



Diagnostic accuracy of interleukin-6 and procalcitonin in patients with periprosthetic joint infection: a systematic review and meta-analysis

Jung-Ro Yoon¹ · Se-Hyun Yang¹ · Young-Soo Shin¹

Received: 19 July 2017 / Accepted: 18 December 2017 / Published online: 2 January 2018
© SICOT aisbl 2018

Abstract

Purpose Many studies have found associations between laboratory biomarkers and periprosthetic joint infection (PJI), but it remains unclear whether these biomarkers are clinically useful in ruling out PJI. This meta-analysis compared the performance of interleukin-6 (IL-6) versus procalcitonin (PCT) for the diagnosis of PJI.

Methods In this meta-analysis, we reviewed studies that evaluated IL-6 or/and PCT as a diagnostic biomarker for PJI and provided sufficient data to permit sensitivity and specificity analyses for each test. The major databases MEDLINE, EMBASE, the Cochrane Library, Web of Science, and SCOPUS were searched for appropriate studies from the earliest available date of indexing through February 28, 2017. No restrictions were placed on language of publication.

Results We identified 18 studies encompassing a total of 1,835 subjects; 16 studies reported on IL-6 and 6 studies reported on PCT. The area under the curve (AUC) was 0.93 (95% CI, 0.91–0.95) for IL-6 and 0.83 (95% CI, 0.79–0.86) for PCT. The pooled sensitivity was 0.83 (95% CI, 0.74–0.89) for IL-6 and 0.58 (95% CI, 0.31–0.81) for PCT. The pooled specificity was 0.91 (95% CI, 0.84–0.95) for IL-6 and 0.95 (95% CI, 0.63–1.00) for PCT. Both the IL-6 and PCT tests had a high positive likelihood ratio (LR); 9.3 (95% CI, 5.3–16.2) and 12.4 (95% CI, 1.7–89.8), respectively, making them excellent rule-in tests for the diagnosis of PJI. The pooled negative LR for IL-6 was 0.19 (95% CI, 0.12–0.29), making it suitable as a rule-out test, whereas the pooled negative LR for PCT was 0.44 (95% CI, 0.25–0.78), making it unsuitable as a rule-out diagnostic tool.

Conclusions Based on the results of the present meta-analysis, IL-6 has higher diagnostic value than PCT for the diagnosis of PJI. Moreover, the specificity of the IL-6 test is higher than its sensitivity. Conversely, PCT is not recommended for use as a rule-out diagnostic tool.

Keywords Interleukin-6 · Procalcitonin · Diagnosis · Meta-analysis

Introduction

Periprosthetic joint infection (PJI) remains one of the most challenging complications of total joint arthroplasty (TJA), with an incidence ranging between 0.5 and 2.0%

[1, 2]. Since the symptoms of PJI are often nonspecific and there are no gold standard thresholds or criteria for currently available laboratory tests, PJI is difficult to diagnose precisely and quickly, thus hindering the effective management of PJI [3, 4]. For example, serum biomarkers such as erythrocyte sedimentation rate (ESR), C-reactive protein (CRP) and white blood cell (WBC) count are not sufficiently specific to diagnose PJI [5, 6]. Although synovial fluid biomarkers such as synovial fluid WBC count and differential count may help in the diagnosis of PJI, these biomarkers are normally increased by associated edema in the extremity and/or expected inflammation around the surgical site in the early post-operative period [7]. Furthermore, previous studies evaluating the accuracy of microbiologic cultures for

Research of this study was performed at Veterans Health Service Medical Center.

✉ Young-Soo Shin
sysoo3180@naver.com

¹ Department of Orthopaedic Surgery, Veterans Health Service Medical Center, 61 Jinhwangdoro-gil, Gangdong-Gu, Seoul 134-791, Republic of Korea

detecting PJI found that sample contamination may result in false-positive results. Moreover, microbiologic cultures may yield false-negative results, especially if patients have been on antibiotics prior to sampling [8, 9]. Therefore, diagnostic markers or panels of markers with the desirable properties of high sensitivity, high specificity, and clinical meaning are needed to realize rapid diagnosis and differentiation in various clinical settings. Recent follow-up studies focusing on the diagnostic accuracy of novel biomarkers suggested that both interleukin-6 (IL-6) and procalcitonin (PCT) were useful biomarkers. These two biomarkers could be assessed rapidly and had high sensitivity and specificity [10, 11], but it is still unknown whether they are clinically useful in ruling out PJI. To date, no published systematic review has compared the diagnostic performances of IL-6 and PCT. Therefore, in this study we compared the diagnostic performance of IL-6 versus that of PCT for the diagnosis of PJI.

Materials and methods

Data and literature sources

This study was conducted according to the guidance of the Cochrane Review Methods. Multiple comprehensive databases, including MEDLINE (January 1, 1976 to February 28, 2017), EMBASE (January 1, 1985 to February 28, 2017), Web of Science (January 1, 1980 to February 28, 2017), SCOPUS (January 1, 1980 to February 28, 2017), and the Cochrane Library (January 1, 1987 to February 28, 2017) were searched for studies that evaluated the diagnostic value of IL-6 or PCT for PJI diagnosis. There were no restrictions on language. Search terms were used in the title, abstract, MeSH, and keyword fields ('prosthesis-related infections' [Mesh] OR 'arthroplasty, replacement' [Mesh] OR 'joint prosthesis' [Mesh] OR 'periprosthetic joint infection [tiab] 'prosthetic

Fig. 1 Preferred reporting items for systemic reviews and meta-analyses (PRISMA) flow diagram of literature selection

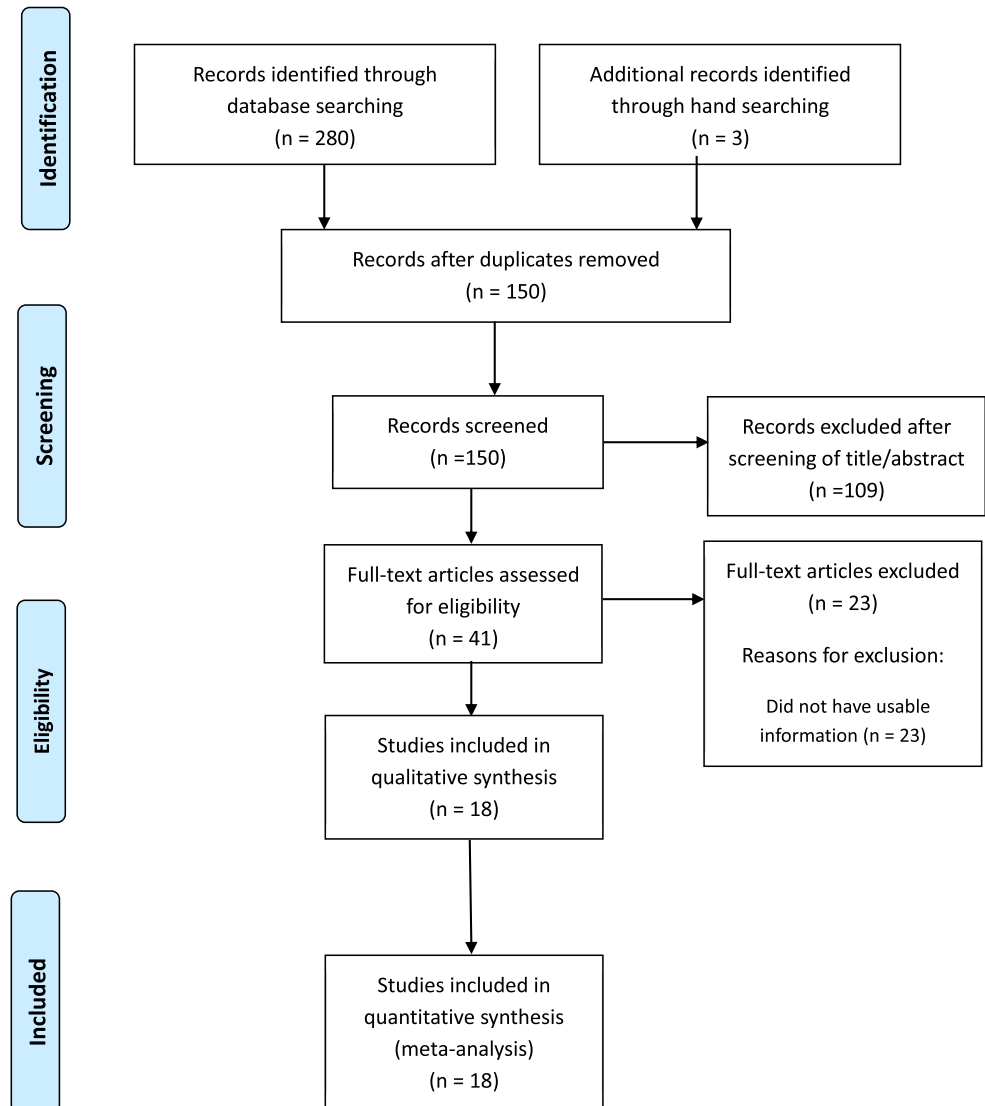


Table 1 Summary of study characteristics

Study	Year	A/B	Gender (male/ female)	Mean age (years)	Inflammatory markers	Joint type	Cut-off values	Diagnosis standard
Abou El-Khier et al. [17]	2013	11/40	21/19	58	Serum IL-6	Hip or knee	10.4 pg/mL	MSIS
Bottner et al. [20]	2007	21/78	41/78	64	Serum IL-6; Serum PCT	Hip or knee	12.0 pg/mL; 0.3 ng/mL	Other
Deirmengian et al. [7]	2010	14/51	23/51	65	Synovial fluid IL-6	Hip or knee	13,350 pg/mL	Other
Deirmengian et al. [21]	2014	29/95	44/95	67	Synovial fluid IL-6	Hip or knee	2.3 ng/mL	MSIS
Di Cesare et al. [10]	2005	17/58	25/58	63	Serum IL-6	Hip or knee	10 pg/mL	Other
Elgeidi et al. [19]	2014	11/40	21/40	59	Serum IL-6	Hip or knee	10.4 pg/mL	Other
Etinger et al. [25]	2015	41/98	42/56	67	Serum IL-6; Serum PCT	Hip, knee or shoulder	5.12 pg/mL; 0.025 ng/mL	MSIS
Frangiamore et al. [24]	2016	31/90	39/90	64	Synovial fluid IL-6	Hip or knee	8671 pg/mL	MSIS
Glehr et al. [22]	2013	78/124	NA	66	Serum IL-6; Serum PCT	Hip or knee	2.55 pg/mL; 0.35 ng/mL	MSIS
Gollwitzer et al. [18]	2013	15/35	12/35	70	Serum IL-6; Synovial fluid IL-6	Hip or knee	1.89 pg/mL; 1897 pg/mL	Other
Jacovides et al. [16]	2011	31/74	33/74	65	Synovial fluid IL-6	Hip or knee	4270 pg/mL	Other
Lenski et al. [2]	2014	22/40	NA	N/S	Synovial fluid IL-6	Hip, knee or elbow	30,750 pg/mL	MSIS
Nilsdotter-Augustinsson et al. [30]	2007	25/131	64/67	68	Synovial fluid IL-6	Hip	10,000 pg/mL	Other
Randau et al. [26]	2014	48/120	47/73	68	Serum IL-6; Synovial fluid IL-6; Serum PCT	Hip or knee	2.6 pg/mL; 2,100 pg/mL; 46 ng/mL	MSIS
Saeed et al. [29]	2013	8/14	NA	73	Synovial fluid PCT	Hip or knee	2.5 µg/L	Other
Tang et al. [27]	2016	21/52	22/52	64	Serum IL-6	Hip or knee	7.5 ng/L	MSIS
Worthington et al. [23]	2016	16/46	25/46	72	Serum IL-6	Hip	9 pg/mL	Other
Yuan et al. [28]	2015	25/74	22/74	65	Serum PCT	Hip	0.5 ng/mL	Other

A number of total infection, B number of total evaluated subjects, IL interleukin, PCT procalcitonin, MSIS musculoskeletal infection society, NA not available

infection' [tiab] OR 'arthroplasty' [tiab]) AND 'interleukin-6' [tiab] OR 'IL-6' [tiab] OR 'interleukin-6' [Mesh] OR 'inflammatory markers' [tiab] OR 'procalcitonin' [tiab]). After the initial electronic search, relevant articles and their bibliographies were searched manually.

Study selection

Two reviewers independently selected relevant studies for full review by searching through titles and abstracts. A full text copy of each article was reviewed if the abstract did not provide enough data to make a decision. Studies were included in the meta-analysis if they: (1) assessed IL-6 or PCT as an index test to evaluate the presence of PJI; (2) had the reference standard of PJI defined as a communicating sinus tract with the

prosthesis, a positive microbiological culture, or histopathological findings; and (3) fully reported cases in absolute numbers of true-positive, false-positive, false-negative, and true-negative outcomes.

Data extraction and quality assessment

Two reviewers independently recorded data from each study using a predefined data extraction form and resolved any differences by discussion. Data entry by these two reviewers was checked by a third investigator. Variables recorded were: (1) type of inflammatory marker (IL-6 and/or PCT) and sample size; (2) sensitivity, specificity, positive likelihood ratio (LR), negative LR, diagnostic odds ratio (DOR), area under the curve (AUC), location of arthroplasty, and diagnostic test

Table 2 Diagnostic performance of screening tests for the detection of IL-6 and PCT in patients with PJI

Study	Number of subjects per test result				Sensitivity (95% CI)	Specificity (95% CI)
	TP	FP	FN	TN		
Studies evaluating the IL-6 as the screening test						
Abou El-Khier et al. [17]	11	3	0	26	1.000 (0.715–1.000)	0.897 (0.726–0.978)
Bottner et al. [20]	20	7	1	50	0.952 (0.762–0.999)	0.877 (0.763–0.949)
Deirmengian et al. [7]	14	0	14	37	0.500 (0.306–0.694)	1.000 (0.905–1.000)
Deirmengian et al. [21]	26	2	3	64	0.897 (0.726–0.978)	0.970 (0.895–0.966)
Di Cesare et al. [10]	17	2	0	39	1.000 (0.805–1.000)	0.951 (0.835–0.994)
Elgeidi et al. [19]	11	3	0	26	1.000 (0.715–1.000)	0.897 (0.726–0.978)
Ettinger et al. [25]	33	7	8	51	0.805 (0.651–0.912)	0.879 (0.767–0.950)
Frangiamore et al. [24]	25	2	6	57	0.806 (0.625–0.925)	0.966 (0.883–0.996)
Glehr et al. [22]	63	15	15	31	0.808 (0.703–0.888)	0.674 (0.520–0.805)
Gollwitzer et al. [18] (Serum)	7	1	8	19	0.467 (0.213–0.734)	0.950 (0.751–0.999)
Gollwitzer et al. [18] (Synovial fluid)	9	1	6	19	0.600 (0.323–0.837)	0.950 (0.751–0.999)
Jacovides et al. [16]	27	0	4	43	0.871 (0.702–0.964)	1.000 (0.918–1.000)
Lenski et al. [2]	28	2	3	36	0.903 (0.742–0.980)	0.947 (0.823–0.994)
Nilsdotter-Augustinsson et al. [30]	17	7	8	99	0.680 (0.465–0.851)	0.934 (0.869–0.973)
Randau et al. [26] (Serum)	38	30	10	42	0.792 (0.650–0.895)	0.583 (0.461–0.698)
Randau et al. [26] (Synovial fluid)	30	10	18	62	0.625 (0.427–0.789)	0.857 (0.715–0.946)
Tang et al. [27]	20	15	1	16	0.952 (0.762–0.999)	0.516 (0.331–0.698)
Worthington et al. [23]	13	7	3	23	0.813 (0.544–0.960)	0.767 (0.577–0.901)
Summary estimates					0.830 (0.740–0.890)	0.910 (0.840–0.950)
Studies evaluating the PCT as the screening test						
Bottner et al. [20]	7	1	14	56	0.333 (0.146–0.570)	0.982 (0.906–1.000)
Ettinger et al. [25]	37	42	4	16	0.902 (0.769–0.973)	0.276 (0.167–0.409)
Glehr et al. [22]	37	0	41	46	0.474 (0.360–0.591)	1.000 (0.923–1.000)
Randau et al. [26]	6	0	42	72	0.125 (0.047–0.252)	1.000 (0.950–1.000)
Saeed et al. [29]	6	1	2	5	0.750 (0.349–0.968)	0.833 (0.359–0.996)
Yuan et al. [28]	20	12	5	34	0.800 (0.593–0.932)	0.739 (0.589–0.857)
Summary estimates					0.580 (0.310–0.810)	0.950 (0.630–1.000)

TP true-positive, FP false-positive, FN false-negative, TN true-negative, IL interleukin, PCT procalcitonin, PJI periprosthetic joint infection, CI confidence interval

conducted; (3) and cutoff value. Regarding (3), if a study presented several cutoff values for an index test, the data with the best estimates were extracted. If a study presented different cutoff values of an index test for serum versus synovial fluid, data from different biological fluids were analyzed as separate studies. Two reviewers independently assessed methodological quality based on quality assessment of diagnostic accuracy studies (QUADAS-2) tool [12]. Two reviewers resolved all differences by discussion; their consensus was checked by a third investigator.

Data synthesis and analysis

Measurements of diagnostic performance, such as sensitivity, specificity, positive LR, negative LR, DOR, and AUC, are reported as point estimates with 95% CIs for the diagnosis of PJI. The positive LR, negative LR, and post-test probability were calculated to evaluate the clinical utility of IL-6 and PCT using the summary estimates of sensitivity and specificity. The positive LR is the ratio of the probability of a positive test result if the subject has PJI to the probability of a positive test result if the subject does

Table 3 QUADAS-2 evaluation of included studies

Study	QUADAS score ^a														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Abou El-Khier et al. [17]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Bottner et al. [20]	1	0	1	1	1	1	1	1	1	1	1	1	1	0	12
Deirmengian et al. [7]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Deirmengian et al. [21]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
Di Cesare et al. [10]	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12
Elgeidi et al. [19]	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12
Ettinger et al. [25]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
Frangiamore et al. [24]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	14
Glehr et al. [22]	1	1	1	1	1	1	1	1	1	1	1	1	0	1	13
Gollwitzer et al. [18]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Jacovides et al. [16]	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12
Lenski et al. [2]	1	0	1	1	1	1	1	1	1	1	1	1	1	0	12
Nilsdotter-Augustinsson et al. [30]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Randau et al. [26]	1	1	1	1	1	1	1	1	1	1	1	1	0	0	12
Saeed et al. [29]	1	1	1	1	1	1	1	1	1	1	1	1	0	1	13
Tang et al. [27]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Worthington et al. [23]	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13
Yuan et al. [28]	1	1	1	1	1	1	1	1	1	1	1	0	1	0	12

^a A number 1 indicates “yes,” and 0 indicates “no”

The numbers in the top row correspond to the following questions:

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described?
3. Is the reference standard likely to correctly classify the target condition?
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test (i.e., the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit its replication?
10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/ intermediate test results reported?
14. Were withdrawals from the study explained?

not have PJI. Likewise, the negative LR is the ratio of the probability of a negative test result if the subject has PJI to the probability of a negative test result if the subject does not have PJI. Positive LR values greater than 5 and negative LR values less than 0.2 indicate strong diagnostic evidence for ruling in/ruling out diagnoses [13]. To describe the Fagan plot results, the pre-test probability and post-test probability were connected by a straight line crossing the likelihood ratio. The DOR was calculated to evaluate the diagnostic effectiveness of IL-6 and PCT, with higher values suggesting better discriminatory test performance [14]. An AUC value greater than 0.8 indicates a good diagnostic ability [2]. The bivariate random-effects model was used to incorporate the

correlation that might exist between sensitivity and specificity, resulting from the use of different thresholds across studies. In addition, we used the model to create hierarchical summary receiver operating characteristic (HSROC) curves. All statistical analyses were performed with Stata version 14.2 static software; the metandi and midas commands were utilized for all analyses. Publication bias was assessed by using the effective sample size funnel plot and associated regression test of asymmetry described by Deeks et al. [15]. Heterogeneity was determined by estimating the proportion of between-study inconsistencies due to actual differences between studies, rather than differences due to random error or chance, using the I^2 statistic. Values of 25%, 50%, and 75% were considered to

Fig. 2 Funnel plot for publication bias assessment of included studies. **a** Interleukin-6 and **b** procalcitonin, for the detection of PJI. *ESS* effective sample size

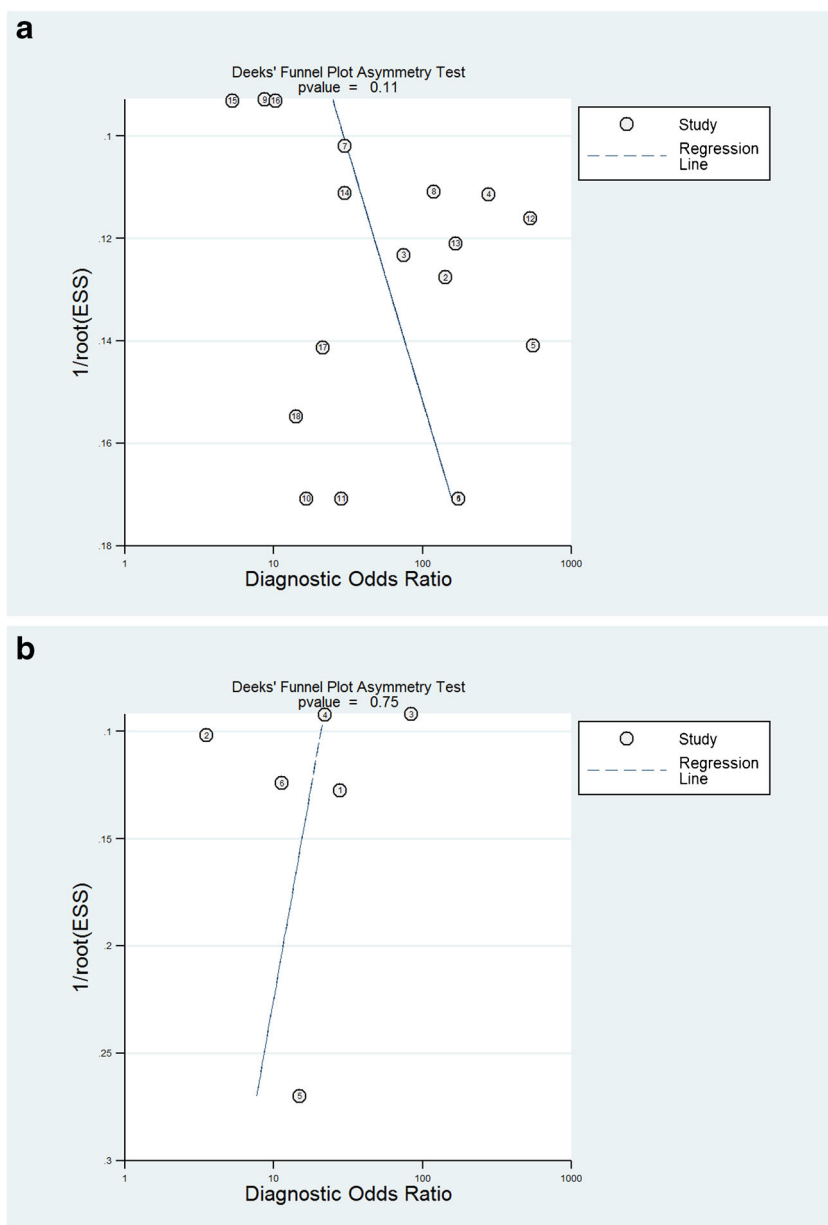


Table 4 Subgroup analysis for the diagnostic performance of screening tests according to specific study designs

Variable	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)	Positive LR (95% CI)	Negative LR (95% CI)	AUC (95% CI)	DOR (95% CI)	I ² (95% CI)
Overall analysis								
IL-6	18	0.83 (0.74–0.89)	0.91 (0.84–0.95)	9.3 (5.3–16.2)	0.19 (0.12–0.29)	0.93 (0.91–0.95)	49 (24–103)	96 (94–99)
PCT	6	0.58 (0.31–0.81)	0.95 (0.63–1.00)	12.4 (1.7–89.8)	0.44 (0.25–0.78)	0.83 (0.79–0.86)	28 (6–143)	99 (98–99)
Studies evaluating both the IL-6 and the PCT								
IL-6	4	0.83 (0.75–0.88)	0.77 (0.62–0.88)	3.7 (2.0–6.7)	0.22 (0.14–0.35)	0.86 (0.83–0.89)	16 (6–45)	27 (0–100)
PCT	4	0.48 (0.17–0.80)	0.98 (0.59–1.00)	28.2 (1.4–569.2)	0.53 (0.28–1.03)	0.82 (0.79–0.85)	53 (4–682)	99 (98–99)

IL interleukin, PCT procalcitonin, CI confidence interval, LR likelihood ratio, AUC area under the curve, DOR diagnostic odds ratio, I² I-square

indicate low, moderate, and high heterogeneity, respectively. When statistical heterogeneity was substantial, we conducted meta-regression to identify potential sources of

bias, such as the location, diagnostic standard, cutoff value, and location where samples were obtained. The number, age, and sex of the study subjects were also

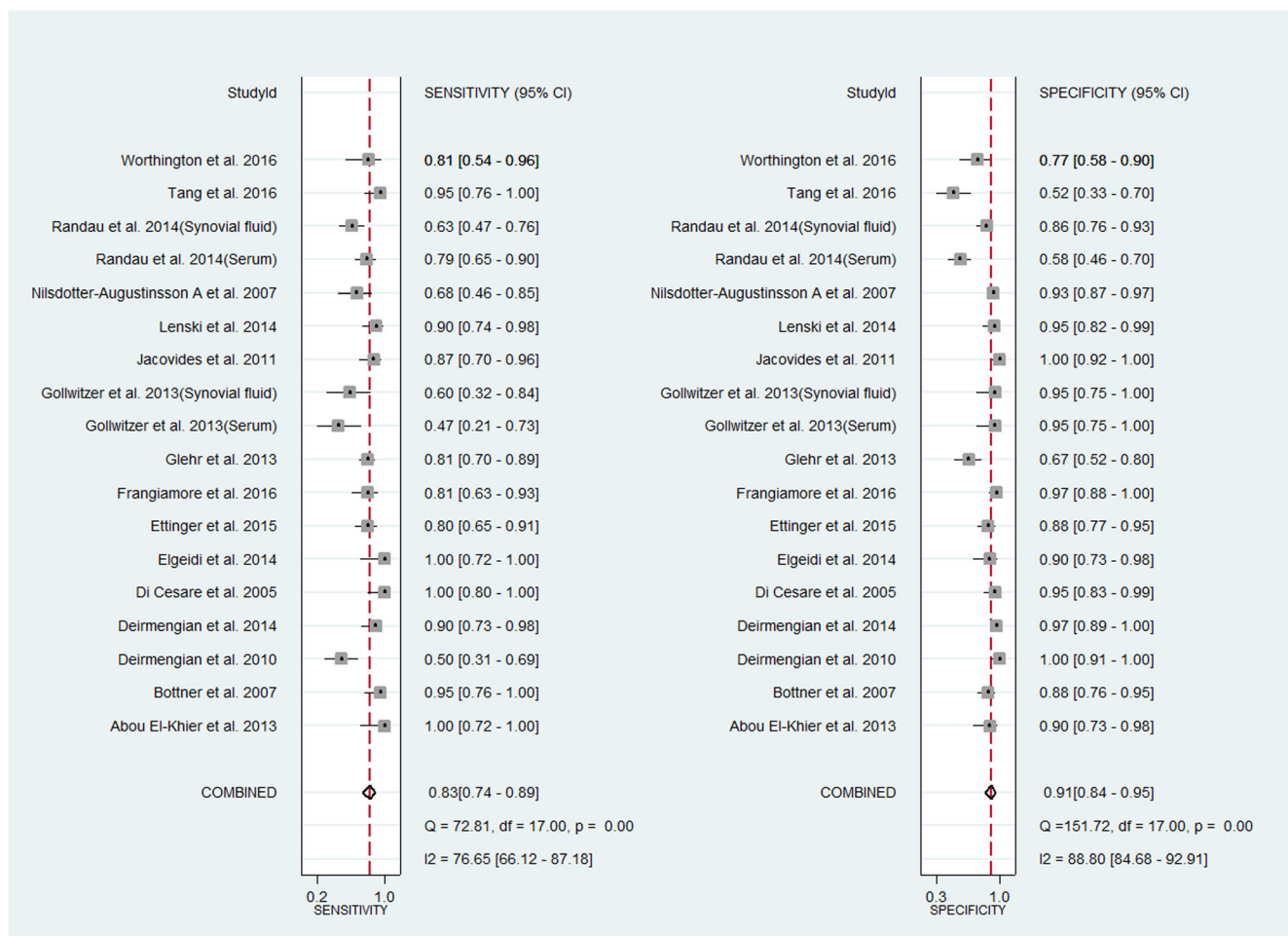


Fig. 3 Paired forest plots of the sensitivity and specificity of interleukin-6 for the detection of PJI

considered. Subgroup analyses were performed for studies evaluating both IL-6 and PCT in an attempt to explore potential sources of heterogeneity.

Results

Identification of studies

The details of study identification, inclusion, and exclusion are summarized in Fig. 1. An electronic search yielded 63 studies of interest in PubMed (MEDLINE), 75 in EMBASE, 80 in Web of Science, 54 in SCOPUS, and eight in the Cochrane Library. Three additional publications were identified through manual searching. After removing 132 duplicates, 150 studies remained; of these, 109 were excluded after reading the abstracts and full-text articles, and an additional 23 studies were excluded because they had unusable information. This process eventually resulted in 18 studies that were included in the final meta-analysis [2, 7, 10, 16–30].

Study characteristics and patient samples

The 18 studies we examined included 1,327 subjects in whom the diagnostic accuracy of IL-6 for PJI was examined and six studies included 508 subjects in whom the diagnostic accuracy of PCT for PJI was examined. For IL-6, 16 studies included

patients who had hip or knee arthroplasty and two included patients who had shoulder or elbow arthroplasty. The mean patient age ranged from 58 to 72 years, and 30 to 53% of the patients were male. For PCT, five studies included patients who had hip or knee arthroplasty and one included patients who had shoulder arthroplasty. The mean patient age ranged from 64 to 73 years, and 34 to 43% of the patients were male. Two studies reported on IL-6 using different cutoff values for serum and synovial fluid. The clinical and methodological characteristics, as well as the main results of each study, are summarized in Tables 1 and 2.

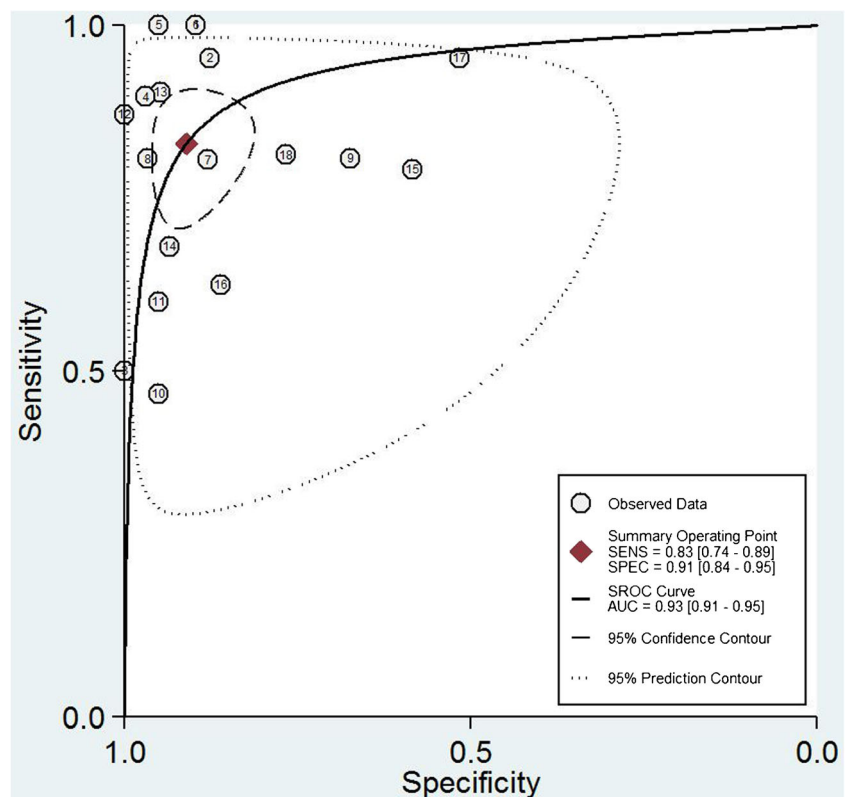
Quality and publication biases of the included studies

The methodological quality of the included studies was evaluated using the QUADAS-2 tool [12]. This tool includes four domains that assess patient selection, index tests, reference standards, and flow and timing. Overall, the quality of the studies was deemed satisfactory (Table 3). The funnel plots and regression tests indicated no significant publication bias ($P=0.11$ and $P=0.75$ for IL-6 and PCT, respectively) (Fig. 2a, b).

Diagnostic accuracy of interleukin-6 for PJI

The pooled sensitivity across studies for IL-6 was 0.83 (95% CI, 0.74–0.89), the pooled specificity was 0.91 (95% CI,

Fig. 4 Hierarchical summary receiver operating characteristic curves of interleukin-6 for the detection of PJI. *AUC* area under the curve, *SENS* sensitivity, *SPEC* specificity, *SROC* summary receiver operating characteristic



0.84–0.95), the pooled positive LR was 9.3 (95% CI, 5.3–16.2), the pooled negative LR 0.19 (95% CI, 0.12–0.29), and the pooled DOR was 49 (95% CI, 24–103) (Table 4). The positive LR was sufficiently high to qualify IL-6 testing as a rule-in diagnostic tool. Similarly, the negative LR was sufficiently low to qualify IL-6 testing as a rule-out diagnostic tool. The I^2 statistics for sensitivity and specificity were 77% (95% CI, 66–87%) and 79% (95% CI, 85–93%), indicating that there was substantial heterogeneity (Fig. 3). The HSROC curve for the index test indicated that the AUC was 0.93 (95% CI, 0.91–0.95) for IL-6 (Fig. 4). The Fagan plot showed that a positive result on the IL-6 test increased the probability of PJI from 38 to 85% and a negative result in the IL-6 test decreased the probability of PJI to 10% (Fig. 5).

Diagnostic accuracy of procalcitonin for PJI

The pooled sensitivity across studies for PCT was 0.58 (95% CI, 0.31–0.81), the pooled specificity was 0.95 (95% CI, 0.63–1.00), the pooled positive LR was 12.4 (95% CI, 1.7–89.8), the pooled negative LR was 0.44 (95% CI, 0.25–0.78), and the pooled DOR was 28 (95% CI, 6–143) (Table 4). The positive LR was sufficiently high to qualify PCT testing as a rule-in diagnostic tool, whereas the relatively high negative LR (greater than 0.2) was not sufficiently low to qualify PCT testing as a rule-out diagnostic tool. The I^2 statistics for sensitivity and specificity were 93% (95% CI, 88–97%) and 97% (95% CI, 95–98%), indicating that there was substantial heterogeneity (Fig. 6). The HSROC curve for the index test indicated that the AUC was 0.83 (95% CI, 0.79–0.86) for PCT (Fig. 7). The Fagan plot showed that a positive result on the PCT test increased the probability of PJI from 44 to 91% and a negative result in the PCT test decreased the probability of PJI to 26% (Fig. 8).

Meta-regression and subgroup analysis

Between-study heterogeneity was shown for sensitivity and specificity among studies of both index tests. Thus, univariate meta-regression analysis was performed to identify potential sources of heterogeneity (Table 5). For studies evaluating IL-6, the location where the samples were obtained was the most probable source of heterogeneity ($P < 0.01$). The location of study publication ($P = 0.03$) and the age of the study subjects ($P = 0.01$) also explained some of this heterogeneity. For both index tests, neither the diagnostic standard nor the cutoff values were the main source of heterogeneity. In addition, we performed subgroup analysis for the four studies that evaluated both IL-6 and PCT in the same patients (Table 4). When analyzing only these four studies, the negative LRs of IL-6 and PCT increased to 0.22 (95% CI 0.14–0.35) and 0.53 (95% CI 0.28–1.03), respectively, indicating that PJI is unlikely when both IL-6 and PCT are evaluated in a single patient.

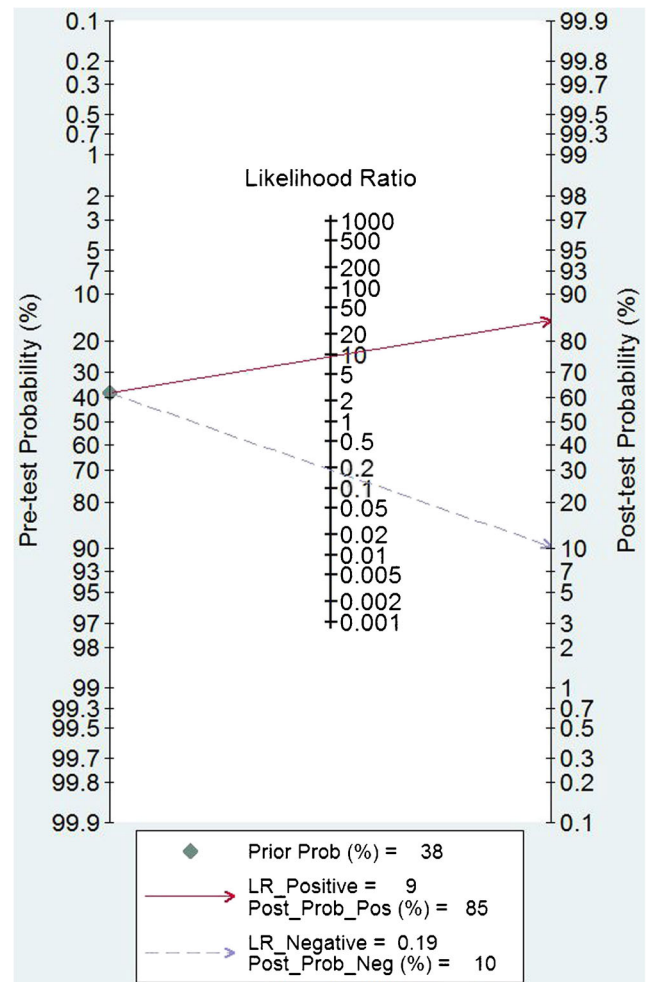


Fig. 5 Pre-test probabilities and likelihood ratios for IL-6. LR likelihood ratio

Pooled results from these studies showed a greatly reduced specificity for IL-6, 0.77 (95% CI 0.62–0.88), and a sensitivity for PCT of 0.48 (95% CI 0.17–0.80).

Discussion

The main finding of the present meta-analysis is that IL-6 has a higher diagnostic value than PCT for distinguishing PJI from other causes of failure. Specifically, the AUC value was 0.93 for IL-6 and 0.83 for PCT. Both biomarkers had an optimal positive LR that was sufficiently high to qualify testing as a rule-in diagnostic tool. Conversely, the PCT had a suboptimal negative LR, making it insufficient to function as a rule-out biomarker.

IL-6 is released by monocytes in response to a local infection and is the main stimulator of CRP production in liver cells. Thus, the IL-6 response to infection is much more rapid than that of CRP and the IL-6 levels quickly return to normal after surgery [31]. Previous studies have shown that IL-6

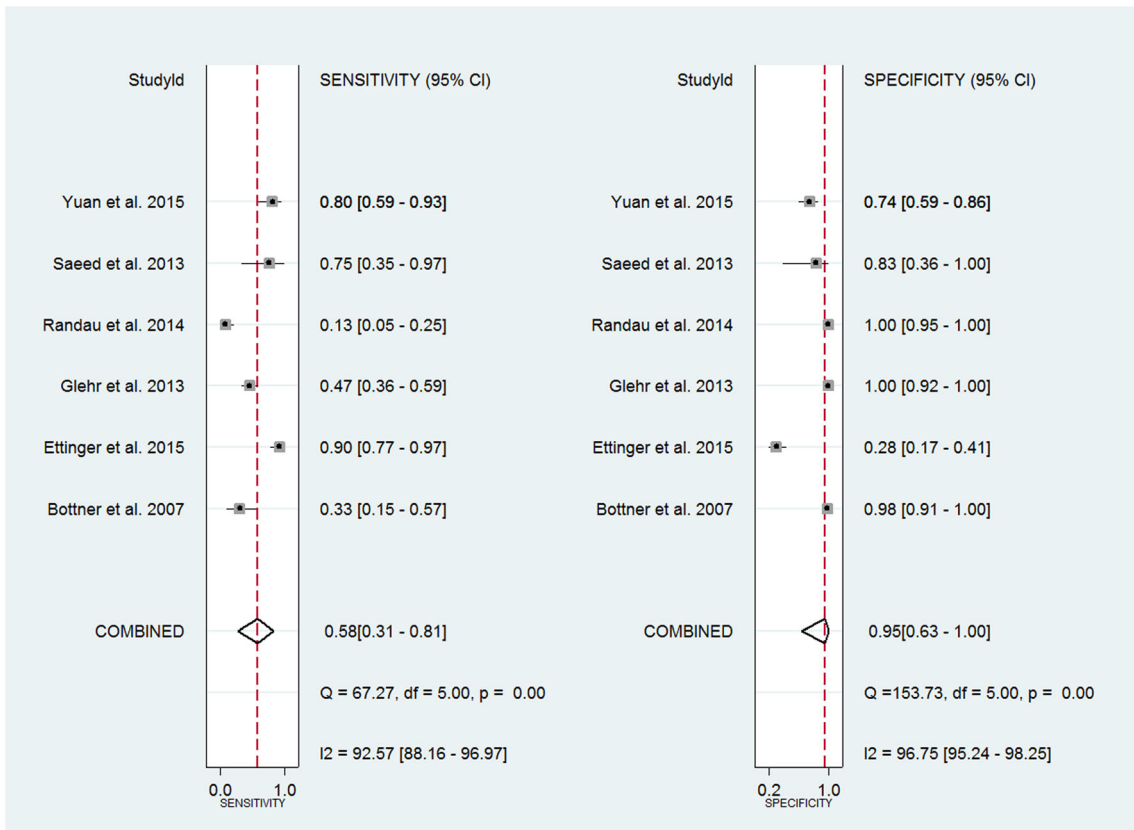
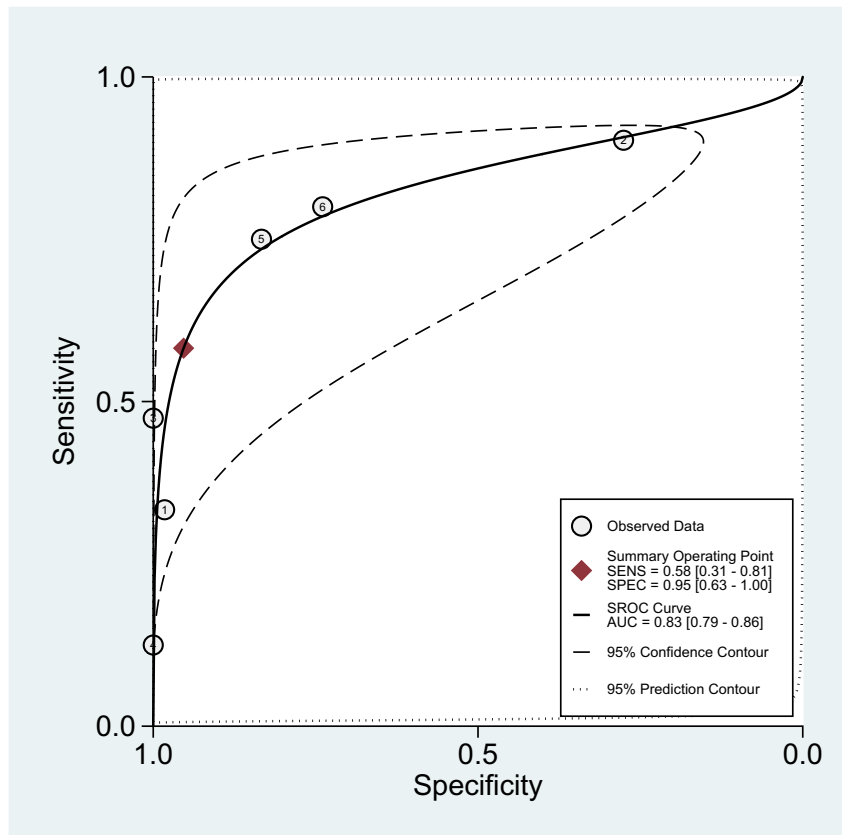


Fig. 6 Paired forest plots of the sensitivity and specificity of procalcitonin for the detection of PJI

Fig. 7 Hierarchical summary receiver operating characteristic curves of procalcitonin for the detection of PJI. *AUC* area under the curve, *SENS* sensitivity, *SPEC* specificity, *SROC* summary receiver operating characteristic



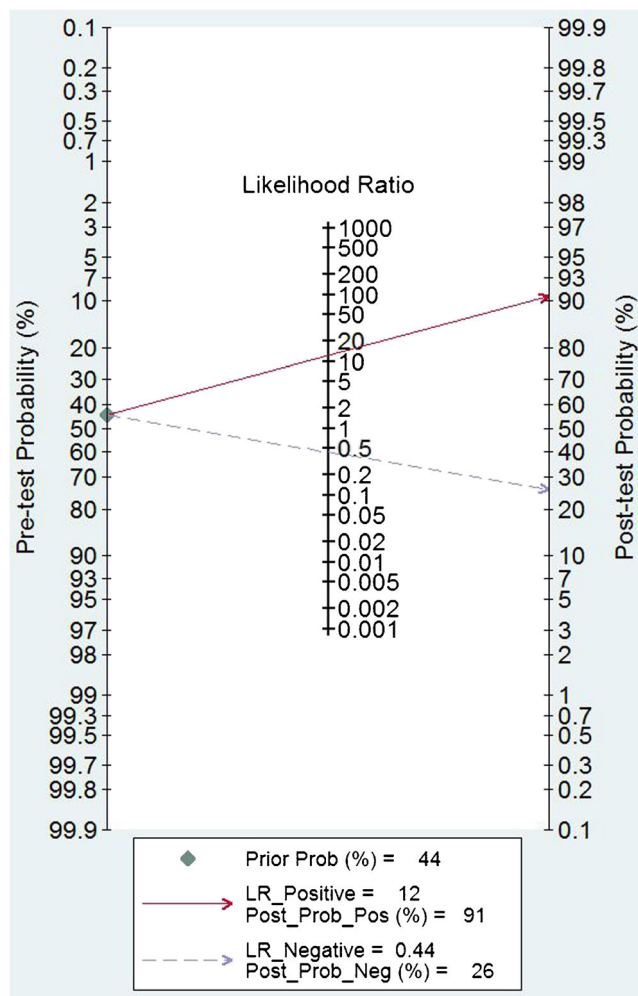


Fig. 8 Pre-test probability and likelihood ratios for procalcitonin. *LR* likelihood ratio

levels are not increased in patients with aseptic loosening [31], whereas high concentrations of IL-6 have been detected in interface tissue from patients with loosening of prosthesis, but no infection [32]. Indeed, our findings from subgroup analysis of studies evaluating both IL-6 and PCT in the same patients suggested that IL-6 has much lower specificity (77%) compared with the findings from overall analysis (91%). These results may be attributable to the fact that IL-6 levels in peripheral blood appear to be elevated in patients with aseptic loosening of total hip arthroplasty, because monocytes respond to polyethylene particles by producing IL-6 [20]. Another factor that could explain these results are differences in patients such as those with chronic inflammatory disease and those with Paget disease or immunodeficiency syndromes. Much more IL-6 can be detected in some of these diseases, which may be responsible for the limited accuracy and subsequent reduction of specificity of IL-6 [26].

Although both biomarkers are readily available and can be serially monitored, synovial fluid biomarkers should

theoretically lead to more reliable and accurate diagnosis of PJI compared with serum biomarkers because synovial fluid biomarkers are obtained directly from the affected joint [33]. One study that evaluated the sensitivity, specificity, and accuracy of 24 synovial fluid biomarkers in patients with PJI versus aseptic disease reported that synovial fluid IL-6 had excellent diagnostic performance, with accuracy above 0.9 for the diagnosis of hip and knee PJI [7]. This finding corresponds well with the results of a recent study reporting that synovial fluid IL-6 had an accuracy of 0.89 with very high sensitivity, subsequently leading to strong diagnostic strength [24]. In contrast, another study investigating the serum and synovial PCT levels in 42 patients with arthritis found that serum PCT was the best biomarker to discriminate patients with septic arthritis from patients with non-septic arthritis, whereas synovial PCT was not helpful for distinguishing between infectious and non-infectious arthritis [34]. Considering the possible influence of the location where the samples were obtained from the study subjects on diagnostic performance, we further evaluated this issue by meta-regression analyses. For IL-6, whether samples were collected from synovial fluid or not appeared to be the most probable source of heterogeneity, whereas this was not the case for PCT. As expected, this discrepancy was likely due to the lower number of synovial fluid studies in the PCT group compared to the IL-6 group. Intriguingly, the results of this meta-analysis did not support previous findings, in that suboptimal sensitivity (58%) and a negative LR (0.44) was found for the six PCT studies, including five that used serum and one that used synovial fluid for diagnosing PJI. The poor outcomes for PCT in detecting PJI may be explained by the fact that PCT is a more accurate marker for systemic bacterial infection and an inaccurately low PCT level is seen during localized infections. These findings suggest that PCT should not be utilized as a rule-out diagnostic tool in patients with localized infections [20, 23].

Investigating the sources of heterogeneity is key for determining whether our conclusions can be applied across different studies. In the diagnostic studies, the threshold effect was regarded as a major cause of heterogeneity. The I^2 results of both IL-6 and PCT suggested that the current meta-analysis had remarkable heterogeneity. The calculated Spearman correlation coefficient for IL-6 was 0.191 ($P=0.447$), which does not support that heterogeneity resulted from the threshold effect, whereas the coefficient for PCT was 0.943 ($P=0.005$). In addition, some possible causes of heterogeneity, including study location, number of patients, age, sex, diagnostic standard, cut-off value, and location where samples were obtained were explored by meta-regression. Unfortunately, we did not identify the source of PCT heterogeneity in this meta-analysis. However, we did identify study location ($P=0.03$), age ($P=0.01$), and location where samples were obtained ($P<0.01$) as sources of heterogeneity

Table 5 Univariate meta-regression analysis for identifying potential sources of heterogeneity in the diagnostic performance of screening tests

Variable (<i>P</i> -value)	Number of studies	Sensitivity		Specificity	
		Adjusted	<i>P</i> -value	Adjusted	<i>P</i> -value
Studies evaluating the IL-6 as the screening test					
Study location (<i>P</i> = 0.03)					
Western countries	13	0.83	0.15	0.94	<0.01
Other countries	5	0.82		0.77	
Number of patients (<i>P</i> = 0.84)					
≤70	9	0.84	0.08	0.92	0.10
≥70	9	0.82		0.90	
Age, mean, year (<i>P</i> = 0.01)					
≤65	8	0.91	<0.01	0.94	0.04
≥65	10	0.76		0.89	
Men, % (<i>P</i> = 0.36)					
≤48	13	0.80	0.97	0.92	0.14
≥48	5	0.90		0.89	
Diagnostic standard (<i>P</i> = 0.16)					
MSIS	9	0.85	0.06	0.94	<0.01
Others	9	0.81		0.86	
Cutoff value of the IL-6, pg/mL (<i>P</i> = 0.35)					
Synovial fluid ≤ 10,000 or serum ≤ 10	12	0.81	0.51	0.89	0.65
Synovial fluid ≥ 10,000 or serum ≥ 10	6	0.87		0.94	
Where to obtain samples (<i>P</i> < 0.01)					
Blood serum	10	0.88	<0.01	0.82	0.97
Joint aspirate	8	0.76		0.96	
Studies evaluating the PCT as the screening test					
Study location (<i>P</i> = 0.42)					
Western countries	3	0.69	0.36	0.83	<0.01
Other countries	3	0.46		0.99	
Number of patients (<i>P</i> = 0.85)					
≤70	1	0.76	0.63	0.87	0.01
≥70	5	0.55		0.96	
Age, mean, year (<i>P</i> = 0.92)					
≤65	2	0.58	0.95	0.94	0.03
≥65	4	0.58		0.96	
Men, % (<i>P</i> = 0.87)					
≤48	4	0.60	0.72	0.96	0.05
≥48	2	0.52		0.96	
Diagnostic standard (<i>P</i> = 0.87)					
MSIS	3	0.53	0.72	0.98	0.01
Others	3	0.63		0.92	
Cutoff value of the PCT, ng/mL (<i>P</i> = 0.09)					
Synovial fluid ≤ 460 or serum ≤ 30	5	0.68	0.01	0.89	0.04
Synovial fluid ≥ 460 or serum ≥ 30	1	0.12		1.00	
Where to obtain samples (<i>P</i> = 0.85)					
Blood serum	5	0.55	0.63	0.96	0.01
Joint aspirate	1	0.76		0.87	

IL interleukin, PCT procalcitonin, MSIS musculoskeletal infection society

for IL-6. Thus, the heterogeneity of IL-6 in the included studies was likely caused by these three factors.

This study has several limitations. First, half of the studies had a small sample size (< 70 subjects), which could have led

to overestimation of the diagnostic accuracy of both IL-6 and PCT for detecting PJI. However, our meta-regression analysis showed that sample size did not have much impact on the study outcomes. Second, positive and negative LR_s were

calculated from binary data. The results of both biomarkers were either positive or negative, meaning that useful information could have been lost because their concentrations increase as disease severity increases [35]. To obtain more precise information regarding test reliability, it will be necessary to calculate LR_s based on multiple cutoffs. Finally, the IL-6 and PCT detection assays were different among the studies, which could have negatively affected the assessment of diagnostic accuracy.

Conclusions

Based on the results of the present meta-analysis, IL-6 has a higher diagnostic value than PCT for the diagnosis of PJI, and the IL-6 test has higher specificity than sensitivity. Conversely, PCT is not recommended for use as a rule-out diagnostic tool.

Compliance with ethical standards

Conflict of interest The authors declare that they have to no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

There are no conflicts of interest between any of the contributors to this manuscript.

References

- Matar WY, Jafari SM, Restrepo C et al (2010) Preventing infection in total joint arthroplasty. *J Bone Joint Surg Am* 92(Suppl 2):36–46
- Lenski M, Scherer MA (2014) Synovial IL-6 as inflammatory marker in periprosthetic joint infections. *J Arthroplast* 29(6):1105–1109
- Parvizi J, Ghanem E, Menashe S et al (2006) Periprosthetic infection: what are the diagnostic challenges? *J Bone Joint Surg Am* 88(suppl 4):138–147
- Zimmerli W, Trampuz A, Ochsner PE (2004) Prosthetic-joint infections. *N Engl J Med* 351(16):1645–1654
- Austin MS, Ghanem E, Joshi A et al (2008) A simple, cost-effective screening protocol to rule out periprosthetic infection. *J Arthroplast* 23(1):65–68
- Ghanem E, Antoci V, Pulido L et al (2009) The use of receiver operating characteristics analysis in determining erythrocyte sedimentation rate and C-reactive protein levels in diagnosing periprosthetic infection prior to revision total hip arthroplasty. *Int J Infect Dis* 13(6):e444–e449
- Deirmengian C, Hallab N, Tarabishy A et al (2010) Synovial fluid biomarkers for periprosthetic infection. *Clin Orthop Relat Res* 468(8):2017–2023
- Parvizi J, Fassihi SC, Enayatollahi MA (2016) Diagnosis of periprosthetic joint infection following hip and knee Arthroplasty. *Orthop Clin N Am* 47(3):505–515
- Parvizi J, Cavanaugh PK, Diaz-Ledezma C (2013) Periprosthetic knee infection: ten strategies that work. *Knee Surg Relat Res* 25(4):155–164
- Di Cesare PE, Chang E, Preston CF et al (2005) Serum interleukin-6 as a marker of periprosthetic infection following total hip and knee arthroplasty. *J Bone Joint Surg Am* 87(9):1921–1927
- Maharajan K, Patro DK, Menon J et al (2013) Serum procalcitonin is a sensitive and specific marker in the diagnosis of septic arthritis and acute osteomyelitis. *J Orthop Surg Res* 8(1):19
- Whiting PF, Rutjes AW, Westwood ME et al (2011) QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 155(8):529–536
- Deeks JJ (2001) Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 323(7305):157–162
- Glas AS, Lijmer JG, Prins MH et al (2003) The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 56(11):1129–1135
- Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 58(9):882–893
- Jacovides CL, Parvizi J, Adeli B et al (2011) Molecular markers for diagnosis of periprosthetic joint infection. *J Arthroplast* 26(Suppl 6):99–103
- Abou El-Khier NT, El Ganainy AR, Elgeidy A et al (2013) Assessment of interleukin-6 and other inflammatory markers in the diagnosis of Egyptian patients with periprosthetic joint infection. *Egypt J Immunol* 20(2):93–99
- Gollwitzer H, Dombrowski Y, Prodingner PM et al (2013) Antimicrobial peptides and proinflammatory cytokines in periprosthetic joint infection. *J Bone Joint Surg Am* 95(7):644–651
- Elgeidi A, Elganainy AE, Abou Elkhier N et al (2014) Interleukin-6 and other inflammatory markers in diagnosis of periprosthetic joint infection. *Int Orthop* 38(12):2591–2595
- Bottner F, Wegner A, Winkelmann W et al (2007) Interleukin-6, procalcitonin and TNF-alpha: markers of peri-prosthetic infection following total joint replacement. *J Bone Joint Surg Br* 89(1):94–99
- Deirmengian C, Kardos K, Kilmartin P et al (2014) Diagnosing periprosthetic joint infection: has the era of the biomarker arrived? *Clin Orthop Relat Res* 472(11):3254–3262
- Glehr M, Friesenbichler J, Hofmann G et al (2013) Novel biomarkers to detect infection in revision hip and knee arthroplasties. *Clin Orthop Relat Res* 471(8):2621–2628
- Worthington T, Dunlop D, Casey A et al (2010) Serum procalcitonin, interleukin-6, soluble intercellular adhesion molecule-1 and IgG to shortchain exocellular lipoteichoic acid as predictors of infection in total joint prosthesis revision. *Br J Biomed Sci* 67(2):71–76
- Frangiamore SJ, Siqueira MB, Saleh A et al (2016) Synovial cytokines and the MSIS criteria are not useful for determining infection resolution after periprosthetic joint infection explantation. *Clin Orthop Relat Res* 474(7):1630–1639
- Ettinger M, Calliess T, Kielstein JT et al (2015) Circulating biomarkers for discrimination between aseptic joint failure, low-grade infection and high-grade septic failure. *Clin Infect Dis* 61:332–341
- Randau TM, Friedrich MJ, Wimmer MD et al (2014) Interleukin-6 in serum and in synovial fluid enhances the differentiation between periprosthetic joint infection and aseptic loosening. *PloS One* 9(2):e89045
- Tang X, Wang Q, Wang H et al (2016) [Comparison and estimation of different diagnostic methods in detecting the presence of periprosthetic joint infection]. *Zhonghua Wai Ke Za Zhi* 54(4):251–257
- Yuan K, Li W-D, Qiang Y et al (2015) Comparison of procalcitonin and C-reactive protein for the diagnosis of periprosthetic joint infection before revision total hip arthroplasty. *Surg Infect* 16(2):146–150

29. Saeed K, Dryden M, Sitjar A et al (2013) Measuring synovial fluid procalcitonin levels in distinguishing cases of septic arthritis, including prosthetic joints, from other causes of arthritis and aseptic loosening. *Infection* 41(4):845–849
30. Nilsson-Augustinsson A, Briheim G, Herder A et al (2007) Inflammatory response in 85 patients with loosened hip prostheses: a prospective study comparing inflammatory markers in patients with aseptic and septic prosthetic loosening. *Acta Orthop* 78(5): 629–639
31. Wirtz DC, Heller KD, Miltner O et al (2000) Interleukin-6: a potential inflammatory marker after total joint replacement. *Int Orthop* 24(4):194–196
32. Waddell J, Pritzker KP, Boynton EL (2005) Increased cytokine secretion in patients with failed implants compared with patients with primary implants. *Clin Orthop Relat Res* 434(434):170–176
33. Ahmad SS, Shaker A, Saffarini M et al (2016) Accuracy of diagnostic tests for prosthetic joint infection: a systematic review. *Knee Surg Sports Traumatol Arthrosc* 24(10):3064–3074
34. Parvizi J, Zmistowski B, Berbari EF et al (2011) New definition for periprosthetic joint infection: from the workgroup of the musculoskeletal infection society. *Clin Orthop Relat Res* 469(11):2992–2994
35. Fischer JE, Bachmann LM, Jaeschke R (2003) A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med* 29(7):1043–1051