pISSN: 2093-940X, eISSN: 2233-4718
Journal of Rheumatic Diseases Vol. 25, No. 1, January, 2018
https://doi.org/10.4078/jrd.2018.25.1.3

Review Article

Overview of the Process of Conducting Meta-analyses of the Diagnostic Test Accuracy

Young Ho Lee
Department of Rheumatology, Korea University College of Medicine, Seoul, Korea

Diagnosis is a critical step for clinical treatment. Many individual studies have been conducted to determine the accuracy of various diagnostic tests, but they had small sample sizes and correspondingly inadequate statistical strength. Combining the results from several such studies can help increase the statistical strength and precision of their results. Meta-analysis is a useful tool for evaluating the accuracy of diagnostic tests and can be used to obtain precise estimates when multiple small studies for a given test and subject pool are available. The need for meta-analysis on studies examining diagnostic test accuracy has increased noticeably, and more meta-analyses on diagnostic test accuracy studies are being published. A meta-analysis of diagnostic test accuracy studies differs from a typical meta-analysis because diagnostic test accuracy studies report a pair of statistics, such as sensitivity and specificity, rather than a single statistic. Therefore, meta-analyses of the diagnostic test accuracy need to deal with two summary statistics simultaneously. More complex statistical methods are required for conducting meta-analyses using diagnostic test accuracy studies compared to that required for conventional meta-analysis. This is because the sensitivity and specificity are generally inversely correlated due to a threshold effect, and there is considerable heterogeneity in the results of test accuracy studies. This review provides an overview of the process of meta-analysis of the diagnostic test accuracy. **(J Rheum Dis 2018;25:3-10)**

Key Words. Diagnostic test, Accuracy, Meta-analysis

INTRODUCTION

Diagnostic tests have been used to identify the presence or absence of a disease in a patient for the purpose of the treatment. Accurate diagnosis is therefore the cornerstone of good clinical care and provides the basis for proper treatments. Meta-analysis is a statistical technique used to combining results from different studies on the same topic and is becoming a popular method for resolving discrepancies in results regarding diagnostic test accuracy [1,2]. The basic principle of a meta-analysis is that the limitations of individual studies, such as small sample sizes and correspondingly inadequate statistical strength, can be overcome by combining the results from several studies to increase the statistical strength and precision in estimating effects [1]. Therefore, meta-analysis of studies examining the accuracy of diagnostic tests can provide more precise assessments when multiple small studies addressing the same test and patients are available [3]. This technique also examines the discrepancies in the results of different studies by addressing between-study heterogeneity, thus providing a more precise measurement of diagnostic test accuracy [4].

The accuracy of a diagnostic test is a measure of how well the test discriminates between patients with and those without the target condition [5]. Studies examining the accuracy of diagnostic tests report a pair of statistics, such as sensitivity and specificity, rather than a single one. Therefore, a meta-analysis of diagnostic test accuracy studies differs from typical meta-analyses, because the method used for conducting a meta-analysis for diagnostic test accuracy deals with two summary statistics si-

Received : August 17, 2017, **Revised :** September 13, 2017, **Accepted :** September 16, 2017

Corresponding to : Young Ho Lee, Department of Rheumatology, Korea University College of Medicine, 73 Inchon-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail : lyhcgh@korea.ac.kr

multaneously, and considerable heterogeneity in the results of the test accuracy studies is common [6]. Since sensitivity and specificity are generally inversely correlated and may be affected by a threshold effect, more complex statistical methods are required for conducting meta-analyses of diagnostic test accuracy studies [7].

The need for such meta-analyses has obviously increased recently, and more meta-analyses of diagnostic test accuracy studies are being published. In this review, we aimed to describe an overview of the process of conducting a meta-analysis using studies examining the accuracy of diagnostic tests for providing guidance for conducting and understanding meta-analysis of diagnostic test accuracy in the future.

MAIN SUBJECTS

Diagnostic test accuracy

Diagnostic test accuracy refers to the ability of a test to distinguish between patients with the disease and those without. Diagnostic test accuracy studies compare the accuracy of the diagnostic test of interest (the 'index test') to that of an existing diagnostic test (the 'reference test'); thus, the index test(s) is the diagnostic test whose accuracy is being investigated and the reference test is the 'gold standard' test to which the results of the index test will be compared [8].

The accuracy of a diagnostic test can be measured in a number of ways and is commonly reported using several statistical parameters: sensitivity and specificity, positive and negative predictive values (PPV and NPV), positive and negative likelihood ratios (PLR and NLR), diagnostic odds ratio (DOR), or receiver operating characteristic (ROC) curve [5,9]. The sensitivity of a test refers to the probability that the index test result will be positive in a patient, while the specificity is the probability that the index test result will be negative in a control [10]. In other words, sensitivity is the probability that a person with the condition of interest will have a positive result, while specificity is the probability of a person without the condition having a negative result. The PPV is the probability that a patient with a positive test result is diseased, while the NPV is the probability that a control with a negative test result is non-diseased. Likelihood ratio (LR) assesses the probability or likelihood that the test result obtained would be expected in patients with disease than in those without disease. PLR is considered as the best indicator for ruling-in the diagnosis: the higher the PLR, the more indicative of a disease is the test. NLR is a good indicator for ruling-out the diagnosis [11]. Diagnostic evidence based on PLRs and NLRs is defined as follows: PLRs >10 and NLRs <0.1, conclusive diagnostic evidence; PLRs >5 and NLRs <0.2, strong diagnostic evidence; PLRs of $2\sim5$ and NLRs of $0.2\sim0.5$, weak diagnostic evidence; and PLRs of $1\sim2$ and NLRs $0.5\sim1$, negligible evidence [11]. The DOR summarizes the diagnostic accuracy of the test with a single number that describes how many times higher the odds of obtaining a positive result are in a diseased patient relative to that in a non-diseased control. However, its use is limited because it cannot be used directly in clinical practice [12].

A ROC curve is useful for evaluating the performance of a diagnostic test accuracy [13]. The ROC curve represents the relationship between the sensitivity and specificity of the test at various thresholds. The ROC curve is obtained by varying the positivity threshold across all possible values and plotting the sensitivity (true positive rate) against 1-specificity (false positive rate) [13]. The summary ROC (sROC) curve is the estimate of an ordinary ROC curve adjusted for the study outcomes in the ROC space [10] and displays the results of individual studies in the ROC space. The sROC curve is recommended for evaluating the accuracy of a diagnostic test based on data from a meta-analysis [14]. The area under the curve (AUC) and the index Q are useful summaries of the curve [15]. The AUC is the probability that a diseased individual will have a higher test result than a non-diseased individual for a randomly selected pair of individuals, where the test result is 1 for a perfect test and 0.5 for a completely uninformative test [15]. The AUC can also be interpreted as the average sensitivity of the test taken over all specificity values (or, equivalently, as the average specificity over all sensitivity values) [13]. An AUC of >0.9, $0.7\sim0.9$, and $0.5\sim0.7$ can be regarded as a high, moderate, and low diagnostic accuracy, respectively (or $0.9\sim1$=excellent, $0.8\sim0.9$=very good, $0.7\sim0.8$=good, $0.6\sim0.7$=sufficient, $0.5\sim0.6$=bad) [16]. The Q* index is another useful global estimate of test accuracy for comparing sROC curves and is defined at the point where sensitivity equals specificity on a sROC curve [13]. A Q* value of 1.0 indicates 100% accuracy (i.e., sensitivity and specificity of 100%) [13].

Method of conducting a meta-analysis of diagnostic test accuracy

1) Step 1. Searching for heterogeneity

Heterogeneity between studies is especially common in meta-analyses on diagnostic test accuracy, due to differences in study populations and testing procedures. Before statistically pooling the data from the included studies, between-study heterogeneity should be tested [17]. The heterogeneity test examines the null hypothesis, i.e., that there are no differences between the findings of primary studies. Cochran's Q test is used to determine whether variations between primary studies represent true differences or are due to chance [18], and it is calculated by summing the squared deviation of each study's estimate from the overall estimate and then comparing it with the chi-squared distribution for $\kappa-1$ degrees of freedom (df), where κ is the number of studies [18]. Due to the low statistical strength of Cochran's Q test, a p-value <0.10 (not 0.05) is considered to indicates the presence of heterogeneity [19]. Another common indicator of heterogeneity is the I^2 value, which quantifies the effect of heterogeneity and does not depend on the number of studies or the type of outcome data. I^2 values range from 0 to 100% and represent the proportion of inter-study variability that can be attributed to heterogeneity rather than to chance ($I^2=100\%\times[Q-df]/Q$) [20]. I^2 values of 25%, 50%, and 75% are interpreted as low, moderate, and high estimates, respectively. However, Cochrane Q test results or I^2 statistics alone may not provide complete informative, as they do not account for heterogeneity due to threshold effects.

2) Step 2. Testing the threshold effect

Sensitivity and specificity of the studies are combined in an integrated value of all studies (pooling) by the weighted mean (by sample size or inverse variance of each study). However, this is often inappropriate due to the difference of threshold of the index test, because there is a relationship between the cut-off point and the sensitivity and specificity. Increasing the threshold increases the specificity, but decreases the sensitivity [21]. Different studies may use different cut-off points that influence the estimation of summary points, and variations in the diagnostic test accuracy may be partly due to such variations in cut-off points. When a threshold effect exists, there is a *negative* correlation between sensitivities and specificities, which leads to a typical pattern of "shoulder arm" plot in a sROC space [22]. Spearman's correlation coefficient between the sensitivity and specificity of all studies can test for the presence of a threshold effect [22]. Spearman's correlation coefficient $r\geq0.6$ generally indicates a threshold effect [12].

3) Step 3. Deciding the model for statistical pooling

Meta-analysis combines the effect sizes of the included studies by weighting the data according to the sample size and variability within each study. The choice of statistical method for meta-analysis depends on the heterogeneity observed in the results [11]. The fixed effect model assumes that genetic factors have similar effects on disease susceptibility in all the studies and that the observed variations are caused by chance alone [23]. The random effects model assumes that different studies exhibit substantial diversity and assesses both intra-study sampling errors and inter-study variances [24]. In the absence of heterogeneity, a fixed effects model is used for meta-analysis. When a significant Q value ($p<0.10$) is calculated, indicating the existence of heterogeneity in the studies, a random effects model is applied for the meta-analysis. Both models offer similar results for homogeneous study groups; however, if heterogeneity is present, the random effects model usually provides wider confidence intervals than the fixed effects model.

The paired nature of diagnostic test accuracy data makes meta-analysis complicated. Due to the threshold effect, sensitivity and specificity are expected to be heterogeneous. Meta-analyses of studies reporting pairs of sensitivity and specificity estimates have often used the linear regression model for the construction of sROC curves, as proposed by Moses et al. [22]. The Moses-Littenberg method is the most commonly used simple model for deriving a sROC in meta-analysis of diagnostic tests. However, this is a kind of fixed-effects model, since it does not consider the heterogeneity between studies. Thus, it is usually used for exploratory purposes. More advanced methods such as the hierarchical sROC model and bivariate analysis have been proposed to overcome the limitations of the Moses-Littenberg model [7,25]. These models account for both within- and between-study heterogeneity [7,25].

4) Step 4. Dealing with heterogeneity

It is also important to determine the possible causes of heterogeneity when heterogeneity exists in the studies included in the meta-analysis. The heterogeneity can be explained by analyzing study subgroups or by meta-regression. Subgroup analysis performs meta-analysis based

on important potential confounders such as patient characteristics, test methods, and study design, and an assessment is made to determine how much the factors affect the test accuracy. Subgroup analysis can detect homogeneous subgroups with respect to important potential confounders. Meta-regression is a regression analysis that explores possible factors contributing to heterogeneity [4]. The DOR is normally used for meta-regression, as it is a unitary measure of diagnostic performance that encompasses sensitivity and specificity or PLR and NLR [13].

An example of a meta-analysis of diagnostic test accuracy

Monosodium urate crystals precipitate on the articular surface of hyaline cartilage and this precipitate forms a hyperechoic, bright band that parallels the hyperechoic bony cortex, forming a "double contour" of hyperechoic bone and bright-appearing monosodium urate deposits, as visualized on ultrasound images [26]. We performed a meta-analysis using published data of the sensitivity and specificity of ultrasound according to the double-contour sign for the diagnosis of gout in order to assess the diagnostic capability of ultrasound. Studies were selected for the analysis if they included (i) case-control, cross-sectional, or cohort studies, (ii) sufficient data to calculate the sensitivity and specificity of US according to the double contour sign, and (iii) patients with gout diagnosed on the basis of the classification criteria [27,28] or the demonstration of monosodium urate crystals in a joint aspirate. Within- and between-study variations and heterogeneities were assessed using Cochran's Q-statistic and I^2 value, respectively. We used a random effects model to combine the sensitivity, specificity, PLR, NLR, and DOR estimates due to heterogeneity and analyzed sROC curves and Q* index. To examine the potential source of heterogeneity observed in the meta-analysis, subgroup analysis and meta-regression were performed with the following covariates: (i) sample size, (ii) study design, and (iii) diagnostic criteria. Statistical manipulations for

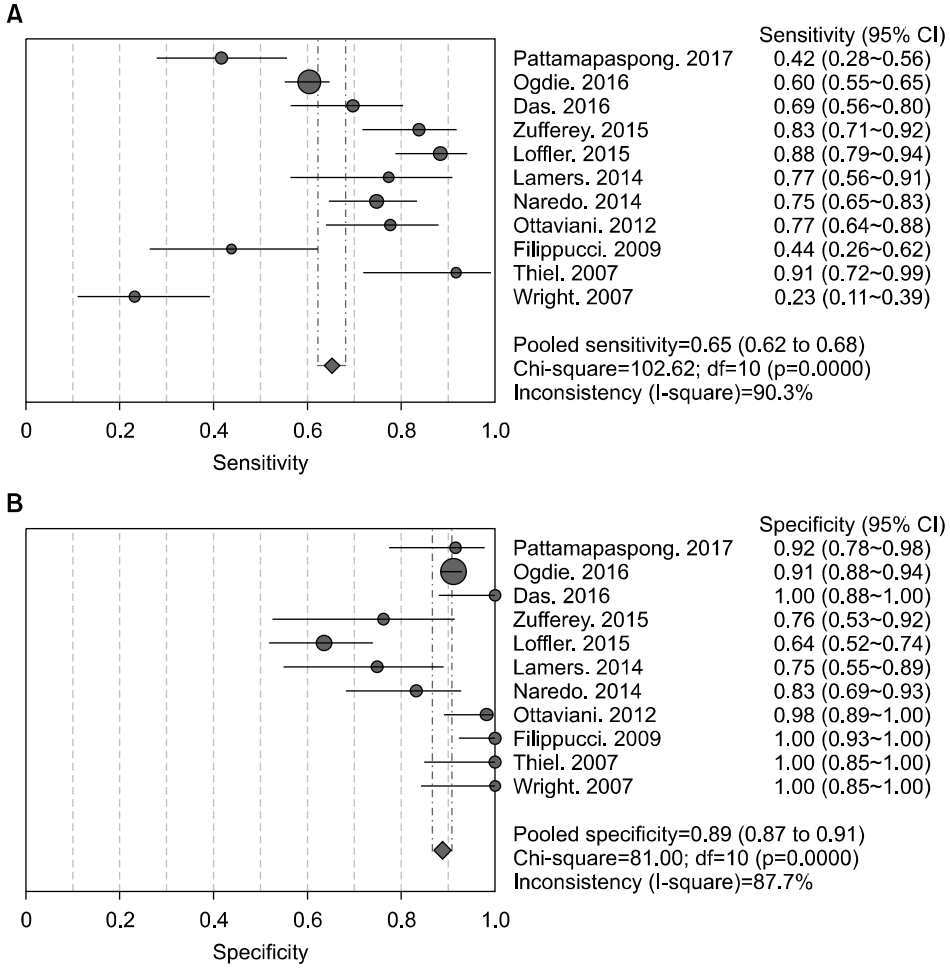

Figure 1. (A) Sensitivity and (B) specificity estimates for ultrasound used for the diagnosis of gout. Circles and lines represent point estimates and 95% confidence intervals (CI), respectively. Circled areas represent relative study sizes. Df: degrees of freedom.

Table 1. Summarized results of the meta-analysis

Subgroup	Population	Study number	Number		Sensitivity (95% CI)	Specificity (95% CI)	PLR (95% CI)	NLR (95% CI)	DOR (95% CI)
			Gout	Control					
All combined	Overall	11	978	788	0.651 (0.620～0.682)	0.890 (0.866～0.911)	5.889 (3.365～10.30)	0.359 (0.266～0.485)	17.61 (11.11～17.92)
Subject number	＞100	4	643	580	0.672 (0.634～0.708)	0.876 (0.846～0.902)	5.503 (2.354～12.86)	0.293 (0.194～0.442)	17.21 (9.924～29.87)
	＜100	7	295	208	0.607 (0.548～0.663)	0.928 (0.884～0.959)	6.783 (2.911～15.80)	0.399 (0.256～0.621)	20.70 (8.290～51.69)
Study design	Prospective	8	779	649	0.635 (0.601～0.669)	0.915 (0.891～0.936)	6.089 (3.563～10.40)	0.380 (0.275～0.525)	18.60 (11.35～30.18)
	Retrospective	3	159	139	0.730 (0.653～0.797)	0.770 (0.691～0.837)	4.918 (1.332～18.15)	0.248 (0.259～1.038)	17.70 (4.450～70.45)
Diagnostic criteria	MSU	9	853	704	0.674 (0.401～0.705)	0.881 (0.854～0.904)	5.585 (3.054～10.21)	0.307 (0.211～0.446)	18.47 (11.26～30.30)
	ACR	2	85	84	0.424 (0.317～0.536)	0.424 (0.317～0.536)	10.35 (1.236～86.74)	0.608 (0.502～0.737)	17.17 (1.910～154.49)

CI: confidence interval, PLR: positive likelihood ratio, NLR: negative likelihood ratio, DOR: diagnostic odds ratio, MSU: monosodium urate, ACR: American College of Rheumatology.

Table 2. Characteristics of the individual studies included in the meta-analysis

Study	Country	Gout	Control	Gout		Control	Diagnostic criteria	Study design	Scanned joints	Sensitivity (%)	Specificity (%)
				Duration (yr)	Age (yr) (mean±SD)	Age (yr) (mean±SD)					
Pattamapaspong, 2017 [30]	Thailand	53	36	NA	65.4±10.6	64.7±16.8	ACR	Retrospective	Selected joints	42	92
Ogdie, 2017 [31]	Multi-national	416	408	58.48±36.4*	60.2±14.6	59.5±16.0	MSU	Prospective	Affected joints	60.1	91.4
Das, 2017 [32]	India	62	30	NA	49.1±9.1	47.6±10.6	MSU	Prospective	1st MTP, knee	69.4	100
Zufferey, 2015 [33]	Switzerland	60	21	NA	65±12	67±10	MSU	Prospective	1st MTP, ankle, knee	84	78
Löffler, 2015 [34]	Germany	83	80	NA	69±12	76±11	MSU	Retrospective	Affected joints	87.8	64.1
Lamers-Karnebeek, 2014 [35]	Netherlands	26	28	NA	NA	NA	MSU	Prospective	1st MTP, ankle, knee	77	75
Naredo, 2014 [36]	Spain	91	42	NA	56.4±11.5	56.6±13.5	MSU	Prospective	26 joints	75	83
Ottaviani, 2012 [37]	France	53	50	9.2±10.7	59.7±15.8	59.5±15.3	MSU	Prospective	MTP, knee, MCP	77	98
Filippucci, 2009 [38]	Italy	32	48	NA	65±11.6	66±13.6	ACR	Prospective	Knee	44	99
Thiele, 2007 [39]	USA	23	23	NA	NA	NA	MSU	Retrospective	MTP, ankle, knee, MCP	92	100
Wright, 2007 [40]	UK	39	22	12±8	52±11	53±16	MSU	Prospective	1st MTP	22	100

NA: not available, SD: standard deviation, ACR: American College of Rheumatology, MSU: monosodium urate, MTP: metatarsophalangeal joint, MCP: metacarpophalangeal joint.
*Months.

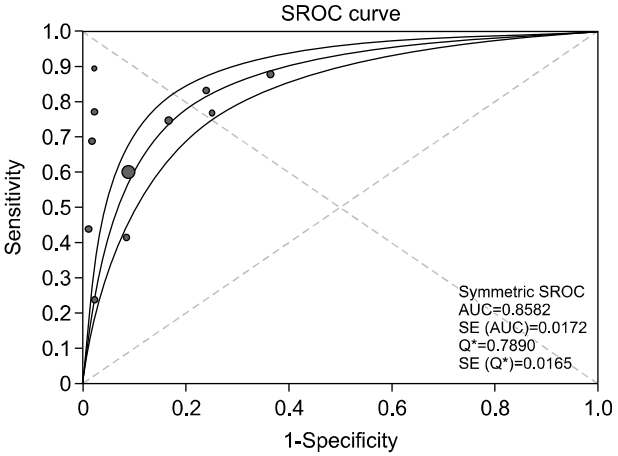

Figure 2. Summary receiver-operating characteristic curves for ultrasound for the diagnosis of gout. Solid circles represent the individual studies included in this meta-analysis. The curve shown is a regression line that summarizes the overall diagnostic accuracy. SE (AUC): standard error of the area under the curve, Q*: an index defined by the point on the summary receiver operating characteristic (sROC) curve, where the sensitivity and specificity are equal, SE (Q*): Q* index standard error.

this meta-analysis were performed using Meta-DiSc, version 1.4 (Unit of Clinical Biostatistics team, Hospital Universitario Ramón y Cajal, Madrid, Spain) [29].

Eleven studies including 938 patients with gout and 788 controls (patients with non-gout inflammatory arthritis) were included in the meta-analysis (Figure 1). The pooled sensitivity and specificity of US were 65.1% (95% confidence interval [CI], 62.0～68.2) and 89.0% (95% CI, 96.6～91.1), respectively (Table 1, Figure 1). The PLR, NLR, and DOR were 5.889 (3.365～10.30), 0.359 (0.266～0.485), and 17.61 (11.11～17.92), respectively (Table 2) [30-40]. The AUC of US was 0.858 and the Q* index was 0.789, indicating good diagnostic accuracy (Table 1, Figure 2). Some between-study heterogeneity was found in the meta-analysis. Meta-regression showed that the sample size, study design, and diagnostic criteria were not sources of heterogeneity. A similar pattern was found in the subgroup analysis according to diagnostic criteria (Table 1). In conclusion, the meta-analysis indicated that US offers very good diagnostic accuracy and can play an important role in the diagnosis of gout.

CONCLUSION

Meta-analysis is a useful tool for summarizing research on diagnostic test accuracy by combining data from multiple studies using statistical techniques, thus increasing the precision and statistical power of the evaluations of diagnostic test accuracy in the primary research. It is therefore necessary for clinicians to be able to understand the results of meta-analyses of diagnostic test accuracy studies. Such meta-analyses differ from typical meta-analysis, because meta-analysis of diagnostic test accuracy studies need to deal with two summary statistics simultaneously. Since sensitivity and specificity are generally inversely correlated due to a threshold effect, and considerable heterogeneity in the results of test accuracy studies is to be expected, more complex statistical methods are required for conducting meta-analyses of diagnostic test accuracy studies. This review provides an overview of the process of a meta-analysis conducted using studies investigating the accuracy of diagnostic tests to help in conducting and understanding meta-analysis of diagnostic test accuracy in the future.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Gotzsche PC. Why we need a broad perspective on meta-analysis. It may be crucially important for patients. BMJ 2000;321:585-6.
2. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995;48:119-30.
3. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997;315:1533-7.
4. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? Stat Med 2002; 21:1559-73.
5. Šimundić AM. Measures of diagnostic accuracy: basic definitions. EJIFCC 2009;19:203-11.
6. Harbord RM, Whiting P, Sterne JA, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. J Clin Epidemiol 2008;61:1095-103.
7. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982-90.
8. Campbell JM, Klugar M, Ding S, Carmody DP, Hakonsen SJ, Jadotte YT, et al. The joanna briggs institute reviewers' manual 2015. The systematic review of studies of diagnostic test accuracy. Adelaide, The Joanna Briggs Institute, 2015, 46 p.
9. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. Clin Microbiol Infect 2014;20:105-

13.
10. Deeks J, Bossuyt P, Gatsonis C. Cochrane handbook for systematic reviews of diagnostic test accuracy version 1.0. 0. Oxford, Cochrane Collaboration, 2009.
11. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ 2001;323:157-62.
12. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol 2003;56:1129-35.
13. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
14. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess 2005;9:1-113, iii.
15. Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. Stat Med 2002;21:1237-56.
16. Okeh UM, Okoro CN. Evaluating measures of indicators of diagnostic test performance: fundamental meanings and formulars. J Biom Biostat 2012;3:132.
17. Trikalinos TA, Salanti G, Zintzaras E, Ioannidis JP. Meta-analysis methods. Adv Genet 2008;60:311-34.
18. Whitehead A, Whitehead J. A general parametric approach to the meta-analysis of randomized clinical trials. Stat Med 1991;10:1665-77.
19. Munafò MR, Flint J. Meta-analysis of genetic association studies. Trends Genet 2004;20:439-44.
20. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-58.
21. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008; 149:889-97.
22. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12:1293-316.
23. Davey Smith G, Egger M. Meta-analyses of randomised controlled trials. Lancet 1997;350:1182.
24. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
25. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 2001;20:2865-84.
26. Terslev L, Gutierrez M, Schmidt WA, Keen HI, Filippucci E, Kane D, et al. Ultrasound as an outcome measure in gout. A validation process by the OMERACT Ultrasound Working Group. J Rheumatol 2015;42:2177-81.
27. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 Gout Classification Criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015;67:2557-68.
28. Wallace SL, Robinson H, Masi AT, Decker JL, McCarty DJ, Yü TF. Preliminary criteria for the classification of the acute arthritis of primary gout. Arthritis Rheum 1977;20:895-900.
29. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31.
30. Pattamapaspong N, Vuthiwong W, Kanthawang T, Louthrenoo W. Value of ultrasonography in the diagnosis of gout in patients presenting with acute arthritis. Skeletal Radiol 2017;46:759-67.
31. Ogdie A, Taylor WJ, Neogi T, Fransen J, Jansen TL, Schumacher HR, et al. Performance of ultrasound in the diagnosis of gout in a multicenter study: comparison with monosodium urate monohydrate crystal analysis as the gold standard. Arthritis Rheumatol 2017;69:429-38.
32. Das S, Ghosh A, Ghosh P, Lahiri D, Sinhamahapatra P, Basu K. Sensitivity and specificity of ultrasonographic features of gout in intercritical and chronic phase. Int J Rheum Dis 2017;20:887-93.
33. Zufferey P, Valcov R, Fabreguet I, Dumusc A, Omoumi P, So A. A prospective evaluation of ultrasound as a diagnostic tool in acute microcrystalline arthritis. Arthritis Res Ther 2015;17:188.
34. Löffler C, Sattler H, Peters L, Löffler U, Uppenkamp M, Bergner R. Distinguishing gouty arthritis from calcium pyrophosphate disease and other arthritides. J Rheumatol 2015;42:513-20.
35. Lamers-Karnebeek FB, Van Riel PL, Jansen TL. Additive value for ultrasonographic signal in a screening algorithm for patients presenting with acute mono-/oligoarthritis in whom gout is suspected. Clin Rheumatol 2014;33:555-9.
36. Naredo E, Uson J, Jiménez-Palop M, Martínez A, Vicente E, Brito E, et al. Ultrasound-detected musculoskeletal urate crystal deposition: which joints and what findings should be assessed for diagnosing gout? Ann Rheum Dis 2014;73: 1522-8.
37. Ottaviani S, Richette P, Allard A, Ora J, Bardin T. Ultrasonography in gout: a case-control study. Clin Exp Rheumatol 2012;30:499-504.
38. Filippucci E, Riveros MG, Georgescu D, Salaffi F, Grassi W. Hyaline cartilage involvement in patients with gout and calcium pyrophosphate deposition disease. An ultrasound study. Osteoarthritis Cartilage 2009;17:178-81.
39. Thiele RG, Schlesinger N. Diagnosis of gout by ultrasound. Rheumatology (Oxford) 2007;46:1116-21.
40. Wright SA, Filippucci E, McVeigh C, Grey A, McCarron M, Grassi W, et al. High-resolution ultrasonography of the first metatarsal phalangeal joint in gout: a controlled study. Ann Rheum Dis 2007;66:859-64.