures (change from baseline) - BASFI (follow-up 3 weeks; Better indicated by lower values)										
2 (Lubrano 2006 and 2007)	observ ational	very serious[1]	N/A	serious[2]	not serious	none	71	71	MD -1.25 (-2.28 to -0.2)	VERY LOW
Composite measures (change from baseline) - BASFI (follow-up 6 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -1.34 (-1.64 to -1.04)	VERY LOW
Composite measures (change from baseline) - BASFI (follow-up 12 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -0.92 (-1.21 to -0.63)	VERY LOW
Composite measures (change from baseline) - BASDAI (follow-up 3 weeks; Better indicated by lower values)										
1 (Lubrano	observ	very	N/A	serious[2]	serious[3]	none	19	19	MD -0.71 (-1.49	VERY

Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Inpatient programme	No treatment	Absolute (95% CI)	
2006)	ational	serious[1]							lower to +0.07)	LOW
Composite measures (change from baseline) - Revised Leeds Disability Questionnaire (0-3) (follow-up 3 weeks; Better indicated by lower values)										
2 (Lubrano 2006 and 2007)	observ ational	very serious[1]	N/A	serious[2]	not serious	none	71	71	MD -0.38 (-0.60 to -0.17)	VERY LOW
Composite measures (change from baseline) - Revised Leeds Disability Questionnaire (0-3) (follow-up 6 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -0.40 (-0.57 to -0.23)	VERY LOW
Composite measures (change from baseline) - Revised Leeds Disability Questionnaire (0-3) (follow-up 12 weeks; Better indicated by lower values)										
1 (Lubrano 2007)	case series	very serious[1]	N/A	serious[2]	not serious	none	52	52	MD -0.30 (-0.49 to -0.11)	VERY LOW

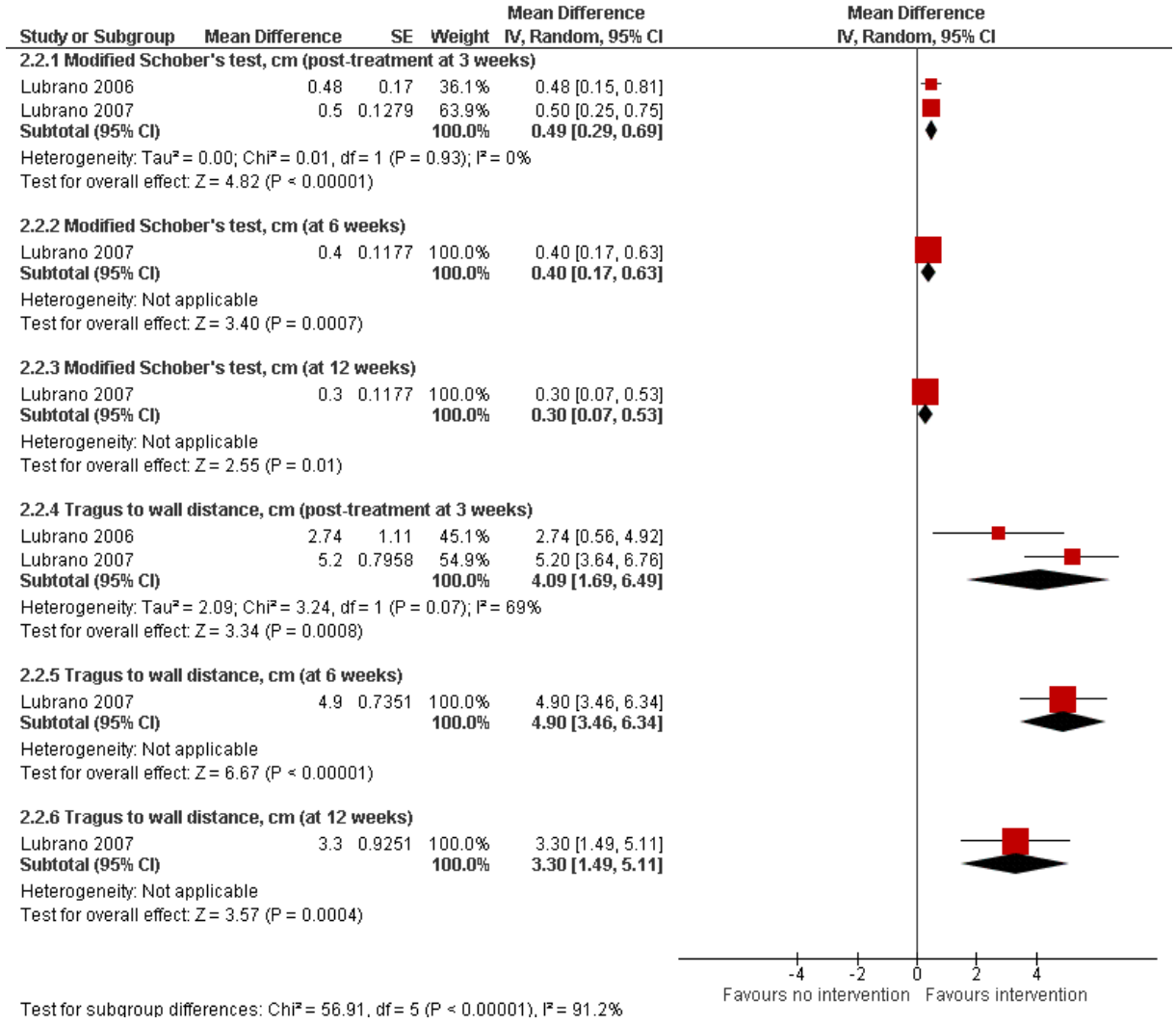
[1] Small prospective case series of patients with active ankylosing spondylitis; no details were provided of the methods of outcome assessments; no comparative group
[2] Intervention comprised combination of exercise and manual therapy
[3] Not a statistically significant difference

Forest plots for group and individualised multimodal inpatient programme including manual therapy vs. no treatment

Pain (data from prospective case series, Lubrano 2007)



Joint mobility (data from Lubrano 2006 and 2007)



Composite measures (data from Lubrano 2006 and 2007)

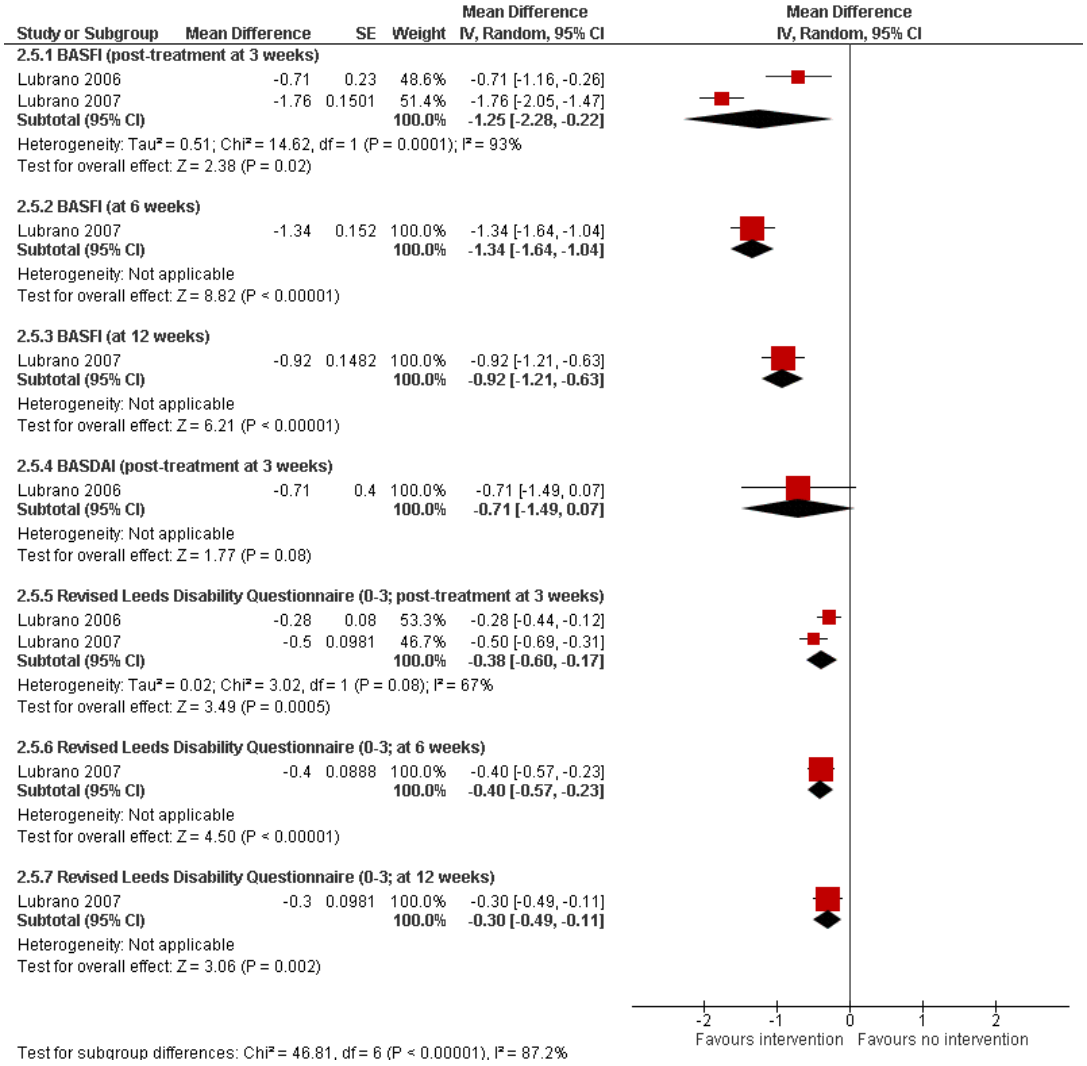


Table 81 GRADE profile for group and individualised multimodal inpatient programme including manual therapy vs. no treatment (data from a retrospective case series)

Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Group and individualised programme	No treatment	Absolute (95% CI)	
Joint mobility: Finger to floor distance (cm) (follow-up 2 weeks; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	not serious	none	49	49	MD -6 (-11.29 to -0.71)	VERY LOW
Composite measures: BASFI (follow-up 2 weeks; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	not serious	none	59	59	MD -0.8 (-1.5 to -0.1)	VERY LOW
Composite measures: BASFI (follow-up mean 9.3 months; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	serious[3]	none	48	48	MD -0.3 (-1.23 to +0.63)	VERY LOW
Composite measures: BASDAI (follow-up 2 weeks; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	not serious	none	59	59	MD -1.2 (-1.98 to -0.42)	VERY LOW
Composite measures: BASDAI (follow-up mean 9.3 months; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	serious[3]	none	48	48	MD -0.3 (-1.2 to +0.6)	VERY LOW
Composite measures: BASMI (follow-up 2 weeks; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	not serious	none	87	87	MD -0.9 (-1.61 to -0.19)	VERY LOW
Composite measures: BASMI (follow-up mean 9.3 months; Better indicated by lower values)										
(Eppeland 2013)	case series	very serious[1]	N/A	serious[2]	serious[3]	none	48	48	MD -0.6 (-1.62 to +0.42)	VERY LOW

[1] Retrospective case series including participants likely to benefit from a 2-week inpatient rehabilitation programme; unclear whether the physiotherapist administering the

Number of studies	Design	Quality assessment					Number of people		Effect	Quality
		Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Group and individualised programme	No treatment	Absolute (95% CI)	
intervention also assessed the outcomes; there were substantial missing data for all the outcomes (except BASMI) [2] Intervention comprised exercises (including water-based exercises) in a group setting and individual physiotherapy consisting of massage, stretching, mobilisation/articulation and advice on body posture enhancing exercises; delivered by a multidisciplinary team [3] Not a statistically significant difference										

Forest plots for group and individualised multimodal inpatient programme including manual therapy vs. no treatment (data from a retrospective case series)

Composite measures (data from retrospective case series, Eppeland 2013)

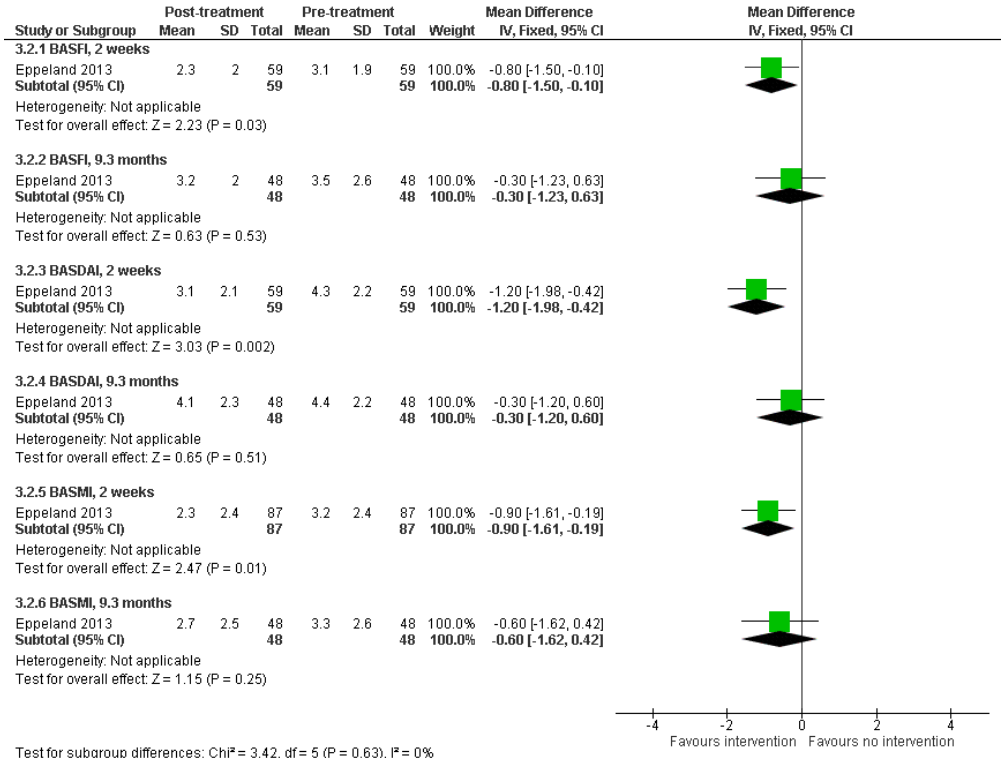


Table 82: GRADE table for retrospectively analysed prospective cohort study with self-reported physiotherapy

Quality assessment						No	Effect		
Studies	**Pop**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Total**	**Units**	**Effect**	**Quality**
Improvement of 20% in BASFI at 6 months (unadjusted)									
Escalas 2016	Axial SpA	serious[1]	not serious	serious[2]	serious[3]	689	RR (95% CI)	0.96 (0.77, 1.18)	VERY LOW
Improvement of 20% in BASFI at 6 months (propensity matched)									
Escalas 2016	Axial SpA	serious[1]	not serious	serious[2]	serious[3]	689	RR (95% CI)	1.15 (0.91, 1.45)	VERY LOW
Improvement of 20% in BASFI at 12 months (propensity matched)									
Escalas 2016	Axial SpA	serious[1]	not serious	serious[2]	serious[3]	671	RR (95% CI)	0.94 (0.80, 1.11)	VERY LOW
Improvement of 20% in BASFI at 24 months (propensity matched)									
Escalas 2016	Axial SpA	serious[1]	not serious	serious[2]	serious[3]	629	RR (95% CI)	1.09 (0.90, 1.33)	VERY LOW

[1] Observational study design

[2] Study evaluated physiotherapy and did not explicitly describe any manual therapy components

[3] Using GRADE default MID interval for dichotomous outcomes of (0.8, 1.25)

G.3.2 Exercise for spondyloarthritis

Review Question 15

What is the effectiveness of structured exercise compared with standard care for managing spondyloarthritis?

Table 83 GRADE profile for unsupervised structured home exercise vs standard care

Quality assessment						No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Unsupervised structured home exercise	Standard care	Absolute	
Pain (Better indicated by lower values)									
Kraag (1990), Rodriguez-Lozano (2013), Sweeney (2002),	randomised trials	serious [1]	serious [2]	not serious	serious [3]	478	481	MD 0.12 lower (0.63 lower to 0.39 higher)	Very low
BASDAI (Better indicated by lower values)									
Rodriguez-Lozano (2013), Sweeney (2002), Fang (2016), Hseih (2014), Jennings (2015)	randomised trials	serious [1]	not serious	not serious	serious [3]	521	513	MD 0.14 lower (0.38 lower to 0.1 higher)	Low
BASFI (Better indicated by lower values)									
Rodriguez-Lozano (2013), Sweeney (2002), Fang (2016), Hseih (2014), Jennings (2015)	multiple methodologies	serious [1]	not serious	not serious	not serious	521	513	MD 0.33 lower (0.53 to 0.12 lower)	Moderate
BASG (Better indicated by lower values)									
Kraag (1990), Hseih (2014),	multiple methodologies	not serious	not serious	not serious	serious [3]	84	90	MD 0.05 higher (0.77 lower to 0.88 higher)	Moderate
BASMI (Better indicated by lower values)									
Fang (2016), Jennings (2015)	multiple methodologie	serious [1]	not serious	not serious	serious [3]	56	48	MD 0.05 lower (0.9	Low

Quality assessment						No of patients		Effect	Quality
	s							lower to 0.79 higher)	
HAQ-S (Better indicated by lower values)									
Jennings (2015)	randomised trial	not serious	N/A	not serious	serious[3]	35	35	MD 0.08 lower (0.36 lower to 0.2 higher)	Moderate
ASQoL (Better indicated by lower values)									
Rodriguez-Lozano (2013)	randomised trial	not serious	N/A	not serious	not serious	381	375	MD 0.75 lower (1.18 to 0.32 lower)	High
Finger-floor distance (Better indicated by lower values)									
Kraag (1990)	randomised trial	not serious	N/A	not serious	not serious	22	26	MD 10 lower (14.14 to 5.16 lower)	High

[1] One study had high rate (20-25%) of loss to follow up and did not clearly report allocation concealment and method of randomisation.
[2] Moderate level of heterogeneity reported (33% =< I^2 <66%))
[3] Not a statistically significant difference

Table 84 GRADE profile for supervised structured exercise (outpatient) vs standard care

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised individual structured exercise (outpatient)	Standard care	Absolute (95% CI)	Quality
BASMI (Better indicated by lower values)										
Karapolat (2009) – 2 comparisons pooled	randomised trials	serious[1]	Serious[2]	no serious indirectness[3]	serious imprecision[4]	none	25	12	MD 0.41 lower (2.99 lower to 2.18 higher)	VERY LOW
Pain (Better indicated by lower values)										
Karapolat (2009) – 2	randomi	serio	Serious[2]	no serious	serious	none	26	12	MD 0.70 higher	

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised individual structured exercise (outpatient)	Standard care	Absolute (95% CI)	Quality
comparisons pooled	sed trials	us[1]		indirectness[3]	imprecision[4]				(22.77 lower to 24.18 higher)	VERY LOW
Finger-floor distance (Better indicated by lower values)										
Ince (2006), Karapolat (2009) – 2 comparisons pooled	randomised trials	very serious[5]	no serious inconsistency[6]	no serious indirectness[3]	serious imprecision[4]	none	41	27	MD 2.43 lower (9.17 lower to 4.31 higher)	VERY LOW

[1] Article has multiple errors and inconsistencies which may undermine the reliability of the results
[2] Serious inconsistency (33%<i^2<66%)
[3] No indirectness as population, intervention and outcome were as specified in the review protocol
[4] Not a statistically significant difference
[5] Allocation concealment unclear in one study. Multiple reporting errors with the other study.
[6] No serious inconsistency (i^2<33%)

Table 85 GRADE profile for supervised individual structured (inpatient) exercise vs standard care

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised individual structured exercise (inpatient)	Standard care	Absolute (95% CI)	Quality
BASDAI (Better indicated by lower values)										
Kjeken (2013)	randomised trials	no serious risk of bias[1]	N/A	no serious indirectness[2]	serious[3]	none	46	49	MD 5.8 lower (15.01 lower to 3.41 higher)	MODERATE
BASMI (Better indicated by lower values)										
Kjeken (2013)	randomised trials	no serious risk of bias[1]	N/A	no serious indirectness[2]	serious[3]	none	46	49	MD 0.4 lower (1.29 lower to 0.49 higher)	MODERATE

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised individual structured exercise (inpatient)	Standard care	Absolute (95% CI)	Quality
BASFI (Better indicated by lower values)										
Kjeken (2013)	randomised trials	no serious risk of bias[1]	N/A	no serious indirectness[2]	serious[3]	none	46	49	MD 3.2 higher (4.85 lower to 11.25 higher)	MODERATE

[1] No substantial risk of bias detected
[2] No indirectness as population, intervention and outcome were as specified in the review protocol
[3] Not a statistically significant difference

Table 86 GRADE profile for supervised structured group exercise vs home exercise

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised structured group exercise	Unsupervised structured home exercise	Absolute (95% CI)	Quality
BASFI (Better indicated by lower values)										
Analay (2003)	randomised trials	no serious risk of bias[1]	N/A	no serious indirectness[2]	serious imprecision[3]	none	23	22	MD 4.13 lower (14.17 lower to 5.91 higher)	MODERATE
Finger-floor distance (Better indicated by lower values)										
Analay (2003, Cagliyan (2007))	randomised trials	serious[5]	no serious inconsistency[5]	no serious indirectness[2]	serious imprecision[3]	none	46	45	MD 3.68 lower (10.01 lower to 2.65 higher)	LOW
Stiffness (Better indicated by lower values)										
Analay (2003)	randomised trials	no serious risk of bias[1]	N/A	no serious indirectness[2]	serious imprecision[3]	none	23	22	MD 11.5 lower (32.84 lower to 9.84 higher)	MODERA

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised structured group exercise	Unsupervised structured home exercise	Absolute (95% CI)	Quality
										TE
Pain (Better indicated by lower values)										
Analay (2003, Cagliyan (2007))	randomised trials	serious[4]	no serious inconsistency[5]	no serious indirectness[2]	serious imprecision[3]	none	46	45	MD 0.27 lower (1.44 lower to 0.91 higher)	LOW

[1] No substantial risk of bias, though few RCTs for this question were able to blind participants to treatment allocation
[2] No indirectness as population, intervention and outcome were as specified in the review protocol
[3] Not a statistically significant difference
[4] One study at high risk of bias due to multiple issues
[5] No evidence of inconsistency (i^2<33%)

Table 87 GRADE profile for supervised structured group exercise vs standard care

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised structured group exercise	Standard care	Absolute (95% CI)	Quality
BASDAI (Better indicated by lower values)										
Altan (2012); Maseiro 2014	RCTs	serious[1]	not serious	not serious[3]	serious[4]	none	51	46	MD 1.09 lower (1.92 to 0.27 lower)	LOW
BASMI (Better indicated by lower values)										
Altan (2012); Maseiro 2014	RCTs	serious[1]	serious[2]	not serious[3]	serious[4]	none	51	46	MD 0.37 lower (1.02 lower to 0.27 higher)	VERY LOW
BASFI (Better indicated by lower values)										

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Supervised structured group exercise	Standard care	Absolute (95% CI)	Quality
Altan (2012); Maseiro 2014	RCTs	serious[1]	not serious	not serious[3]	not serious	none	51	46	MD 0.78 lower (1.32 to 0.24 lower)	MODERATE
ASQoL (Better indicated by lower values)										
Altan (2012)	RCTs	not serious	N/A	not serious[3]	serious[4]	none	30	25	MD 0.5 higher (0.89 lower to 1.89 higher)	MODERATE

[1] Included studies at high risk of bias
[2] Serious inconsistency (i^2>33%)
[3] No indirectness as population, intervention and outcome were as specified in the review protocol
[4] Not a statistically significant difference

G.3.3 Hydrotherapy for spondyloarthritis

Review Question 16

- What is the effectiveness of hydrotherapy compared with standard care for managing spondyloarthritis?

GRADE tables for RTCs

Table 88: GRADE profile for active hydrotherapy vs standard care in people with axial symptoms (RCT)

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy	Standard care	Absolute (95% CI)	Quality
BASMI (Better indicated by lower values)										
Ciprian (2013)	RCTs	serious[1]	N/A	serious[2]	serious[3]	none	15	15	MD 0.04 lower (1.76 lower to 1.68 higher)	VERY LOW
BASDAI (Better indicated by lower values)										
Ciprian (2013)	RCTs	serious[1]	N/A	serious[2]	serious[3]	none	15	15	MD 0.2 lower (1.17 lower to 0.77 higher)	VERY LOW
Quality of Life HAQ (Better indicated by lower values)										
Ciprian (2013)	RCTs	serious[1]	N/A	serious[2]	serious[3]	none	15	15	MD 0.15 lower (0.55 lower to 0.25 higher)	VERY LOW
Pain (VAS) (Better indicated by lower values)										
Ciprian (2013)	RCTs	serious[1]	N/A	serious[2]	serious[3]	none	15	15	MD 6.26 lower (15.01 lower to 2.49 higher)	VERY LOW

[1] Information about treatment allocation method not available. knowledge of intervention not prevented during study.
[2] Active hydrotherapy delivered as part of a spa therapy package, and was preceded by mud pack application and passive thermal water immersion.
[3] Not a statistically significant difference

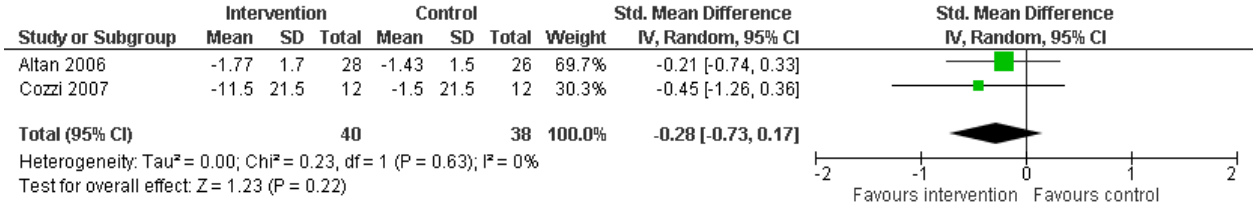
Table 89 GRADE profile for passive hydrotherapy vs standard care in people with axial symptoms (RCT)

Quality assessment	No of patients	Effect	Quality

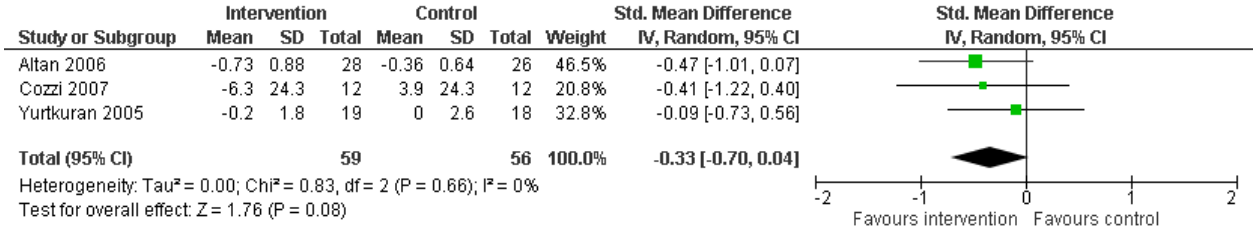
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Passive hydrotherapy	Standard care	Absolute (95% CI)	
BASDAI (Better indicated by lower values)										
Altan (2006), Cozzi (2007)	RCTs	very serious[1]	no serious inconsistency[2]	serious[3]	serious[4]	none	40	38	SMD 0.28 lower[9] (0.73 lower to 0.17 higher)	VERY LOW
Finger-floor distance (Better indicated by lower values)										
Yurtkuran (2005)	RCTs	very serious[5]	N/A	serious[6]	serious[4]	none	19	18	MD 0.4 lower (3.4 lower to 2.6 higher)	VERY LOW
BASFI/Dougados FI (Better indicated by lower values)										
Altan (2006), Cozzi (2007), Yurtkuran (2005)	RCTs	very serious[7]	no serious inconsistency[2]	serious[6]	serious[4]	none	59	56	SMD 0.33 lower[10] (0.7 lower to 0.04 higher)	VERY LOW
Pain (Better indicated by lower values)										
Altan (2006), Cozzi (2007), Yurtkuran (2005)	RCTs	very serious[7]	serious[8]	serious[6]	serious[4]	none	59	56	MD 4.17 lower (12.07 lower to 3.74 higher)	VERY LOW
QoL(NHP) (Better indicated by lower values)										
Altan (2006)	RCTs	very serious[5]	N/A	serious[6]	serious[4]	none	28	26	MD 3.10 lower (40.66 lower to 34.46 higher)	VERY LOW

[1] Both studies had omissions of detail required to assess adequacy of randomisation and allocation concealment. One study (Altan) additionally had some discrepancies in the reporting of results.
[2] No inconsistency detected (I^2<33%)
[3] Both studies looked at passive hydrotherapy (bathing)
[4] Not a statistically significant difference
[5] Study lacked clarity across a number of bias-assessment domains, with some reporting discrepancies
[6] Study of passive hydrotherapy
[7] All studies had risk of bias issues
[8] Serious inconsistency (I^2=44%)
[9] SMD equates to MD of 0.44 on a BASDAI 0-10 scale
[10] SMD equates to MD of 0.32 on a BASFI 0-10 scale

BASDAI



BASFI or Dougados functional index



Pain

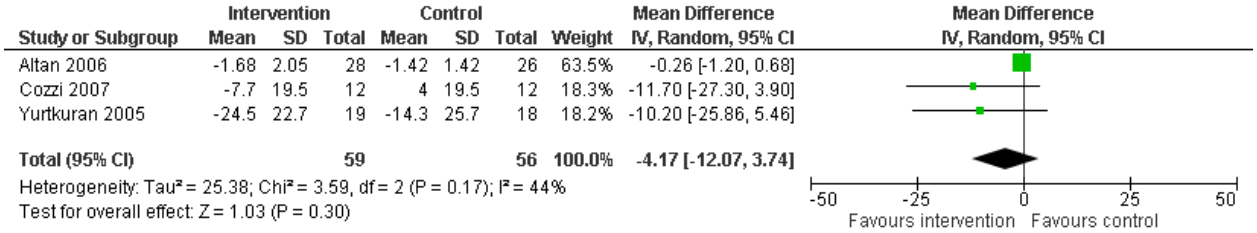


Table 90 GRADE profile for passive hydrotherapy with electrical current vs standard care in people with axial and peripheral symptoms (RCT)

Quality assessment	No of patients	Effect	Quality

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Passive hydrotherapy+electrical current	Standard care	Absolute (95% CI)	
BASMI (Better indicated by lower values)										
Gurcay (2008)	RCTs	serious[1]	N/A	very serious[2]	no serious imprecision	none	29	28	MD 0.56 lower (0.94 to 0.18 lower)	VERY LOW
BASFI (Better indicated by lower values)										
Gurcay (2008)	RCTs	serious[1]	N/A	very serious[2]	no serious imprecision	none	29	40	MD 1.36 lower (1.83 to 0.89 lower)	VERY LOW
BASDAI (Better indicated by lower values)										
Gurcay (2008)	RCTs	serious[1]	N/A	very serious[2]	no serious imprecision	none	29	28	MD 1.61 lower (2.18 to 1.04 lower)	VERY LOW
ASQoL (Better indicated by lower values)										
Gurcay (2008)	RCTs	serious[1]	N/A	very serious[2]	no serious imprecision	none	29	28	MD 2.07 lower (3.00 to 1.14 lower)	VERY LOW

[1] No detail on method used to generate allocation sequence
[3] Intervention involved passive hydrotherapy/bathing while an electrical current was administered.

GRADE tables and results for observational studies

Table 91 GRADE profile for active hydrotherapy in people with axial or axial and peripheral symptoms (observational studies

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy	Standard care	Absolute (95% CI)	Quality
BASFI (Better indicated by lower values)										
Robertson (2004)	Cohort	very serious[1]	N/A	not serious	serious[2]	None	17	n/a	Mean change 3.98 (-5.0 to 12.9)	VERY LOW

1. *Retrospective observational study, no comparison group, 34% of potential cases excluded for missing outcome data*
2. *Not a statistically significant difference*

Table 92 GRADE profiles for passive hydrotherapy in people with axial symptoms (observational studies)

Tichler 1995

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Passive hydrotherapy	Standard care	Absolute (95% CI)	Quality
Morning stiffness (Better indicated by lower values)										
Tishler (1995)	Non-randomised intervention	very serious[1]	N/A	not serious	not serious	All participants received intervention: no comparison group	14	n/a	Mean change -23 (SD 7)	LOW
Finger-floor distance (Better indicated by lower values)										
Tishler (1995)	Non-randomised intervention	very serious[1]	N/A	not serious	not serious	All participants received intervention: no comparison group	14	n/a	Mean change -14 (SD 4)	LOW

1. *Participants randomly selected but no detail on method. Selective outcome reporting: outcomes displayed as graphs and only largest results presented numerically*

Annegret 2013

Quality assessment							No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Passive hydrotherapy	Standard care	Absolute (95% CI)	
BASFI (Better indicated by lower values)										
Annegret (2013)	Control group of randomised trial	serious [1]	N/A	not serious	serious imprecision[2]	-	19	n/a	Mean change 0.22 (SD 1.01)	LOW
Self-assessed pain (NRS) (Better indicated by lower values)										
Annegret (2013)	Control group of randomised trial	serious [1]	N/A	not serious	serious imprecision[2]	-	19	n/a	Mean change 5.50 (SD 22.18)	LOW

1. *No serious risk of bias detected in study design, but no eligible comparison group available for our analysis*
2. *Not a statistically significant change*

Table 93 GRADE profiles for active hydrotherapy as part of a complex intervention in people with axial symptoms (observational studies)

Colina, 2009

Quality assessment							No of patients	Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy +physical therapy	Absolute (95% CI)	
BASFI (Better indicated by lower values)									
Colina (2009)	Non-randomised intervention	very serious [1]	N/A	serious[2]	not serious	-	30	Mean change 2.1 (no SD), p<0.05	VERY LOW
EQ-5D									
Colina (2009)	Non-randomised intervention	very serious [1]	N/A	serious[2]	not serious	-	30	Mean change 33 (no SD), p<0.05	VERY LOW

1. *Patients self-selected into intervention or control group, according to whether they found the proposed exercise programme acceptable*
2. *Hydrotherapy only one component of a complex exercise programme*

Aydemir 2010

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy +physical therapy	Standard care	Absolute (95% CI)	Quality
BASMI (Better indicated by lower values)										
Aydemir (2010)	Non-randomised intervention	very serious[1]	N/A	serious[2]	serious[3]		28	n/a	Mean change -1.06 (No SD), p=0.48	VERY LOW
BASDAI (Better indicated by lower values)										
Aydemir (2010)	Non-randomised intervention	very serious[1]	N/A	serious[2]	serious[3]		28	n/a	Mean change -0.4 (No SD), p>0.05	VERY LOW
BASFI (Better indicated by lower values)										
Aydemir (2010)	Non-randomised intervention	very serious[1]	N/A	serious[2]	serious[3]		28	n/a	Mean change 0.2 (no SD) p not reported	VERY LOW
SF-36 pain (Better indicated by lower values)										
Aydemir (2010)	Non-randomised intervention	very serious[1]	N/A	serious[2]	serious[3]		28	n/a	Mean change -0.89 (no SD), p=0.575	VERY LOW
SF-36 physical function (Better indicated by lower values)										
Ayedemir (2010)	Non-randomised intervention	very serious[1]	N/A	serious[2]	serious[3]		28	n/a	Mean change -1.85 (no SD), p=0.412	VERY LOW

1. *No control group, no detail on how participants were recruited.*
2. *Hydrotherapy only one component of a complex exercise programme*
3. *Not a statistically significant difference*

Eppeland 2013

Quality assessment							No of patients	Effect	Qualit

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy +physical therapy	Standard care	Absolute (95% CI)	y
BASMI (Better indicated by lower values)										
Eppeland (2013)	Retrospective case series	very serious[1]	N/A	serious[2]	not serious	n/a	87	n/a	Mean change -0.9 (SD 2.4) p<0.001	VERY LOW
BASDAI (Better indicated by lower values)										
Eppeland (2013)	Retrospective case series	very serious[3]	N/A	serious[2]	not serious	n/a	59	n/a	Mean change -0.8 (SD 2.2) p<0.001	VERY LOW
BASFI (Better indicated by lower values)										
Eppeland (2013)	Retrospective case series	very serious[3]	N/A	serious[2]	not serious	n/a	57	n/a	Mean change -0.8 (SD 2.0) p<0.001	VERY LOW
Finger-floor distance (Better indicated by lower values)										
Eppeland (2013)	Retrospective case series	very serious[3]	N/A	serious[2]	not serious	n/a	49	n/a	Median change -11 (IQR 25) p<0.001	VERY LOW

1. *No control group, retrospective study*
2. *Hydrotherapy only one component of a complex exercise programme*
3. *No control group, retrospective study, missing data*

Van Tubergen 2001

Quality assessment							No of patients		Effect	Quality
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy +physical therapy	Standard care	Absolute	
BASFI (Better indicated by lower values)										
Van Tubergen (2001)	Control group of randomised trial	serious[1]	N/A	serious[2]	serious[3]	n/a	39	n/a	Mean change -0.1(1.3)	LOW

Quality assessment							No of patients		Effect	
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Active hydrotherapy +physical therapy	Standard care	Absolute	Quality
BASDAI (Better indicated by lower values)										
Van Tubergen (2001)	Control group of randomised trial	serious[1]	N/A	serious[2]	serious[3]	n/a	39	n/a	Mean change 0.4 (1.5)	LOW
Morning stiffness (Better indicated by lower values)										
Van Tubergen (2001)	Control group of randomised trial	serious[1]	N/A	serious[2]	serious[4]	n/a	39	n/a	Median change 0 (IQR -1.3 to 1.4)	LOW
ASQoL										
Van Tubergen (2001)	Control group of randomised trial	serious[1]	N/A	serious[2]	serious[4]	n/a	39	n/a	Median change 0.0 (IQR -1.0 to 1.8)	LOW

1. *No serious risk of bias detected in study design, but no eligible comparison group available for our analysis*
2. *Hydrotherapy only one component of a complex exercise programme*
3. *Not a statistically significant difference*
4. *Inter-quartile range around median crosses 0*

Table 94 Summary of results from observational studies

Author	Year	Study type	Outcome	n	Baseline effect	Change from baseline or effect at follow up	Duration of follow up
Active hydrotherapy alone, in people with axial symptoms							
Robertson et al	2004	Retrospective cohort	BASFI	74		mean change: 3.98 (-5.0 to 12.9), p=0.4	3 to 5 years
Passive hydrotherapy alone, in people with axial symptoms							
Annegret et al	2013	Control group of an RCT	BASFI (mean(sd))	19	3.9(2.3)	change score (sd): 0.22(0.92)	9 months

Author	Year	Study type	Outcome	n	Baseline effect	Change from baseline or effect at follow up	Duration of follow up
Tishler et al	1995	Non-controlled intervention study	morning stiffness, mins (mean (sd))	14	38(7)	15(4)	2 weeks
			Finger-floor distance, cm (mean (sd))		27(3)	13(4)	4 weeks
Active hydrotherapy as part of a complex intervention in people with axial symptoms							
Aydemir et al	2010	Non-controlled, non-randomised intervention study	Pain, SF-36 domain (mean)	28	43.48	42.59 (p value of change: 0.575)	1 month
			Physical function, SF-36 domain (mean)		48.33	46.48 (p value of change: 0.412)	
			BASDAI (mean)		5.3	4.9 (p value of change: >0.05)	
			BASFI (mean)		4	4.2 (p value not reported)	
			BASMI (mean)		3.23	2.29 (p value of change: 0.48)	
Colina et al	2009	Intervention group from a non-randomised controlled study	BASFI (mean (sd))	30	6.9 (1.6)*	2.1 (no SD, p<0.05)	8 months from study start, 6 months from start of exercise intervention
			EQ-5D		16 (4.8)*	33 (no SD, p<0.05)	
Eppeland et al	2013	Retrospective case series	Finger-floor distance, cm (median (IQR))	49	11.0(25)	0(16)	2 weeks
			BASFI (mean (sd))	59	3.1(1.9)	2.3(2.0)	2 weeks
				48	3.2(2.)	3.5(2.6)	mean 9.3 months (sd=6.9)
			BASDAI (mean (sd))	57	4.3(2.2)	3.1(2.1)	2 weeks
				48	41(2.3)	4.4(2.2)	mean 9.3 months (sd=6.9)
			BASMI (mean (sd))	87	3.2(2.4)	2.3(3.4)	2 weeks
				48	3.3(2.6)	2.7(2.5)	mean 9.3 months (sd=6.9)
van Tubergen et	2001	Control group of an	pain, VAS (mean(sd))	39	4.8(2.8)	change -0.2(2.1)	40 weeks

Author	Year	Study type	Outcome	n	Baseline effect	Change from baseline or effect at follow up	Duration of follow up
al		RCT					
			morning stiffness, mins (median (IQR))		30 (10;60)	change 0 (-13;14)	
			ASQoL (median, IQR)		8.0 (3.0;11.8)	change 0.0 (-1.0; 1.8)	
			BASFI (mean (sd))		4.2 (2.1)	change -0.1 (1.3)	
			BASDAI (mean (sd))		4.5 (2.0)	change 0.4 (1.5)	
*Also included baseline values of the control group							

G.3.4 Acupuncture for spondyloarthritis

Review Question 17

- What is the effectiveness of acupuncture compared with sham acupuncture and standard care for managing spondyloarthritis?

GRADE tables

Table 95 Acupuncture vs sham acupuncture

Quality assessment							No of patients		Effect	
Study	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	Sham acupuncture	Absolute (95% CI)	Quality
Stiffness (better indicated by lower values)										
Emery (1986)	RCTs	very serious [1]	N/A	no serious indirectness[2]	serious imprecision[3]	none	5	5	MD 2.5 lower (16.63 lower to 11.63 higher)	VERY LOW
Pain (better indicated by lower values)										
Emery (1986)	RCTs	very serious [1]	N/A	no serious indirectness[2]	serious imprecision[3]	none	5	5	MD 0.2 lower (16.93 lower to 16.53 higher)	VERY LOW

[1] Inadequate reporting of baseline characteristics makes it difficult to assess whether randomisation was successful at eliminating selection bias, or demonstrating whether trial participants were representative of the patient population. Report stated that chest expansion and spinal movement were assessed, but these were not reported.
[2] No indirectness as population, intervention and outcome were as specified in the review protocol
[3] Not a statistically significant difference

Table 96 Acupuncture vs standard care

Quality assessment							No of patients		Effect	
Study	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	Standard care	Absolute (95% CI)	Quality
Finger-floor distance (better indicated by lower values)										
Jia (200	randomised trials	serious [1]	N/A	no serious indirectness[2]	no serious imprecision	none	30	30	MD 4.91 lower (9.32 to 0.5 lower)	MO

Quality assessment							No of patients		Effect	Quality
Study	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Acupuncture	Standard care	Absolute (95% CI)	
6)										DERATE
Swollen and painful peripheral joins (better indicated by lower values)										
Jia (2006)	randomised trials	serious [1]	N/A	no serious indirectness [2]	serious imprecision [3]	none	30	30	MD 0.03 lower (0.23 lower to 0.17 higher)	LOW
Morning stiffness (better indicated by lower values)										
Jia (2006)	randomised trials	serious [1]	N/A	no serious indirectness [2]	serious imprecision [3]	none	30	30	MD -1.40 lower (-16.47 lower to 13.67 higher)	LOW

[1] Limited reporting of baseline participant characteristics - hard to assess potential for selection bias. No details of blinding or allocation method reported. No details of missing outcome or baseline data reported
[2] No indirectness as population, intervention and outcome were as specified in the review protocol
[3] Not a statistically significant difference

G.3.5 Physical aids for spondyloarthritis

Review Question 18

- What is the effectiveness of physical aids (for example, braces) compared with standard care for managing spondyloarthritis?

No evidence was identified for this review

G.4 Surgical Interventions

Review Questions 34 and 35

- What factors predict clinical improvement after spinal surgery (including osteotomy and fusion) in people with axial inflammation?
- What factors predict clinical improvement after joint replacement surgery?

GRADE profiles for Q34: predictors of successful spinal surgery

No studies identified

GRADE profiles for Q35: predictors of successful joint replacement surgery

Table 97 Hip arthroplasty in people with ankylosing spondylitis: predictors of arthroplasty revision due to loosening of prosthetic components

Quality assessment							No of patients	Effect		
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	Quality
Age (per year)										
Lehtimaki (2001)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	none	n/a	76 operations in 54 patients	HR (95% CI)	0.98 (0.95 to 1.01)	VERY LOW
								P value	0.2	
Female sex										
Lehtimaki (2001)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	76 operations in 54 patients	HR (95% CI)	1.70 (0.66 to 4.40)	VERY LOW
								P value	0.3	
Weight (per kg)										
Lehtimaki	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	none	n/a	76 operations in 54 patients	HR (95% CI)	1.03 (0.99 to 1.07)	VERY LOW

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
(2001)								P value	0.2	
Steroids										
Lehtimaki (2001)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	76 operations in 54 patients	HR (95% CI)	1.23 (0.82 to 1.83)	VERY LOW
								P value	0.3	
Bleeding >median										
Lehtimaki (2001)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	76 operations in 54 patients	HR (95% CI)	0.85 (0.37 to 1.98)	VERY LOW
								P value	0.7	

[1]Risk of bias due to observational and retrospective nature of study, some limitations in quality of reporting, and potential confounders not controlled for in the analysis
[2]Outcome not directly relevant to review protocol
[3]95% confidence interval contains multiple qualitatively different possible clinical results

Table 98 Hip arthroplasty in people with ankylosing spondylitis: predictors of postoperative function (flexion)

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Female sex (diagnostic test accuracy)										
Zhang (2014)	Ankylosing spondylitis	very serious[1]	N/A	none[2]	none	n/a	167 hips in 100 patients	sensitivity	22.2% (14.0-30.4%)	VERY LOW
								specificity	86.8% (78.7-94.8%)	
Acetabular profusion (diagnostic test accuracy)										

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Zhang (2014)	Ankylosing spondylitis	very serious[1]	N/A	none[2]	none	n/a	167 hips in 100 patients	sensitivity	12.1% (5.7-18.6%)	LOW
								specificity	95.6% (90.7-100%)	
Ankylosis (diagnostic test accuracy)										
Zhang (2014)	Ankylosing spondylitis	very serious[1]	N/A	none[2]	none	n/a	167 hips in 100 patients	sensitivity	51.5% (41.7-61.4%)	LOW
								specificity	35.3% (23.9-46.7%)	
Preoperative C-reactive protein level										
Zhang (2014)	Ankylosing spondylitis	very serious[3]	N/A	serious[4]	none	n/a	167 hips in 100 patients	multivariate OR (95% CI)	0.981 (0.968 to 0.994)	VERY LOW
								P value	0.004	
Heterotopic ossification (diagnostic test accuracy)										
Zhang (2014)	Ankylosing spondylitis	very serious[1]	N/A	none[2]	none	n/a	167 hips in 100 patients	sensitivity	35.4% (25.9-44.8%)	LOW
								specificity	35.3% (23.9-46.7%)	
Heterotopic ossification										
Zhang (2014)	Ankylosing spondylitis	very serious[3]	N/A	serious[4]	none	n/a	167 hips in 100 patients	multivariate OR (95% CI)	0.237 (0.106 to 0.530)	VERY LOW
								P value	<0.001	
Use of a 32-mm femoral head (diagnostic test accuracy)										
Zhang (2014)	Ankylosing spondylitis	very serious[1]	N/A	none[2]	none	n/a	167 hips in 100 patients	sensitivity	74.8% (66.2-83.3%)	LOW

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
								specificity	75.0% (64.7-85.3%)	
Use of a 32-mm femoral head										
Zhang (2014)	Ankylosing spondylitis	very serious[3]	N/A	serious[4]	none	n/a	167 hips in 100 patients	multivariate OR (95% CI)	3.902 (1.817 to 8.377)	VERY LOW
								P value	<0.001	

[1] *Risk of bias due to observational and retrospective nature of study, some limitations in quality of reporting, and potential confounders not controlled for in the analysis*
[2] *Outcome directly relevant to review protocol*
[3] *Risk of bias due to observational and retrospective nature of study and some limitations in quality of reporting*
[4] *Outcome not directly relevant to review protocol*

Table 99 Hip arthroplasty in people with ankylosing spondylitis: predictors of blood loss

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Underweight (diagnostic test accuracy)										
Zhao (2014)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	none	n/a	236	sensitivity	43.4% (36.0-50.7%)	VERY LOW
								specificity	74.6% (63.9-85.4%)	

[1] *Some risk of bias due to observational and retrospective nature of study, and potential confounders not controlled for in the analysis*
[2] *Outcome (blood loss) is not a outcome directly specified in the review protocol*

Table 100 Hip arthroplasty in people with ankylosing spondylitis: predictors of poor healing of surgical incision

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Underweight (diagnostic test accuracy)										
Zhao (2014)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	none	n/a	236	sensitivity	42.9% (16.9-68.8%)	VERY LOW
								specificity	61.7% (55.3-68.1%)	

[1] *Some risk of bias due to observational and retrospective nature of study, and potential confounders not controlled for in the analysis*
[2] *Outcome (blood loss) is not a outcome directly specified in the review protocol*

Table 101 Hip arthroplasty in people with ankylosing spondylitis: predictors of heterotopic ossification

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Age										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	0.90 (0.79 to 1.03)	VERY LOW
Duration of symptoms										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	0.72 (0.39, 1.33)	VERY LOW
Female sex										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	not serious	n/a	47 operations in 24 patients	OR (95% CI)	11.79 (1.89, 73.58)	VERY LOW
Preoperative hip ankylosis										

Quality assessment							No of patients	Effect		Quality
Studies	Clinical population	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Total	Units	Effect	
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	not serious	n/a	47 operations in 24 patients	OR (95% CI)	67.00 (3.44, 1306.20)	VERY LOW
Heterotopic ossification in previous THA										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	not serious	n/a	24 operations in 24 patients	OR (95% CI)	37.86 (1.09, 713.10)	VERY LOW
Preoperative ESR										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	not serious	n/a	47 operations in 24 patients	OR (95% CI)	1.12 (1.03, 1.21)	VERY LOW
Preoperative CRP										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	not serious	n/a	47 operations in 24 patients	OR (95% CI)	1.27 (1.08, 1.48)	VERY LOW
Interval between THAs										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	1.06 (0.97, 1.18)	VERY LOW
Combined spinal epidural (versus general anaesthesia)										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	0.17 (0.02, 1.51)	VERY LOW
Hybrid implant (versus uncemented implant)										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	0.75 (0.10, 5.58)	VERY LOW
Cemented implant (versus uncemented implant)										
Thilak (2015)	Ankylosing spondylitis	very serious[1]	N/A	serious[2]	serious[3]	n/a	47 operations in 24 patients	OR (95% CI)	0.50 (0.06, 4.33)	VERY LOW

[1] *Risk of bias due to observational and retrospective nature of study, some limitations in quality of reporting, and potential confounders not controlled for in the analysis*
[2] *Outcome not directly relevant to review protocol*
[3] *Non-significant result*

G.5 Organisation of care and long-term monitoring

G.5.1 Transistion to adult services for young people with spondyloarthritis

Review question 13

- How should transition from specialist paediatric services to specialist adult rheumatology services be managed for young people between the ages of 16 and 18?

This review was not carried out (see the chapter in the full guideline for details

G.5.2 Monitoring of pharmacological interventions used in spondyloarthritis

Review Question 22

- What is the usefulness of direct access to specialist care, compared with initial primary care access followed by specialist rheumatological care, in the management of flare episodes?

No evidence was identified for this review

G.5.3 Care setting for management of flare episodes

Review Question 29

- What is the usefulness of direct access to specialist care, compared with initial primary care access followed by specialist rheumatological care, in the management of flare episodes?

No evidence was identified for this review

G.5.4 Care setting for long-term management

Review Question 30

- What is the effectiveness of specialist-led long-term management of spondyloarthritis compared with primary-care-led long-term management?

No evidence was identified for this review

G.5.5 Cross-speciality care

Review Question 31

- How should cross-speciality care for people with spondyloarthritis be organised?

No evidence was identified for this review

G.5.6 Complications of spondyloarthritis

Review Question 32

- What are the complications associated with spondyloarthritis?

For a summary of the results from this review, see appendix E, section 5.6 (table 165)

GRADE profiles

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Ischaemic heart disease							
4 (Chou, Brophy, Hung, Haroon)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Edson-Heredia)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Aortic valve insufficiency							
1 (Jantti)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Kaarela)	Reactive arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Stroke/cerebrovascular events							
4 (Brophy, Hung, Keller, Zoller)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Edson-Heredia)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Zoller)	Reactive arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Uveitis/iritis							
1 (Kaarela)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Egeberg)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
2 (Hart, Kaarela)	Reactive arthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
Fracture							
4 (Kang, Maillefert, Munoz-Ortego, Weinstein)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
Osteoporosis/osteopenia							

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
1 (Maillefer)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Inflammatory bowel disease							
1 (Mielants)	Undifferentiated spondyloarthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Edson-Heredia)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Depression							
1 (Shen)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Edson-Heredia)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Psoriasis/pustolosis palmoplantaris							
2 (Jantti, Theander)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Surgery							
1 (Kaarela)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
1 (Kaarela)	Reactive arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW

[1] Multiple possible sources of bias: inconsistent reporting on length of follow up, outcome not well defined, diagnostic criteria were not well defined at baseline
[2] Inconsistent reporting of results between studies
[3] Not possible to calculate meaningful measures of uncertainty

G.5.7 Complications of treatments for spondyloarthritis

Review Question 33

- What are the complications associated with treatments for spondyloarthritis?
- For a summary of the results from this review, see appendix E, section 5.7 (tables 166 and 167)

Table 102 GRADE: Biological DMARDs

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Uveitis							
5 (Baraliakos, Davis, Fouache, Heldman, van der Heijde)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Fouache)	Psoriatic arthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
Infections							
9 (Braun, Carmona, Davis, Deodhar, Gossec, Heldman, Park, Tong, van der Heijde)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
3 (Sieper, Song, Wallis)	Axial spondyloarthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
9 (Bianchi, Carmona, de Vlam, Gladman, Kavanaugh(a), Kavanaugh(b), Mease, Saad, Zisman)	Psoriatic arthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Carmona)	Undifferentiated spondyloarthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Tuberculosis							
5 (Heldman, Jung, Sengupta, van der Heijde)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
2 (Kavanaugh(a), Saad)	Psoriatic arthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
Hepatitis							
1 (Costa)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Depression							
1 (Davis)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Rash							
3 (Davis, Gossec, van der Heijde)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Gladman)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Malignancy							
7 (Deodhar, Gossec, Haynes, Heldman, Hellgren, van der Heijde, Haynes)	Ankylosing spondylitis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
7 (de Vlam, Gladman, Haynes, Hellgren, Kavanaugh(a), Kavanaugh(b), Mease)	Psoriatic arthritis	Very serious[1]	serious[2]	none	serious[3]	Risk of publication bias	VERY LOW
1 (Westhovens)	Undifferentiated spondyloarthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Cardiovascular adverse events							

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
1 (Kavanaugh(b))	Psoriatic arthritis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW
Demyelinating disease							
1 (van der Heijde)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[3]	Risk of publication bias	VERY LOW

[1] Poor reporting of study designs and outcomes; no control group to compare outcomes to.
[2] Inconsistent results between studies
[3] Not possible to calculate meaningful measures of uncertainty

Table 103 GRADE: standard DMARDs

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Infections							
1 (Wallis)	Axial spondyloarthritis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW
1 (Zisman)	Psoriatic arthritis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW

[1] Poor reporting of study designs and outcomes; no control group to compare outcomes to.
[2] Not possible to calculate meaningful measures of uncertainty

Table 104 GRADE: NSAIDs

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Cardiovascular adverse events							
1 (Kristensen)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW
1 (Kristensen)	Undifferentiated spondyloarthritis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW
Renal adverse events							
1 (Kristensen)	Ankylosing spondylitis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW
1 (Kristensen)	Undifferentiated spondyloarthritis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW

[1] Poor reporting of study designs and outcomes; no control group to compare outcomes to.
[2] Not possible to calculate meaningful measures of uncertainty

Table 105 GRADE: Corticosteroids

Quality assessment							
No of studies	**Clinical population**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Quality**
Infections							
1 (Wallis)	Axial spondyloarthritis	Very serious[1]	N/A	none	serious[2]	Risk of publication bias	VERY LOW

[1] Poor reporting of study designs and outcomes; no control group to compare outcomes to.
[2] Not possible to calculate meaningful measures of uncertainty

G.6 Information for people with spondyloarthritis

G.6.1 Information for people with spondyloarthritis

Review Question 27

- What information on treatment, long-term complications and self-management do young people and adults with spondyloarthritis find useful?

Table 106 GRADE profile

Quality assessment									
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Number of patients**	**Findings**	**Quality**
Summaries on latest research and medications									
(Cooksey 2012)	Survey	No serious concern[1]	N/A	No serious concerns[2]	N/A	none	155 (Internet)	95/155 (61)	LOW
							211 (Written material)	138/211 (65)	
							"Generally greater information on the cause of AS and the known treatments available. Plus what new treatments are coming onto the market or will be available in the near future." (Male, aged 46)		
Stories and experiences from other AS patients									
(Cooksey 2012)	Survey	No serious concern[1]	N/A	No serious concerns[2]	N/A	none	155 (Internet)	66 (43)	LOW
							211 (Written material)	90 (43)	
							"Swapping stories and self help, get AS sufferers to socialise with each other." (Male, aged 34)		
Opportunity to ask a doctor questions									
(Cooksey 2012)	Survey	No serious concern[1]	N/A	No serious concerns[2]	N/A	none	155 (Internet)	66 (43)	LOW
							211 (Written	74 (35)	

Quality assessment									
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Number of patients**	**Findings**	**Quality**
							material)		
							"The main issue is access to specialists. GPs often seem to know little about conditions such as AS and my consultants AS clinic (which is very good) only takes place every 3–6 months. There is a need to be able to discuss issues arising from flare-ups while they are occurring - not weeks or months later," (Male, aged 37).		
AS networking									
(Cooksey 2012)	Survey	Very serious concern[3]	N/A	No serious concerns[2]	N/A	none	155 (Internet)	39 (25)	VERY LOW
							211 (Written material)	56 (27)	
							"Regular emails to provide recent findings and other peoples experiences," (Male, aged 36).		
Diagnosis, medication, exercises and how to improve performance of daily activities									
(Giacomelli 2015)	Survey	Serious concern[4]	N/A	No serious concerns[2]	N/A	none	743	446 (60)	VERY LOW
Information on disease									
(Leung 2009)	Survey	No serious concern[1]	N/A	No serious concerns[2]	N/A	none	105	72 (68)	LOW
Advice on exercise									
(Leung 2009)	Survey	No serious concern[1]	N/A	No serious concerns[2]	N/A	none	105	77 (73)	LOW
Use of alternative medicine									
(Leung	Survey	No serious	N/A	No serious	N/A	none	105	35 (33)	LOW

Quality assessment									
No of studies	**Design**	**Risk of bias**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**Number of patients**	**Findings**	**Quality**
2009)		concern[1]		concerns[2]					
Managing pain (scale 0 – 24 : higher scores indicate greater need)									
(Dragoi 2013)	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 14.29 (6.69)	VERY LOW
Arthritis process (scale 0 – 28 : higher scores indicate greater need)									
(Dragoi 2013)	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 19.44 (6.89)	VERY LOW
Treatments (scale 0 – 28 : higher scores indicate greater need)									
(Dragoi 2013)	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 15.90 (7.59)	VERY LOW
Self-help measures (scale 0 – 24 : higher scores indicate greater need)									
(Dragoi 2013)	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 15.76 (5.90)	VERY LOW
Movement (0 - 20 : higher scores indicate greater need)									
(Dragoi 2013	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 9.79 (5.67)	VERY LOW
Feelings (scale 0 – 16 : higher scores indicate greater need)									
(Dragoi 2013	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 8.68 (4.73)	VERY LOW
Support systems (scale 0 – 16 : higher scores indicate greater need)									
(Dragoi 2013	Survey	Serious concern[3]	N/A	No serious concerns[2]	N/A	none	125	Mean (SD) 6.83 (4.40)	VERY LOW

[1] *Concerns over response rate (50%) and how representative the study population is but overall considered to be a low risk of bias*
[2] *Population and outcomes as specified in the review protocol*
[3] *Unclear methods and reporting*

G.6.2 Information and education for flare management in spondyloarthritis

Review Question 28

- What is the effectiveness of information and education in the management of flare episodes?

No evidence was identified for this review

1