BMJ Open

Diagnostic Value of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033084
Article Type:	Research
Date Submitted by the Author:	20-Jul-2019
Complete List of Authors:	Ni, Jin-Rong; The First Hospital of Lanzhou University Yan, Pei-Jing; Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital Liu, Shi-Dong; The First Hospital of Lanzhou University Hu, Yuan; the First Hospital of Lanzhou University Yang, Kehu; evidence based meidicine center, lanzhou university Lei, Jun-Qiang; The First Hospital of Lanzhou University Song, Bing; The First Hospital of Lanzhou University
Keywords:	Hypertension < CARDIOLOGY, Echocardiography < CARDIOLOGY, Diagnostic radiology < RADIOLOGY & IMAGING

Diagnostic Value of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Jin-Rong Ni[1,3,8,9], Pei-Jing Yan[2,5,6#], Shi-Dong Liu[1,3], Yuan Hu[3], Ke-Hu Yang[2,4,5,6], Jun-Qiang Lei[1,7,8,9*], Bing Song[3*]

[1]The First Hospital (the First Clinical Medical School) of Lanzhou University, Lanzhou 730000, China; [2]Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital, Lanzhou, 730000, China; [3]Department of Cardiovascular Surgery, the First Hospital of Lanzhou University, Lanzhou 730000, China; [4]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China;[5]Evidence Based Social Science Research Center, Lanzhou University, Lanzhou, 730000, China; [6]Key Laboratory of Evidence-based Medicine and Knowledge Translation of Gansu Province, Lanzhou, 730000, China.[7]Department of Radiology, the First Hospital of Lanzhou University, Lanzhou 730000, China. [8]Intelligent Imaging Medical Engineering Research Center of Gansu province, Lanzhou 730000, China; [9]Precision Image and Collaborative Innovation International Scientific and Technological Cooperation Base of Gansu province, Lanzhou 730000, China.

[#]Joint first author

[*]Joint corresponding authors

Corresponding Author: 1. Jun-Qiang Lei, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613919289040

E-mail: leijunqiangldyy@163.com

2. Bing Song, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613993118707

E-mail: songbingldyyxwk@163.com

Funding sources: This study was supported by the Key Laboratory of Evidence Based Medicine and Knowledge Translation Foundation of Gansu Province (Grant No. GSXZYZH2018006).

Statement on potential conflicts of interest: None declared.

Patient and Public Involvement: There is no patient or the public involved in our work.

Data Availability: All data relevant to the study are included in the article.

Key Words: Echocardiography; Pulmonary Hypertension; Catheterization; Diagnosis; Accuracy; Meta-analysis.

Abstract

Objective: To evaluate the diagnostic value of TTE in patients with PH and preliminarily explore the factors that may affect the diagnostic accuracy of TTE.

Design: Systematic review and meta-analysis.

Data sources and eligibility criteria: EMBASE, Cochrane Library for clinical trials, PubMed and Web of Science were searched from inception to June 19, 2019. Studies using both TTE and RHC to diagnose PH were included.

Mean results: 27 studies involving 4386 subjects were eligible for analysis. TTE had a pooled sensitivity 85% (95% CI 81–90%), a pooled specificity of 74% (95% CI 64–81%), a pooled positive likelihood ratio of 3.2 (95% CI 2.3–4.4), a pooled negative likelihood ratio of 0.20 (95% CI 0.15–0.26), a pooled diagnostic odds ratio of 16 (95% CI 10–27), and finally an area under the SROC curve of 0.88 (95%CI 0.85–0.90).

Conclusion: The value of TTE in diagnosing PH is certain, although it cannot yet replace RHC as the gold standard at this stage. The time interval between TTE and RHC, the threshold value of TTE and the disease composition of the study population may all be factors affecting the diagnostic value of TTE.

Review registration number: PROSPERO CRD42019123289.

Strengths and limitations of this study

1. We conducted a systematic and comprehensive search of the main database, included more studies, and obtained a large sample size.

2. Detailed subgroup analysis and sensitivity analysis were performed.

3. The types of pulmonary hypertension included in the studies could not be distinguished.

4. No unified procedure for measuring pulmonary artery pressure by transthoracic echocardiography.

Introduction

The prevalence of pulmonary hypertension (PH) is estimated 1% in the general population, and as high as 10% in the 600 million people older than 65[1] . With the aggravation of population aging, PH will become a global health problem[2]. Early detection and accurate assessment is vital for improved outcomes for PH patients[3]. Right heart catheterization (RHC) is the gold standard for accurate measurements of pulmonary pressures and for the diagnosis of PH[4]. But RHC is invasive and it cannot be used frequently or repeatedly[5]. Transthoracic echocardiography (TTE) is a noninvasive test recommended for use in screening for PH[4]. The feasibility of TTE in evaluating systolic pulmonary artery pressure (SPAP) through tricuspid regurgitation has been confirmed in previous studies[6 7]. TTE is also able to provide crucial information on heart size and function[8].

Despite the frequent use of TTE for screening for PH, its diagnostic accuracy and clinical value has courted much controversy. The latest guideline[4] for PH suggest that peak tricuspid regurgitation velocity (TRV) can be used to determine the likelihood of PH. But clinicians often expect to get a specific numerical value by echocardiography to evaluate the condition, observe the curative effect and judge the prognosis. Therefore echocardiographic method for estimating SPAP by tricuspid regurgitation are still adopted in clinic.

From 2010 to 2013, three reviews[9-11] unanimously concluded that TTE could only be used as a crude screening tool and was not suitable for the diagnosis of PH. However, the studies they included were published before 2010. In recent years, many new original studies related to this topic have emerged. In addition, an ideal evidence system should integrate and evaluate all important research evidence related to specific clinical problems[12]. So high quality meta-

analysis has been increasingly regarded as one of the key tools for achieving evidence[13 14]. Thus, the purpose of this study is to undertake an updated systematic review and quantitative meta-analysis on the value of TTE for diagnosing PH and preliminarily explore the factors that may affect the diagnostic accuracy of TTE.

Methods

The present study is reported as per the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement[15] and the published recommendations[16]. The detailed protocol is accessible in PROSPERO (CRD42019123289) [17 18].

Data sources and search

We performed a systematic search in EMBASE, Cochrane Library, PubMed and Web of Science for relevant literatures from inception to June 19, 2019. Subject words were combined with free words, and the search strategies were developed and adapted for each database (appendix 1). For unpublished trials, we searched ClinicalTrials.gov and the trials registers on the World Health Organization International Clinical Trials Registry Platform. We also reviewed the references of included studies and other systematic reviews and meta-analysis to obtain a comprehensive list of included studies.

Study selection

Studies were selected based on the following inclusion criteria: studies diagnosing PH by TTE; the study population was patients with suspected PH; TTE measurements of SPAP were performed using tricuspid regurgitation; RHC was used as the gold standard for the diagnosis of PH.

Exclusion criteria were: insufficient data to construct a 2×2 table; the tricuspid regurgitation method was not used to calculate pulmonary artery pressure; studies with fewer than 20 subjects; duplicate data is used (in which case, select the largest sample or the latest study).

Two reviewers (JR.N and PJ.Y) independently screened eligible studies for suitability. Disagreements were resolved by consensus. If consensus could not be reached, a third reviewer (SD.L) was deferred to for arbitration. No language restriction was applied.

Data extraction

Two reviewers (JR.N and PJ.Y) extracted data independently as per a predefined data extraction sheet. The following variables were extracted from included studies: lead author, publication year, country of study, study design, study population demographics, sample size, mean age, male ratio, the time interval between TTE and RHC, the cut-off threshold levels for TTE and RHC, and the number of true-positive (TP), false-negative (FN), true-negative (TN) and false-positive (FP) observations. Extracted data was cross-checked and disagreements were resolved via discussion or referral to a third reviewer (Y.H).

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies QUADAS-2 tool was used to assess the risk of bias and clinical applicability concerns of included studies as per the Cochrane Collaboration recommendation[19 20]. Two reviewers (JR.N and PJ.Y) independently evaluated QUADAS-2 items, and all emerging conflicts were resolved by consensus.

Data synthesis and statistical analysis

All statistical analyses were performed with STATA/SE version 15.1 (Stata Corp, College

Station, TX) and Review Manager Version 5.3 software (Copenhagen, Denmark, Nordic Cochrane Centre, Cochrane Collaboration, 2014). All tests were two-tailed. A *P* value< 0.05 was considered statistically significant.

The correlation coefficient between the logarithm of sensitivity and logarithm of one minus specificity (i.e. the false-positive rate) was calculated to test whether the threshold effect was one of the sources of heterogeneity[21]. Deeks' test was used to test for publication bias[22]. The bivariate model for diagnostic meta-analysis was used to obtain pooled estimates of sensitivity and specificity[23]. Statistical heterogeneity among studies was explored using the I^2 statistic.

Pooled sensitivity, specificity, diagnostic odds ratio (DOR), positive (PLR) 、 negative (PLR) likelihood ratios, and the area under the summary receiver operating curve (SROC) were calculated from the number of TPs, FNs, FPs, and TNs. The 95% confidence intervals (CIs) were estimated for each metric.

Subgroup analyses were undertaken based on the following variables: the time interval between TTE and RHC; disease classification of the study population; publication year of the study; study design (prospective or retrospective studies) and cut-off threshold of TTE diagnostic PH. Sensitivity analysis was undertaken by excluding low-quality studies(according to the QUADAS-2 quality assessment) or trials with characteristics different from the others.

Results

Studies Retrieved and Characteristics

Figure 1 presents the PRISMA flow chart of literature screening. A total 27 publications[7 24-49] involving 4386 subjects met our inclusion criteria (Table 1). One study[28] was divided into two independent parts because of the differences between the case group and the control group.

Of the 27 eligible studies, 14 (52%) were published during 2010–2019, and 13 (48%) were published before 2010. 12 (44%) studies were undertaken in the Europe, nine (30%) in the USA, two in the Asia (8%), three in the Middle East (12%) and one in the Australia (4%). Most studies were prospective in design (56%, 15/27) versus 44% (12/27) retrospective.

All included studies used the tricuspid maximal regurgitation velocity (TRVmax) to estimate SPAP; the majority of these studies using classical methods ($4TRVmax^2+RAP$) to calculate SPAP. The right atrial pressure (RAP) was calculated through the diameter and collapse of the inferior vena cava (IVC) during spontaneous respiration in 16(59%) studies, through the jugular vein pressure (JVP) in one study (4%), and using a fixed value (5 or 10 mm Hg) in three studies (11%). Three studies (11%) did not report their methods for calculating RAP. Four studies (15%) used a tricuspid gradient ($4TRVmax^2$) instead of SPAP.

The majority of studies (81%) reported the time interval (mean or maximum) between TTE and RHC while five (19%) did not. 10 studies (37%) had time intervals greater than one week, 13 studies (48%) had time intervals of less than one week. The time interval between TTE and RHC ranged from four hours to three months.

Quality Assessment

The quality assessment of included studies as per the QUADAS-2 inventory is presented in Figure 2. In 20 (74%) study protocols, consecutive subjects were enrolled, with no inappropriate exclusions. The risk of bias during patient recruitment was unclear in the remaining seven (26%) studies, as patient recruitment was not reported. In seven (26%) studies investigators designed the single-blind methods for TTE or RHC. Double blinding in imaging assessment was not mentioned in any of the articles of the studies. Risk of bias on flow and

timing between the index test and reference standard was categorized as unclear in 14 (52%) study protocols that did not explicitly state successful investigation with both index and reference tests in all included patients. Overall, there were low concerns for patient selection, index test, and reference standard. Only five studies (19%) described the number of patients excluded for lack of tricuspid regurgitation, while the rest (81%) only indicated excluding patients without tricuspid regurgitation or using contrast media to enhance tricuspid regurgitation signals.

Quantitative Analysis

The SROC curve for TTE is presented in Figure 3.The AUC was 0.88 (95%CI 0.85–0.90). The pooled sensitivity and specificity for TTE were 85% (95%CI 81%–90%) and 74% (95%CI 64%–81%), respectively. There was significant heterogeneity across studies for the specificity and sensitivity of TTE (Figure 4). The pooled PLR and NLR were 3.2 (95%CI 2.3–4.4) and 0.20 (95%CI 0.15–0.26). The pooled DOR for TTE is 16 (95%CI 10–27).

The results of the subgroup analysis are presented in Table 2. The sensitivity (87%, 95%CI 81%–91%), specificity (74%, 95%CI 62%–83%) and AUC (0.89, 95%CI 0.86–0.91) of TTE for diagnosing PH was higher for studies published in 2010 and later relative to those published before 2010. Among the time interval subgroups, the group with the shortest time interval between TTE and RHC had the best diagnostic effect, with sensitivity specificity and AUC of 88% (95%CI 73%–95%), 90% (95%CI 53%–99%) and 0.94(95%CI 0.92–0.96), respectively. The disease composition of the study population also affected the diagnostic accuracy of TTE. Compared with patients of other diseases, TTE had lower sensitivity (81%, 95%CI 70%–88%), specificity (61%, 95%CI 53%–69%) and AUC (0.73, 95%CI 0.69–0.77) in the subgroup of

patients with definite lung diseases.

Subgroup analysis of different cut-off thresholds for diagnosing PH based on TTE showed that subgroup applied a cut-off threshold of 35mmHg had superior performance than at 40mmHg. The sensitivity, specificity and AUC of the former were respectively 92% (95%CI 88%–94%), 65% (95%CI 43%–83%) and 0.92 (95%CI 0.89–0.94), while the sensitivity, specificity and AUC at 40mmHg were 84% (95%CI 75%–91%), 52% (95%CI 31%–71%) and 0.80 (95%CI 76%–83%).Tricuspid regurgitation pressure gradient (TRPG) calculated by $4TRVmax^2$ has good specificity (81%, 95%CI 70%-89%) in the diagnosis of PH.

The sensitivity analysis results are shown in Table 3. After excluding low-quality studies and studies with specific characteristics, sensitivity analysis did not reveal a source for the heterogeneity for the diagnostic accuracy analysis. Overall, the pooled meta-analysis results for outcomes were robust to our sensitivity analyses.

Discussion

Our study found that TTE has better sensitivity but moderate specificity for the detection of PH. The pooled sensitivity of the TTE was 0.85 at a specificity of 0.74. The AUC of TTE was 0.88. These findings altogether suggest that TTE has clinical value for diagnosing PH. In addition, the time intervals between TTE and RHC, the cut-off threshold of TTE, the basic diseases of the tested patients and other factors may affect the accuracy of TTE to diagnose PH.

There was significant heterogeneity in our study, threshold test proves that threshold effect is not the source of heterogeneity (r=-0.34, *P*=0.12). From SROC diagram, we can see that only four studies fall within the 95% confidence contour. Deeks' test for funnel plot asymmetry suggested no publication bias (*P*=0.69). We speculate the source of heterogeneity to a

combination of variation, including study design, interval time between TTE and RHC, and biases in the review, natural history of PH and population. Despite the significant heterogeneity of the studies included in this paper, our results can still serve as the adamant evidence for the diagnosis of TTE with PH.

Guidelines recommend the use of inferior vena cava width and collapse rate to estimate RAP, which was not used in some of the included studies. Sensitivity analysis for this point showed that studies of using IVC to calculate RAP did not seem to have a higher diagnostic performance. In order to avoid errors caused by RAP estimation, TRVmax is also considered as an indicator to evaluate the possibility of PH. Four studies using TRPG replacing SPAP were grouped into a subgroup. The results showed that this subgroup had high diagnostic specificity, but the overall diagnostic effect was general.

At last year's 6th World Symposium on Pulmonary Hypertension, the new definition of PH surprised the audience and provoked lasting discussion. Scholars represented by Gérald Simonneau[50] pointed out that the normal mean pulmonary artery pressure (MPAP) was 14.0±3.3 mmHg, and two standard deviations higher than the mean value would be MPAP>20 mmHg should be defined as the upper limit of normal values (above 97.5 percentage points) based on scientific methods. However, some opponents[51] argue that it is too early to reduce the MPAP threshold to 20 mmHg because of the risk of over-diagnosis, unclear treatment implications and additional psychological burden on patients. Therefore, we still follow the previous standard that PH was defined as MPAP≥25mmHg at rest. After excluding 7 studies that RHC threshold was not 25mmHg, sensitivity analysis results showed that sensitivity decreased slightly, specificity increased slightly. It should be noted that we must face up to the

increased risk of misdiagnosis due to the reduction of the threshold value of PH diagnostic gold standard. More research is needed to determine whether lowering standards does more harm than good.

Studies included in previous systematic reviews and meta-analysis similar to this study all published before 2010, therefore we conducted subgroup analysis according to the year of publication. The data confirm that studies published after 2010 have a slightly higher diagnostic accuracy than previous studies. We surmised that this could be due to improvement of ultrasonic equipment, the enhancement of ultrasonic operation or clear diagnostic gold standard of RHC.

Does the time interval between the diagnostic test and the standard test affect the diagnostic accuracy? There is no clear consensus. Although PH is a chronic disease, we still believe that the shortest possible time interval is more conducive. Otherwise, changes in the patient's condition and the application of intervention measures will lead to an increase in the deviation of the results of the two examinations. Since most studies (25/27) did not mention whether the subjects were hospitalized patients, we conducted a detailed subgroup analysis according to the time interval between TTE and RHC. As expected, the diagnostic accuracy was highest when the time interval was less than or equal to 24 hours.

Analysis restricted to subgroups disease composition of the population suggested that the diagnostic accuracy of TTE is lower for patients with lung diseases. Changes associated with chronic pulmonary disease, including marked increase in intrathoracic gas, consolidation of lung tissue, expansion of the thoracic cage, and alterations in the position of the heart, adversely affect the imaging quality and parameters measurement of TTE[52]. Therefore, using TTE to

measure pulmonary pressure in patients with lung diseases may not be an ideal choice.

The cut-off threshold of the diagnostic test often affects its sensitivity and specificity. However, guidelines in Europe and the United States differ slightly in the threshold value for the echocardiographic diagnosis of PH. The American guideline advocates for a 40mmHg[53], while the European equivalent recommends 36mmHg[54]. In our review, four studies assessed PH using TRPG (4TRVmax^2) without adding RAP. In the remaining 22 studies, the cut-off thresholds of SPAP ranged from 30–50mmHg. Subgroup analysis showed that the diagnostic accuracy of the group of 35 mmHg was higher. Sensitivity analysis results of studies that excluded high TTE cut-off value showed that high cut-off value increased the specificity and reduced the sensitivity of TTE. Due to the small sample size of the subgroup in this study, the value of the cut-off threshold still needs to be determined by further prospective studies of multi-center and large samples.

There were two types of included literature, prospective studies and retrospective studies. Subgroup analysis presented that TTE had higher diagnostic value in prospective studies than in retrospective studies. A prospective study has predetermined details of experimental implementation, and the quality of research can be improved through multiple controllable links. It is more rigorous and scientific than a retrospective study[55].

The previous three meta-analyses have their own limitations. The studies of Mohammed Taleb[9] and Rui Feng Zhang[11] included few references and the sample size was relatively small. Although Surinder Janda et al[10] included 28 studies, only 13 of them had sufficient data for meta-analysis. Most importantly, they only synthesized the diagnostic accuracy data in their study, but did not conduct subgroup analysis. In this study, we comprehensively included more

literatures and conducted detailed subgroup analysis according to different research parameters to explore the factors that may affect the diagnostic accuracy of TTE. It was expected that more researchers related to PH will benefit from the results of our detailed subgroup analysis.

Limitations

Systematic review and meta-analysis is a secondary research method based on multiple original studies. The quality of the included studies will have an impact on the results of meta-analysis. In addition, there are several limitations in our study. Firstly, the studies included in this review involved several different types of PH, and some studies did not describe the basic disease and PH typing in detail. It is obvious that pulmonary lesions can affect the quality of TTE imaging, leading to underestimated results. Secondly, the blinding procedures were unreported for some of the included studies. Finally, in order to obtain more original studies, we did not specify the type of echocardiography equipment and specific requirements in the operation process. Differences in equipment and ultrasonic inspection implementation process could be a source of interference.

Conclusion

The value of TTE in diagnosing PH is certain, although it cannot yet replace RHC as the gold standard. The diagnostic accuracy of TTE improved with shortening time interval between TTE and RHC and appropriate cut-off threshold. TTE may not be suitable for assessing pulmonary arterial pressure in patients with pulmonary disease. More multicenter, large-sample randomized controlled trials are needed to verify our findings.

Author's contribution

The joint corresponding authors (JQ.L and B.S) are responsible for the design and implementation of the study. As a professional consultant, Y.H carried out quality control on the links of literature inclusion and data extraction. KH.Y provided guidance in literature retrieval and data processing methodology and was responsible for the quality evaluation part. JR.N and PJ.Y have done the systematic review of the literature and extracted data. JR.N has conducted the meta-analyses, and two authors (JR.N, PJ.Y) have substantially contributed to interpretation of data and co-authored the article. All the authors have made repeated revisions to the article. The corresponding authors (JQ.L and B.S) and JR.N take responsibility of the integrity of the analyses.

Reference

1. Hoeper MM, Humbert M, Souza R, et al. A global view of pulmonary hypertension. *The Lancet Respiratory medicine* 2016;4:306-322.
2. Beard JR, Officer A, de Carvalho IA, et al. The World report on ageing and health: a policy framework for healthy ageing. *The Lancet* 2016;387:2145-2154.
3. Deano RC, Glassner-Kolmin C, Rubenfire M, et al. Referral of patients with pulmonary hypertension diagnoses to tertiary pulmonary hypertension centers: the multicenter RePHerral study. *JAMA Intern Med* 2013;173:887-893.
4. Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. *Kardiol Pol* 2015;73:1127-1206.
5. Hoeper MM, Lee SH, Voswinckel R, et al. Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers. *J Am Coll Cardiol* 2006;48:2546-2552.
6. Yock PG, Popp RL. Noninvasive estimation of right ventricular systolic pressure by Doppler ultrasound in patients with tricuspid regurgitation. *Circulation* 1984;70:657-662.
7. Greiner S, Jud A, Aurich M, et al. Reliability of noninvasive assessment of systolic pulmonary artery pressure by Doppler echocardiography compared to right heart catheterization: analysis in a large patient population. *J Am Heart Assoc* 2014;3.
8. Bossone E, D'Andrea A, D'Alto M, et al. Echocardiography in pulmonary arterial hypertension: from diagnosis to prognosis. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography*

2013;26:1-14.

9. Taleb M, Khuder S, Tinkel J, et al. The diagnostic accuracy of Doppler echocardiography in assessment of pulmonary artery systolic pressure: a meta-analysis. *Echocardiography* 2013;30:258-265.

10. Janda S, Shahidi N, Gin K, et al. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. *Heart* 2011;97:612-622.

11. Zhang RF, Zhou L, Ma GF, et al. Diagnostic value of transthoracic Doppler echocardiography in pulmonary hypertension: a meta-analysis. *American journal of hypertension* 2010;23:1261-1264.

12. Haynes RB. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. *Evidence-based medicine* 2006;11:162-164.

13. Tian J, Zhang J, Ge L, et al. The methodological and reporting quality of systematic reviews from China and the USA are similar. *Journal of clinical epidemiology* 2017;85:50-58.

14. Xiu-xia L, Ya Z, Yao-long C, et al. The reporting characteristics and methodological quality of Cochrane reviews about health policy research. *Health policy (Amsterdam, Netherlands)* 2015;119:503-510.

15. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ (Clinical research ed)* 2009;339:b2535.

16. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ (Clinical research ed)* 2015;350:g7647.

17. Wang X, Chen Y, Yao L, et al. Reporting of declarations and conflicts of interest in WHO guidelines can be further improved. *Journal of clinical epidemiology* 2018;98:1-8.

18. Ge L, Tian JH, Li YN, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. *Journal of clinical epidemiology* 2018;93:45-55.

19. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health technology assessment (Winchester, England)* 2004;8(25):iii, 1-234.

20. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC medical research methodology* 2003;3:25.

21. Lee J, Kim KW, Choi SH, et al. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. *Korean journal of radiology* 2015;16:1188-1196.

22. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of clinical epidemiology* 2005;58:882-893.

23. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of clinical epidemiology* 2005;58:982-990.

24. Ahmed M, Elshinawy O, Agmy G, et al. Tricuspid regurgitation velocity versus right ventricular systolic pressure in the echocardiographic evaluation of pulmonary hypertension. *Egyptian Journal of Chest Diseases and Tuberculosis* 2019;68:203-208.

25. Shujaat A, Bajwa AA, Al-Saffar F, et al. Diagnostic accuracy of echocardiography combined with chest CT in pulmonary hypertension. *Clinical Respiratory Journal* 2018;12:948-952.

26. Schneider M, Pistritto AM, Gerges C, et al. Multi-view approach for the diagnosis of pulmonary hypertension using transthoracic echocardiography. *International Journal of Cardiovascular Imaging* 2018;34:695-700.

27. Keir GJ, Wort SJ, Kokosi M, et al. Pulmonary hypertension in interstitial lung disease: Limitations of echocardiography compared to cardiac catheterization. *Respirology* 2018;23:687-94.

28. Habash F, Gurram P, Almomani A, et al. Correlation between echocardiographic pulmonary artery pressure estimates and right heart catheterization measurement in liver transplant candidates. *Journal of Cardiovascular Imaging* 2018;26:75-84.

29. Sohrabi B, Kazemi B, Mehryar A, et al. Correlation between pulmonary artery pressure measured by echocardiography and right heart catheterization in patients with rheumatic mitral valve stenosis (A prospective study). *Echocardiography* 2016;33:7-13.

30. Balcl MK, Arl E, Vayvada M, et al. Assessment of Pulmonary Hypertension in Lung Transplantation Candidates: Correlation of Doppler Echocardiography With Right Heart Catheterization. *Transplantation Proceedings* 2016;48:2797-2802.

31. Nagel C, Henn P, Ehlken N, et al. Stress Doppler echocardiography for early detection of systemic sclerosis-associated pulmonary arterial hypertension. *Arthritis research & therapy* 2015;17:165.

32. Raevens S, Colle I, Reyntjens K, et al. Echocardiography for the detection of portopulmonary hypertension in liver transplant candidates: An analysis of cutoff values. *Liver Transplantation* 2013;19:602-610.

33. Lange TJ, Baumgartner S, Arzt M, et al. Qualitative echocardiography parameters for prediction of pulmonary hypertension. *International Journal of Clinical Practice* 2013;67:5-12.

34. Lafitte S, Pillois X, Reant P, et al. Estimation of pulmonary pressures and diagnosis of pulmonary hypertension by doppler echocardiography: A retrospective comparison of routine echocardiography and invasive hemodynamics. *Journal of the American Society of Echocardiography* 2013;26:457-463.

35. Rajaram S, Swift AJ, Capener D, et al. Comparison of the diagnostic utility of cardiac magnetic resonance imaging, computed tomography, and echocardiography in assessment of suspected pulmonary arterial hypertension in patients with connective tissue disease. *Journal of Rheumatology* 2012;39:1265-1274.

36. Parsaee M, Amin A, Nematollahi MR, et al. Comparison of transthoracic echocardiography and right heart catheterization for assessing pulmonary arterial

pressure in patients with congenital or valvular heart defects. *Iranian Heart Journal* 2012;12:54-61.

37. Hua R, Sun YW, Wu ZY, et al. Role of 2-dimensional Doppler echo-cardiography in screening portopulmonary hypertension in portal hypertension patients. *Hepatobiliary & pancreatic diseases international : HBPD INT* 2009;8:157-61.

38. Nathan SD, Shlobin OA, Barnett SD, et al. Right ventricular systolic pressure by echocardiography as a predictor of pulmonary hypertension in idiopathic pulmonary fibrosis. *Respiratory Medicine* 2008;102:1305-1310.

39. Mogollon Jimenez MV, Escoresca Ortega AM, Cabeza Letran ML, et al. Correlation of Echocardiographic and Hemodynamic Parameters in Pulmonary Hypertension Assessment Prior to Heart Transplantation. *Transplantation Proceedings* 2008;40:3023-3024.

40. Hsu VM, Moreyra AE, Wilson AC, et al. Assessment of pulmonary arterial hypertension in patients with systemic sclerosis: comparison of noninvasive tests with results of right-heart catheterization. *The Journal of rheumatology* 2008;35:458-465.

41. Fisher MR, Criner GJ, Fishman AP, et al. Estimating pulmonary artery pressures by echocardiography in patients with emphysema. *European Respiratory Journal* 2007;30:914-21.

42. Lanzarini L, Fontana A, Campana C, et al. Two simple echo-doppler measurements can accurately identify pulmonary hypertension in the large majority of patients with chronic heart failure. *Journal of Heart and Lung Transplantation* 2005;24:745-754.

43. Mukerjee D, St George D, Knight C, et al. Echocardiography and pulmonary function

as screening tests for pulmonary arterial hypertension in systemic sclerosis. *Rheumatology (Oxford)* 2004;43:461-466.

44. Arcasoy SM, Christie JD, Ferrari VA, et al. Echocardiographic assessment of pulmonary hypertension in patients with advanced lung disease. *American Journal of Respiratory and Critical Care Medicine* 2003;167:735-740.

45. Penning S, Robinson KD, Major CA, et al. A comparison of echocardiography and pulmonary artery catheterization for evaluation of pulmonary artery pressures in pregnant patients with suspected pulmonary hypertension. *American Journal of Obstetrics and Gynecology* 2001;184:1568-1570.

46. Matsuyama W, Ohkubo R, Michizono K, et al. Usefulness of transcutaneous Doppler jugular venous echo to predict pulmonary hypertension in COPD patients. *Eur Respir J* 2001;17:1128-1131.

47. Kim WR, Krowka MJ, Plevak DJ, et al. Accuracy of doppler echocardiography in the assessment of pulmonary hypertension in liver transplant candidates. *Liver Transplantation* 2000;6:453-458.

48. Denton CP, Cailes JB, Phillips GD, et al. Comparison of Doppler echocardiography and right heart catheterization to assess pulmonary hypertension in systemic sclerosis. *Br J Rheumatol* 1997;36:239-243.

49. Laaban JP, Diebold B, Zelinski R, et al. Noninvasive Estimation of Systolic Pulmonary-Artery Pressure Using Doppler Echocardiography in Patients with Chronic Obstructive Pulmonary-Disease. *Chest* 1989;96:1258-1262.

50. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J* 2019;53.

51. Gibbs JSR, Torbicki A. Proposed new pulmonary hypertension definition: is 4 mm(Hg) worth re-writing medical textbooks? *Eur Respir J* 2019;53.

52. Galderisi M, Cosyns B, Edvardsen T, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. *Eur Heart J Cardiovasc Imaging* 2017;18:1301-1310.

53. McLaughlin VV, Archer SL, Badesch DB, et al. ACCF/AHA 2009 expert consensus document on pulmonary hypertension: a report of the American College of Cardiology Foundation Task Force on Expert Consensus Documents and the American Heart Association: developed in collaboration with the American College of Chest Physicians, American Thoracic Society, Inc., and the Pulmonary Hypertension Association. *Circulation* 2009;119:2250-2294.

54. Galie N, Hoeper MM, Humbert M, et al. Guidelines for the diagnosis and treatment of pulmonary hypertension. *Eur Respir J* 2009;34:1219-1263.

55. Ukai T, Shikata S, Nakayama T, et al. A comparison of the results of prospective and retrospective cohort studies in the field of digestive surgery. *Surgery today* 2017;47:789-794.

Table 1 Characteristics of Studies Included in this Meta–analysis

Study	Year	Country	Design	N	Disease Composition of the Population	Mean Age (Years)	Male (%)	Time Interval	TTE Threshold (mmHg)	RHC Threshold (mmHg)	TTE Method
Ahmed	2019	USA	Retrospective	136	Multiple diseases	59±20	35	<3m	SPAP≥40	MPAP≥25	4TRVmax^2+RAP(IVC)
Keir	2018	Australia	Prospective	265	Interstitial lung disease	60.8±16	46	–	TRPG>46	MPAP≥25	4TRVmax^2
Habash- 1	2018	USA	Retrospective	31	Liver transplantation candidates	57±11	42	36.8±13.4d	SPAP>47	MPAP≥25	4TRVmax^2+RAP(IVC)
Habash-2	2018	USA	Retrospective	49	Multiple diseases	59±15	31	16.0±11.6d	SPA>43	MPAP≥25	4TRVmax^2+RAP(IVC)
Schneider	2018	Austria	Prospective	65	Cardiac and lung diseases	67.2	43	<48h	TRPG>32	MPAP≥25	4TRVmax^2
Balcl	2016	Turkey	Prospective	103	Lung transplantation candidates	47.6±10.4	66	<72h	SPAP>35	MPAP≥25	4TRVmax^2+RAP (NR)
Shujaat	2016	USA	Retrospective	87	Multiple diseases	54.3±15.9	29	13d*	SPAP>40	MPAP>25	4TRVmax^2+RAP (NR)
Sohrabi	2016	Iran	Prospective	300	Rheumatic mitral stenosis	59.9	31	<24h	SPAP≥35	MPAP≥25	4TRVmax^2+RAP(IVC)
Nagel	2015	Germany	Prospective	76	Systemic sclerosis	58±14	16	–	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Greiner	2014	Germany	Retrospective	1695	Cardiac disease	63±15	67	<5d	SPAP≥36	MPAP≥25	4TRVmax^2+RAP(IVC)
Lafitte	2013	France	Retrospective	114	Cardiac and lung disease	64.8±15.9	52	<48h	SPAP≥38	MPAP>25	4TRVmax^2+RAP(IVC)
Lange	2013	Germany	Retrospective	231	Multiple diseases	62±13	43	5±4d	SPAP>50	MPAP≥25	4TRVmax^2+RAP (5)
Raevens	2013	Belgium	Retrospective	152	Liver transplantation candidates	58±11	66	–	SPAP>38	MPAP≥25	4TRVmax^2+RAP (NR)
Parsaee	2012	Iran	Prospective	103	Cardiac diseases	41.0±15.8	44	<4h	SPAP≥35	MPAP>25	4TRVmax^2+RAP(IVC)
Rajaram	2012	UK	Retrospective	81	Connective tissue disease	62±14	15	<48h	TRPG≥40	MPAP≥25	4TRVmax^2
Hua	2009	China	Prospective	105	Liver transplantation candidates	49.5±11.8	63	4.2±2.0 d	SPAP≥30	MPAP≥25	4TRVmax^2+RAP(IVC)
Nathan	2008	USA	Retrospective	60	Idiopathic pulmonary fibrosis	62.9±8.6	55	32±78d	SPAP≥40	MPAP>25	4TRVmax^2+RAP(IVC)
Hsu	2008	USA	Prospective	49	Systemic Sclerosis	55	18	<4h	SPAP>47	MPAP≥25	4TRVmax^2+RAP (10)
Mogollon	2008	Spain.	Retrospective	67	Heart transplantation candidates	–	–	–	SPAP>40	MPAP>35	4TRVmax^2+RAP(IVC)
Fisher	2007	USA	Retrospective	63	Emphysema patients	65.6±6.6	60	23d	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Lanzarini	2005	Italy	Prospective	57	Heart failure	52±11	74	<24h	SPAP≥32	SPAP≥35	4TRVmax^2+RAP (IVC)
Mukerjee	2004	UK	Prospective	137	Systemic sclerosis	63	–	<3 m	TRPG>40	MPAP≥25	4TRVmax^2
Arcasoy	2003	USA	Prospective	166	COPD 68%, ILD 28%, PVD 4%	51	43	<72h	SPAP≥45	SPAP≥45	4TRVmax^2+RAP(IVC)
Penning	2001	USA	Retrospective	27	Pregnant women with cardiac diseases	28.6	0	25.8d	SPAP≥40	SPAP≥35	4TRVmax^2+RAP(IVC)
Matsuyama	2001	Japan	Prospective	35	COPD	66	94	–	SPAP≥40	MPAP>25	4TRVmax^2+RAP(IVC)
Kim	2000	USA	Prospective	74	Liver transplantation candidates	54	50	59d	SPAP>50	MPAP≥35	4TRVmax^2+RAP(IVC)
Denton	1997	UK	Prospective	20	COPD	48.6±11.7	30	1.8±2.3m	SPAP≥30	SPAP≥30	4TRVmax^2+RAP(JVP)
Laaban	1989	France	Prospective	27	COPD	63±9	78	<2d	SPAP≥35	SPAP≥35	4TRVmax^2+RAP(5)

TTE, Transthoracic echocardiography; RHC, right cardiac catheter; SPAP, systolic pulmonary artery pressure; MPAP, mean pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient; TRV, tricuspid regurgitation velocity; RAP, right atrial pressure;IVC, Inferior vena cava; JVP, jugular vein pressure; COPD, chronic obstructive pulmonary disease; ILD, Interstitial lung disease; PVD, peripheral vascular disease.
* The median time (other terms are mean time).

Table 2 Subgroup analysis

Group	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
All Studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
Time Interval								
≤24h	4	95(90–99)	0.94(0.92–0.96)	0.88(0.73–0.95)	0.90(0.53–0.99)	8.9(1.5–54.5)	0.13(0.06–0.29)	68(13–348)
≤48h*	7	95(90–99)	0.94(0.91–0.95)	0.88(0.81–0.93)	0.89(0.71–0.96)	7.8(2.8–21.3)	0.13(0.09–0.21)	59(23–148)
≤72h *	9	94(89–99)	0.91(0.89–0.93)	0.87(0.82–0.91)	0.83(0.65–0.93)	5.2(2.4–11.2)	0.15(0.11–0.21)	34(14–82)
≤1 week	13	93(87–99)	0.91(0.88–0.93)	0.87(0.84–0.90)	0.80(0.68–0.88)	4.3(2.7–6.9)	0.16(0.12–0.21)	27(15–48)
>1 week	10	97(95–99)	0.82(0.78–0.85)	0.85(0.73–0.92)	0.60(0.40–0.77)	2.1(1.3–3.4)	0.25(0.14–0.45)	9(4–21)
Unclear	5	82(63–100)	0.85(0.81–0.88)	0.79(0.63–0.99)	0.76(0.61–0.87)	3.4(1.9–5.9)	0.27(0.15–0.51)	12(5–33)
Population Disease								
cardiac diseases	6	94(89–99)	0.90(0.87–0.92)	0.90(0.86–0.93)	0.67(0.29–0.91)	2.7(0.9–8.1)	0.15(0.08–0.30)	18(3–95)
lung diseases	8	90(81–100)	0.73(0.69–9.77)	0.81(0.70–0.88)	0.61(0.53–0.69)	2.1(1.8–2.4)	0.32(0.21–0.48)	7(4–10)
multiple diseases#	6	93(87–99)	0.90(0.87–0.92)	0.89(0.84–0.92)	0.70(0.40–0.89)	3.0(1.3–7.1)	0.16(0.11–0.23)	19(6–60)
Unclear&	8	88(77–100)	0.88(0/85–0.90)	0.80(0.64–0.90)	0.85(0.80–0.89)	5.3(4.0–7.0)	0.23(0.12–0.45)	23(10–51)
Published Year								
≥2010	15	97(95–99)	0.89(0.86–0.91)	0.87(0.81–0.91)	0.74(0.62–0.83)	3.3(2.3–4.9)	0.18(0.13–0.25)	19(11–13)
<2010	13	96(93–99)	0.86(0.83–0.89)	0.84(0.74–0.90)	0.73(0.56–0.85)	3.1(1.8–5.3)	0.22(0.14–0.37)	14(6–33)
Study Design								
Prospective	15	97(95–99)	0.90(0.87–0.92)	0.86(0.77–0.91)	0.79(0.69–0.87)	4.2(2.7–6.4)	0.18(0.11–0.28)	23(12–45)
Retrospective	13	96(92–99)	0.86(0.83–0.89)	0.86(0.80–0.90)	0.65(0.49–0.78)	2.5(1.6–3.7)	0.22(0.15–0.32)	11(6–22)
Cut–off Threshold								

SPAP≥40 mmHg	8	96(93–99)	0.80(0.76–0.83)	0.84(0.75–0.91)	0.52(0.31–0.71)	1.7(1.2–2.5)	0.30(0.21–0.44)	6(3–11)
SPAP≥35 mmHg	4	76(47–100)	0.92(0.89–0.94)	0.92(0.88–0.94)	0.65(0.43–0.83)	2.6(1.4–4.9)	0.13(0.08–0.22)	16(9–28)
TRPG	4	0(0–100)	0.85(0.82–0.88)	0.75(0.58–0.86)	0.81(0.70–0.89)	4.0(2.2–7.3)	0.31(0.17–0.57)	13(4–40)

AUC, Area under curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; TRPG, tricuspid regurgitation pressure gradient.

*Studies with time intervals less than or equal to 24 hours were included in this subgroup.

* Studies with time intervals less than or equal to 24 hours and 48 hours were included in this subgroup.

Studies included a variety of diseases, including heart disease and lung disease.

&Diseases were not specifically identified in the studies (transplant candidates).

Table 3 Sensitivity analysis

Study characteristic	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
all included studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
excluding study of Penning	27	98(97–99)	0.88(0.85–0.91)	0,86(0.81–0.89)	0.75(0.66–0.82)	3.4(2.5–4.6)	0.19(0.14–0.26)	18(11–28)
MPAP≥25	21	98(97–99)	0.87(0.84–0.90)	0.83(0.77–0.88)	0.76(0.67–0.83)	3.5(2.5–4.8)	0.22(0.16–0.30)	16(10–26)
RAP method(IVC)	17	96(93–99)	0.89(0.86–0.91)	0.86(0.82–0.90)	0.73(0.59–0.84)	3.2(2.0–5.1)	0.19(0.13–0.27)	17(8–35)
Excluding high cut-off value*	21	97(95–99)	0.90(0.87–0.92)	0.88(0.85–0.91)	0.72(0.59–0.82)	3.2(2.1–4.8)	0.16(0.12–0.22)	20(11–36)

AUC, Area under curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; SPAP, systolic pulmonary artery pressure; MPAP, mean pulmonary artery pressure; RAP, right atrial pressure; IVC, Inferior vena cava.

* High cut-off value was defined as SPAP greater than 45mmHg or tricuspid regurgitation pressure gradient (TRPG) greater than 40mmHg.

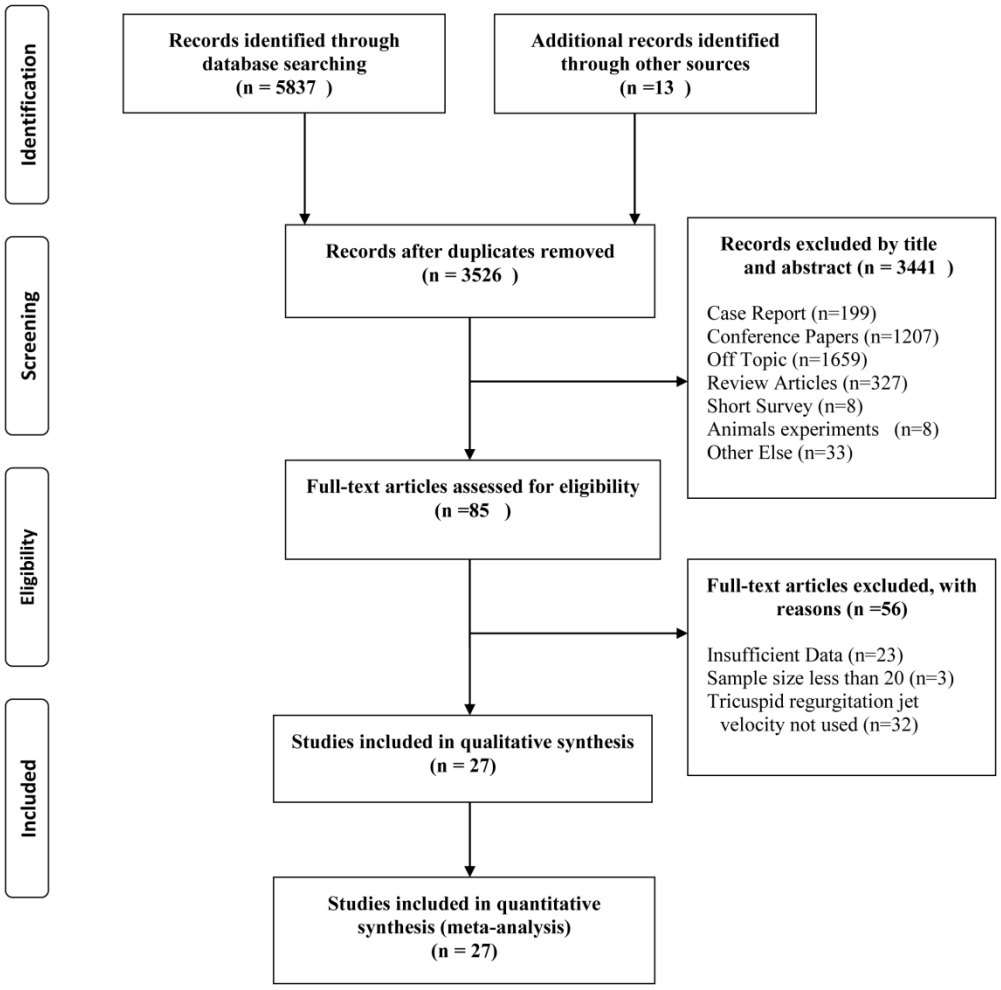

Figure 1 Flowchart for identification of studies

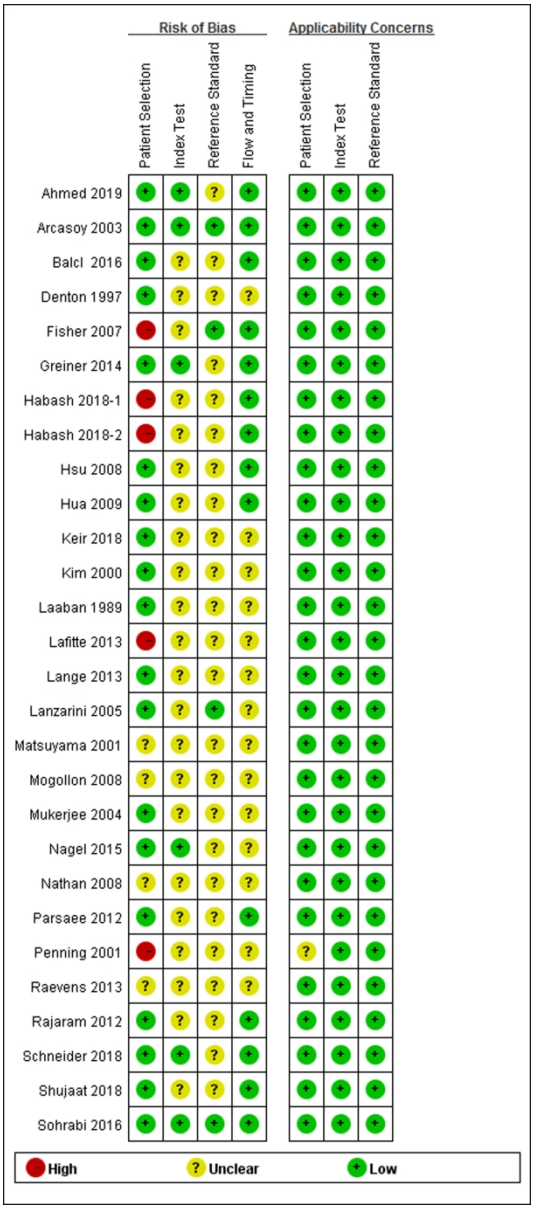

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements regarding each domain for each included study (n=28).

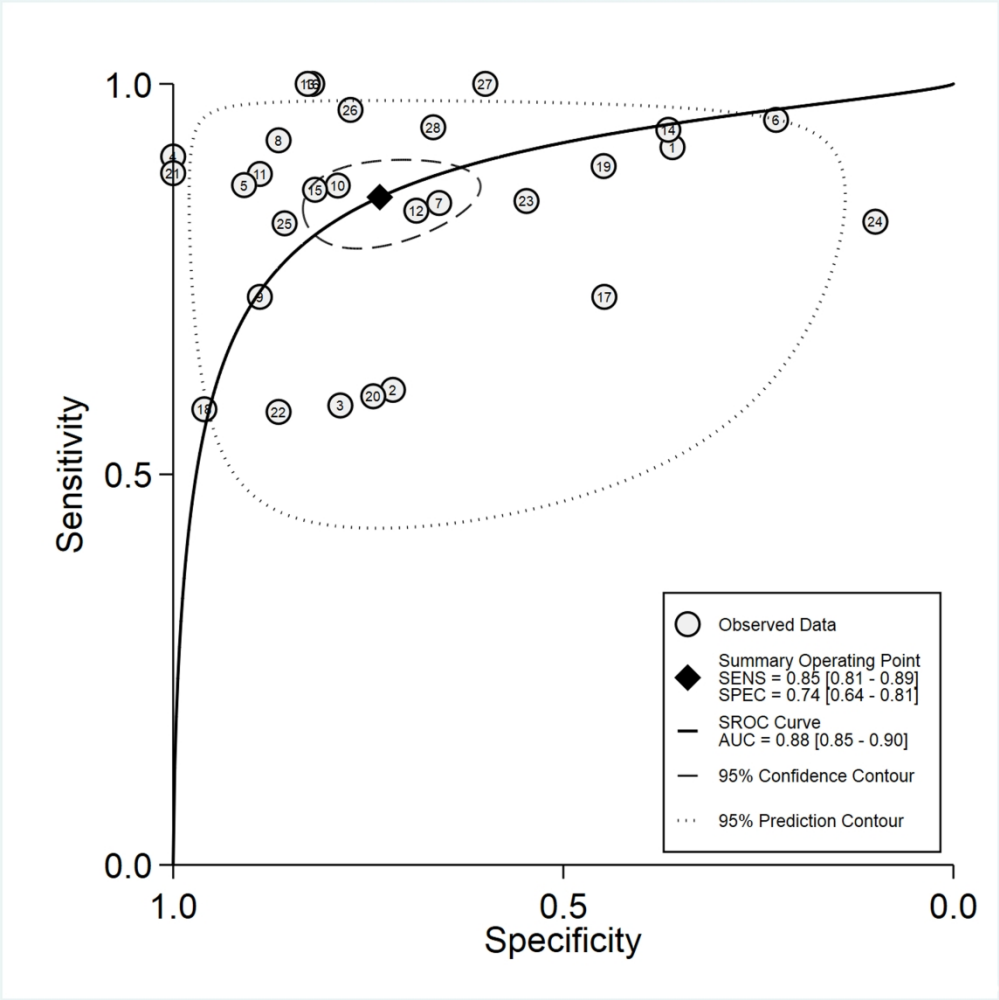

Figure 3 Summary receiver operating characteristic (SROC) graph with 95% confidence region and 95% prediction region for TTE in the diagnosis of pulmonary hypertension (n=28).

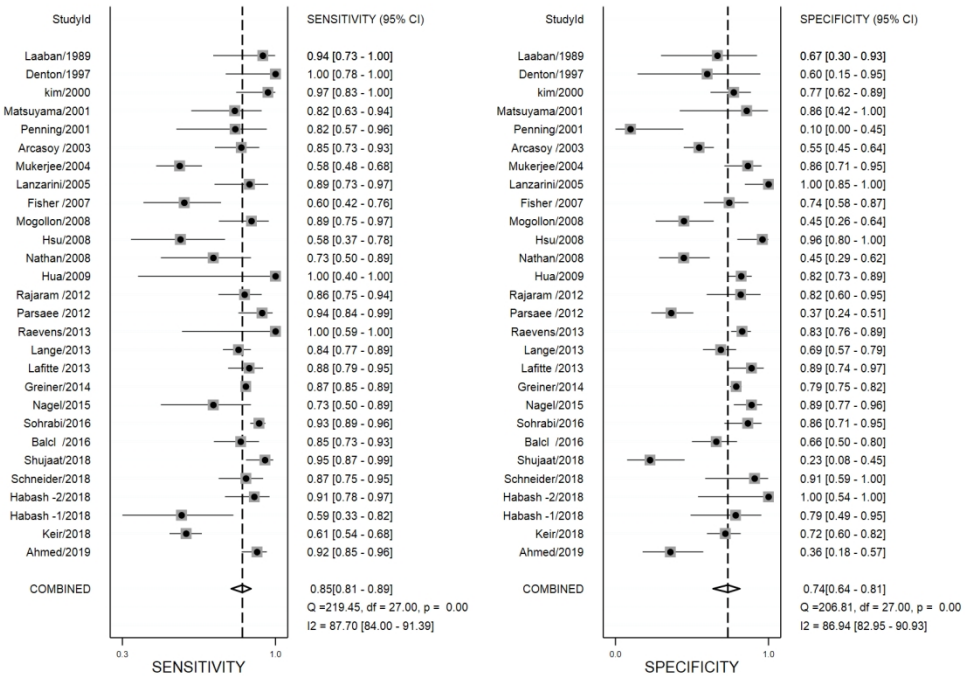

Figure 4 Forest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I2 statistic for heterogeneity (n=28).

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	6
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	6
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	7
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	7
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Supplementary Data
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	30, Figure 1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	8
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	9
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	9
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e)	9

PRISMA-DTA Checklist

		grouping and comparing tests, f) handling of different reference standards	

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	9
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	9
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	30, Figure 1
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	26 Table 1
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	31 Figure 2
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	32 Figure 3
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	33 Figure 4
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	27-29 Table 2-3
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	12
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	16
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	2

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM; The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test

PRISMA-DTA Checklist

Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: **www.prisma-statement.org**.

1. Pubmed

((((((((((((((((Echocardiography[MeSH Terms]) OR Echocardiography[Title/Abstract]) OR ECHO[Title/Abstract]) OR echocardiogram*[Title/Abstract]) OR ultrasound cardiography[Title/Abstract]) OR heart echography[Title/Abstract]) OR cardiac echography[Title/Abstract]) OR Cardiac ultrasound[Title/Abstract]) OR heart ultrasound[Title/Abstract]) OR cardial echography[Title/Abstract]) OR cardio echography[Title/Abstract]) OR echo cardiogram*[Title/Abstract]) OR echo cardiography[Title/Abstract]) OR echo sounding, heart[Title/Abstract]) OR heart echo sounding[Title/Abstract])) AND ((((((((((Hypertension, Pulmonary/diagnosis[MeSH Terms]) OR Hypertension, Pulmonary[Title/Abstract]) OR pulmonary hypertension[Title/Abstract]) OR hypertension, lung[Title/Abstract]) OR hypertensive pulmonary[Title/Abstract]) OR pulmonary arter* hypertension[Title/Abstract]) OR lung arter* hypertension[Title/Abstract]) OR lung hypertension[Title/Abstract]) OR pulmonary fixed hypertension[Title/Abstract]) OR PH[Title/Abstract])) OR pulmonary hypertensive[Title/Abstract])) AND ((((((((((Right heart catheterization[Title/Abstract]) OR Right heart catheterisation[Title/Abstract]) OR right heart catheter*[Title/Abstract]) OR right cardiac catheter*[Title/Abstract]) OR right cardiac catheterization[Title/Abstract]) OR right cardiac catheterisation[Title/Abstract]) OR RHC[Title/Abstract])) OR right - sided heart catheterization[Title/Abstract])) OR right - sided heart catheterisation[Title/Abstract]) Sort by: Best Match

2. Cochrane library

#1 MeSH descriptor: [Echocardiography] explode all trees

#2 Echocardiography

#3 echocardiogram*

#4 Echo

#5 Cardiac ultrasound

#6 heart ultrasound

#7 ultrasound cardiography

#8 heart echography

#9 cardi* echography

#10 echo cardiogra*

#11 echo sounding, heart

#12 echography, heart

#13 heart echo sounding

#14 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13

#15 MeSH descriptor: [Hypertension, Pulmonary] explode all trees

#16 pulmonary hypertension

#17 hypertension, pulmonary

#18 lung hypertension

#19 hypertension, lung

#20 pulmonary arterial hypertension

#21 lung arterial hypertension

#22 lung artery hypertension

#23 pulmonary artery hypertension

#24 pulmonary fixed hypertension

#25 pulmonary hypertensive

#26 hypertensive pulmonary vascular disease*

#27 #15 OR #16 OR #17 OR 15 OR #19 #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26

#28 right cardiac catheterization

#29 Right heart catheterization

#30 right heart catheter*

#31 right cardiac catheter*

#32 right heart catheterisation

#33 right cardiac catheterisation

#34 right - sided heart catheterization

#35 right - sided heart catheterisation

#36 RHC

#37 #28 OR #29 OR #30 #31 #32 OR #33 OR #34 OR #35 OR #36

#38 #14 AND #27 AND #37

3. EMBASE

('echocardiography'/exp OR 'echocardiography':ab OR 'ECHO':ab OR 'echocardiogram*':ab OR 'ultrasound cardiography':ab OR 'heart echography':ab OR 'cardi* echography':ab OR 'cardiac ultrasound':ab OR 'heart ultrasound':ab OR 'echo cardiogra*':ab OR 'echo sounding, heart':ab OR 'echography, heart':ab OR 'heart echo sounding':ab) AND ('pulmonary hypertension'/exp OR 'pulmonary hypertension':ab OR 'hypertension, pulmonary':ab OR 'hypertension, lung':ab OR 'hypertensive pulmonary':ab OR 'pulmonary arter* hypertension':ab OR 'lung arter* hypertension':ab OR 'lung hypertension':ab OR 'pulmonary fixed hypertension':ab OR 'pulmonary hypertensive':ab OR 'PH':ab) AND ('right cardiac catheterisation':ab OR 'right cardiac catheterization':ab OR 'right heart catheterization':ab OR 'right heart catheter*':ab OR 'right cardiac catheter*':ab OR 'RHC':ab)

4. Web of Science

#4 #3 AND #2 AND #1

DocType=All document types; Language=All languages;

#3 TS= (“right cardiac catheterization” or “right heart catheterization” or “ right heart catheter*” or “right cardiac catheter*” or “right heart catheterisation” or “right cardiac catheterisation” or “RHC” or “right - sided heart catheterization” or “right - sided heart catheterisation”)

DocType=All document types; Language=All languages;

#2 TS= (“Pulmonary Hypertension” or “PH” or “lung hypertension” or “pulmonary arter*hypertension” or “lung arter* hypertension” or “pulmonary hypertensive” or “hypertensive pulmonary”)

DocType=All document types; Language=All languages;

#1 TS= (echocardiography or ECHO or echocardiogram* or “ultrasound cardiography” or “heart echography” or “Cardiac ultrasound” or “heart ultrasound” or "cardi* echography" or "echo cardiogra*")

DocType=All document types; Language=All languages;

BMJ Open

Diagnostic Accuracy of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033084.R1
Article Type:	Original research
Date Submitted by the Author:	18-Oct-2019
Complete List of Authors:	Ni, Jin-Rong; The First Hospital of Lanzhou University Yan, Pei-Jing; Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital Liu, Shi-Dong; The First Hospital of Lanzhou University Hu, Yuan; the First Hospital of Lanzhou University Yang, Kehu; evidence based meidicine center, lanzhou university Song, Bing; The First Hospital of Lanzhou University Lei, Jun-Qiang; The First Hospital of Lanzhou University
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Radiology and imaging, Qualitative research, Global health, Cardiovascular medicine
Keywords:	Hypertension < CARDIOLOGY, Echocardiography < CARDIOLOGY, Diagnostic radiology < RADIOLOGY & IMAGING

Diagnostic Accuracy of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Jin-Rong Ni[1,3,8,9], Pei-Jing Yan[2,5,6#], Shi-Dong Liu[1,3], Yuan Hu[3], Ke-Hu Yang[2,4,5,6], Bing Song[3*], Jun-Qiang Lei[1,7,8,9*]

[1]The First Hospital (the First Clinical Medical School) of Lanzhou University, Lanzhou 730000, China; [2]Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital, Lanzhou, 730000, China; [3]Department of Cardiovascular Surgery, the First Hospital of Lanzhou University, Lanzhou 730000, China; [4]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China;[5]Evidence-Based Social Science Research Center, Lanzhou University, Lanzhou, 730000, China; [6]Key Laboratory of Evidence-based Medicine and Knowledge Translation of Gansu Province, Lanzhou, 730000, China.[7]Department of Radiology, the First Hospital of Lanzhou University, Lanzhou 730000, China. [8]Intelligent Imaging Medical Engineering Research Center of Gansu province, Lanzhou 730000, China; [9]Precision Image and Collaborative Innovation International Scientific and Technological Cooperation Base of Gansu province, Lanzhou 730000, China.

[#]Joint first author

[*]Joint corresponding authors

Corresponding Author: 1. Jun-Qiang Lei, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613919289040

E-mail: leijunqiangldyy@163.com

2. Bing Song, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613993118707

E-mail: songbingldyyxwk@163.com

Funding sources: This study was supported by the Key Laboratory of Evidence Based Medicine and Knowledge Translation Foundation of Gansu Province (Grant No. GSXZYZH2018006).

Statement on potential conflicts of interest: None declared.

Patient and Public Involvement: There is no patient or the public involved in our work.

Data Availability: All data relevant to the study are included in the article.

Key Words: Echocardiography; Pulmonary Hypertension; Catheterization; Diagnosis; Accuracy; Meta-analysis.

Abstract

Objective: To evaluate the diagnostic accuracy of transthoracic echocardiography (TTE) in patients with pulmonary hypertension (PH).

Design: Systematic review and meta-analysis.

Data sources and eligibility criteria: EMBASE, Cochrane Library for clinical trials, PubMed and Web of Science were searched from inception to June 19, 2019. Studies using both TTE and right heart catheterization (RHC) to diagnose PH were included.

Main results: A total of 27 studies involving 4386 subjects were considered as eligible for analysis. TTE had a pooled sensitivity of 85% (95% CI 81–90%), a pooled specificity of 74% (95% CI 64–81%), a pooled positive likelihood ratio of 3.2 (95% CI 2.3–4.4), a pooled negative likelihood ratio of 0.20 (95% CI 0.15–0.26), a pooled diagnostic odds ratio of 16 (95% CI 10–27), and finally an area under the summary receiver operating characteristic (SROC) curve of 0.88 (95%CI 0.85–0.90).

Conclusion: TTE has clinical value in diagnosing PH, although it cannot yet replace RHC considered as the gold standard. The accuracy of TTE may be improved by shortening the time interval between TTE and RHC and by developing an appropriate threshold. TTE may not be suitable to assess pulmonary arterial pressure in patients with pulmonary diseases.

Review registration number: PROSPERO CRD42019123289.

Strengths and limitations of this study

1. We conducted a comprehensive search of the main database, included more studies, and obtained a large sample size.

2. Detailed subgroup analysis and sensitivity analysis were performed.

3. The types of pulmonary hypertension included in the studies could not be distinguished.

4. Significant heterogeneity in our study limits the interpretation of results.

Introduction

The prevalence of pulmonary hypertension (PH) is estimated at 1% in the general population, and as high as 10% in the 600 million people older than 65.[1] Early detection and accurate assessment are vital for improved outcomes for PH patients.[2] Right heart catheterization (RHC) is the gold standard in the diagnosis of PH,[3] but it is invasive and cannot be used frequently or repeatedly.[4] The latest guideline for PH recommends transthoracic echocardiography (TTE) as a noninvasive test for screening.[3]

High quality meta-analysis has been considered as one of the key tools for achieving evidence.[5 6] Three systematic reviews and meta-analysis regarding the diagnostic accuracy of TTE for PH were published between 2010 and 2013.[7-9] The studies included in these meta-analysis were all published before 2010. In addition, two of them included fewer studies and performed a simple diagnostic data synthesis.[8 9] The other included a relatively large number of studies, but did not assess a detailed subgroup analysis.[7] In recent years, TTE has still been used in the clinical diagnosis of PH, and many new original studies have been published.[10-13] Therefore, the purpose of our study was to undertake a comprehensive systematic review and quantitative meta-analysis on the accuracy of TTE in the diagnosis of PH.

Methods

The present study is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement and the published recommendations.[14 15] The detailed protocol is accessible in PROSPERO (CRD42019123289).[16 17]

Data sources and search

A systematic search in EMBASE, Cochrane Library for clinical trials, PubMed and Web of Science was performed to find the relevant literature from inception to June 19, 2019. Subject words were combined with free words, and the search strategy was developed and adapted for each database. For unpublished trials, ClinicalTrials.gov and the trials registers on the World Health Organization International Clinical Trials Registry Platform were searched. The references of the included studies and other systematic reviews and meta-analysis were also reviewed to obtain a comprehensive list of included studies.

Study selection

Studies were selected based on the following inclusion criteria: studies that diagnosed PH by TTE; the study population was represented by patients with suspected PH; TTE measurement of systolic pulmonary artery pressure (SPAP) was performed using tricuspid regurgitation; RHC was used as the gold standard for the diagnosis of PH.

The exclusion criteria were the following: insufficient data to construct a 2×2 table; tricuspid regurgitation method was not used to calculate pulmonary artery pressure; studies with less than 20 subjects; duplicate data were used (in this case, the largest sample or the latest study was selected).

Two reviewers (JR.N and PJ.Y) independently screened the eligible studies for suitability. Disagreements were resolved by consensus. If consensus could not be reached, a third reviewer (SD.L) was deferred to arbitration and consensus. No language restriction was applied. If a study is not conducted in the author's language, professional translation software could be used.

Data extraction

Two reviewers (JR.N and PJ.Y) extracted the data independently according to a predefined data extraction sheet. The following variables were extracted from the included studies: lead author, publication year, country of study, study design, study population demographics, sample size, mean age, male ratio, time interval between TTE and RHC, the cut-off threshold levels for TTE and RHC, and number of true-positive (TP), false-negative (FN), true-negative (TN) and false-positive (FP) observations. Extracted data was cross-checked and disagreements were resolved via discussion or referral to a third reviewer (Y.H).

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies QUADAS-2 tool was used to assess the risk of bias and clinical applicability concerns of the included studies according to the Cochrane Collaboration recommendation.[18 19] Two reviewers (JR.N and PJ.Y) independently evaluated QUADAS-2 items, and all emerging conflicts were resolved by consensus.

Data synthesis and statistical analysis

Statistical analysis was performed using STATA/SE version 15.1 (Stata Corp, College Station, TX) and Review Manager Version 5.3 software (Copenhagen, Denmark, Nordic Cochrane Centre, Cochrane Collaboration, 2014). All tests were two-tailed. A p value< 0.05 was considered statistically significant.

The correlation coefficient between the logarithm of sensitivity and logarithm of one minus specificity was calculated to test whether the threshold effect was one of the sources of heterogeneity.[20] Deeks' test was used to test for publication bias.[21] The bivariate model for diagnostic meta-analysis was used to obtain pooled estimates of sensitivity and specificity.[22]

Statistical heterogeneity among studies was explored using the I^2 statistic.

Pooled sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and the area under the summary receiver operating characteristic (SROC) curve were calculated from the number of TPs, FNs, FPs, and TNs. The 95% confidence interval (CI) was estimated for each metric.

Subgroup analyses were undertaken based on the following variables: the time interval between TTE and RHC; disease classification of the study population; publication year of the study; study design (prospective or retrospective) and cut-off threshold of TTE to diagnose PH. Sensitivity analysis was undertaken by excluding low-quality studies (according to the QUADAS-2 quality assessment) or trials with characteristics different from the others.

Results

Studies selection and characteristics

Figure 1 shows the PRISMA flow chart of the literature screening. A total of 27 articles involving 4386 subjects met our inclusion criteria (Table 1).[10-13 23-45] Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2).[27]

Of the 27 eligible studies, fourteen (52%) were published during 2010–2019,[10-13 26 27 30 33 34 39 41 43-45] and thirteen (48%) were published before 2010.[23-25 28 29 31 32 35-38 40 42] Twelve (44%) studies were performed in Europe,[12 24 26 32-35 37-39 43 44] nine (30%) in United States of America (USA),[10 13 23 25 27 28 31 40 42] two (8%) in East Asia,[29 36] three (12%) in Middle East,[11 41 45] and one (4%) in Australia.[30] Most of the studies (15/27, 56%) [11 12 23 24 28-32 35 36 38 39 41 45]were of prospective design versus 44% (12/27)[10 13 25-27 33 34 37 40 42-44] retrospective.

All included studies used the tricuspid maximal regurgitation velocity (TRVmax) to estimate SPAP; the majority of these studies (23/27, 85%) used classical method (4TRVmax^2+RAP) to calculate SPAP.[10 11 13 23-28 31-37 39-45] The right atrial pressure (RAP) was calculated through the diameter and collapse of the inferior vena cava (IVC) during spontaneous respiration in sixteen(59%) studies,[10 23 25-27 31 33 35-37 39-42 44 45] through the jugular vein pressure (JVP) in one study (4%), [24] and using a fixed value (5 or 10 mm Hg) in three studies (11%). [28 32 34] Three studies (11%) did not report their method for calculating RAP. [11 13 43] Four studies (15%) used a tricuspid gradient (4TRVmax ^{2}) instead of SPAP. [12 29 30 38]

The majority of the studies (22/27, 81%) reported the time interval (mean or maximum) between TTE and RHC, [10-13 23-29 31-35 38 40-42 44 45] while five (5/9, 19%) did not. [30 36 37 39 43] Nine studies (33%) considered time intervals greater than one week,[10 13 24 25 27 31 38 40 42] while thirteen studies (48%) considered time intervals of less than one week.[11 12 23 26 29 32-35 37 39 41 44] The time interval between TTE and RHC ranged from four hours to three months.

Quality Assessment

The quality assessment of the included studies according to the QUADAS-2 inventory is shown in Figure 2. In twenty (74%) study protocols, [10-13 23 24 26 28-32 34 35 37-39 41 44 45] consecutive subjects were enrolled, with no inappropriate exclusions. The risk of bias during patient recruitment was unclear in the remaining seven (26%) studies, [25 27 33 36 40 42 43] as patient recruitment was not reported. In six (22%) studies investigators designed the single-blind methods for TTE.[10 12 23 26 39 45] Double blinding in imaging assessment was not mentioned in any studies. The risk of bias on flow and timing between the index test and reference standard was categorized as unclear in 14 (52%) study protocols that did not explicitly state the successful

investigation with both index and reference tests in all included patients.[24 30-40 42 43]

Quantitative Analysis

The SROC curve for TTE is shown in Figure 3. Four studies fall within the 95% confidence interval.[11 26 34 44] The AUC was 0.88 (95%CI 0.85–0.90). The pooled sensitivity and specificity for TTE were 85% (95%CI 81%–90%) and 74% (95%CI 64%–81%), respectively (Figure 4). The pooled PLR and NLR were 3.2 (95%CI 2.3–4.4) and 0.20 (95%CI 0.15–0.26). The pooled DOR for TTE was 16 (95%CI 10–27).

The heterogeneity in our study was significant. The threshold test proved that the threshold effect was not the source of heterogeneity (r=-0.34, *P*=0.12). Deeks' test for funnel plot asymmetry suggested no publication bias (*P*=0.69).The results of the subgroup analysis are presented in Table 2. The sensitivity (87%, 95%CI 81%–91%), specificity (74%, 95%CI 62%–83%) and AUC (0.89, 95%CI 0.86–0.91) of TTE to diagnose PH was higher for studies published in 2010 and later compared to those published before 2010. Among the time interval subgroups, the group with the shortest time interval between TTE and RHC had the best diagnostic effect, with sensitivity, specificity and AUC of 88% (95%CI 73%–95%), 90% (95%CI 53%–99%) and 0.94(95%CI 0.92–0.96), respectively. The disease composition of the study population also affected the diagnostic accuracy of TTE. Compared with patients with other diseases, TTE had lower sensitivity (81%, 95%CI 70%–88%), specificity (61%, 95%CI 53%–69%) and AUC (0.73, 95%CI 0.69–0.77) in the subgroup of patients with definite lung diseases.

Subgroup analysis of different cut-off thresholds to diagnose PH based on TTE showed that the subgroup with a cut-off threshold of 35 mmHg had superior performance than that at 40

mmHg. The sensitivity, specificity and AUC of the former were respectively 92% (95%CI 88%–94%), 65% (95%CI 43%–83%) and 0.92 (95%CI 0.89–0.94), while the sensitivity, specificity and AUC at 40 mmHg were 84% (95%CI 75%–91%), 52% (95%CI 31%–71%) and 0.80 (95%CI 76%–83%).

The sensitivity analysis results are shown in Table 3. After excluding low-quality studies and studies with specific characteristics, sensitivity analysis did not reveal a source for the heterogeneity in the diagnostic accuracy analysis. Overall, the pooled meta-analysis results for outcomes were in accordance to our sensitivity analyses.

Discussion

Our study found that TTE has a better sensitivity but moderate specificity for the detection of PH. In addition, shortening the time interval between TTE and RHC and developing an appropriate threshold could improve the accuracy of TTE. However, the accuracy of TTE to diagnose PH in patients with lung diseases was low.

Although PH is a chronic disease, we still believe that the shortest possible time interval between TTE and RHC is more conducive. Otherwise, changes in the patient's condition and the application of intervention measures would lead to an increase in the deviation of the results of the two examinations. A detailed subgroup analysis was performed according to the time interval between TTE and RHC. As expected, the diagnostic accuracy was highest when the time interval was less than or equal to 24 hours. The results also showed that the efficacy of TTE in the diagnosis of PH was gradually reduced with the extension of the time interval.

Subgroups analysis based on the disease composition of the population suggested that the diagnostic accuracy of TTE was lower in patients with lung diseases. Changes associated with

chronic pulmonary disease, including marked increase in intrathoracic gas, consolidation of lung tissue, expansion of the thoracic cage, and alterations in the position of the heart, adversely affect the imaging quality and parameters measurement of TTE.[46] Therefore, the use of TTE to measure pulmonary pressure in patients with lung diseases might not be an ideal choice.

Guideline recommend the use of IVC width and collapse rate to estimate RAP, [3] which was not used in some of the included studies. Sensitivity analysis for this point showed that studies which calculated RAP through IVC do not seem to have a higher diagnostic performance. In order to avoid errors caused by RAP estimation, TRVmax was also considered as an indicator to evaluate the possibility of PH. Four studies using TRPG (4TRVmax2) instead of SPAP were grouped into a subgroup and showed that this subgroup had good diagnostic specificity but poor sensitivity.

The sensitivity analysis based on the mean pulmonary artery pressure (MPAP) threshold of 25 mmHg did not result in a higher diagnostic value than the whole, indicating that the overall results were stable. It has been suggested that MPAP threshold of 25 mmHg is arbitrarily chosen and that lowering it to 20 mmHg (two standard deviations higher than MPAP for the population) is considered a scientific methods.[47] However, some scientists insist that it is premature to reduce the MPAP threshold to 20 mmHg because of the risk of over-diagnosis, unclear treatment implications and additional psychological burden on patients.[48] Since none of the study we included used MPAP>20mmHg as the diagnostic threshold for RHC, subgroup analysis on the two thresholds of 20mmHg and 25mmHg could not be performed. Therefore, we expect that more studies may be conducted in the future to verify the appropriate threshold of RHC.

In our review, the cut-off thresholds of SPAP ranged from 30 to 50 mmHg. Subgroup analysis showed that the diagnostic accuracy of the group of 35 mmHg was higher. Sensitivity analysis results of studies that excluded high TTE cut-off value showed that a high cut-off value increased the specificity and reduced the sensitivity of TTE. Due to the small sample size of the subgroup in this study, the value of the cut-off threshold still needs to be determined by further prospective studies of multi-center and large samples.

Subgroup analysis according to the publication year confirmed that studies published after 2010 had only a slightly higher diagnostic accuracy than previous studies. With the improvement of TTE technology and instruments in the past ten years, the diagnostic efficiency of PH has not been significantly improved, which forces us to pay attention to other TTE parameters.[49 50] Perhaps, this could be a new direction for future studies on PH diagnosis.

Limitations:

Several limitations are present in our study. Firstly, the systematic review and meta-analysis is a secondary research method based on original research and the quality of the included study affects the results. In addition, the possibility of missing relevant articles objectively exists, and significant heterogeneity may limit the interpretation of the results. Secondly, the accuracy of echocardiography relies heavily on the operator's ability, experience, and operational discipline. In order to obtain more original studies, we did not strictly controlled this aspect. Thirdly, the studies included in this review involve several different types of PH, and some studies do not describe the basic disease and PH type in detail. It is obvious that pulmonary lesions can affect the quality of TTE imaging, leading to underestimated results.

Conclusion

TTE has clinical value in the diagnosis of PH thanks to its better sensitivity and moderate specificity, but it cannot yet replace RHC considered as the gold standard. Shortening the time interval between TTE and RHC and developing an appropriate threshold can improve the accuracy of TTE. TTE may not be suitable to assess pulmonary arterial pressure in patients with pulmonary disease. It may be necessary to combine multiple TTE parameters and conduct multi-center, large-sample studies to further improve the accuracy of TTE in the diagnosis of PH in future research.

Author's contribution

The joint corresponding authors (JQ.L and B.S) are responsible for the design and implementation of the study. SD.L is responsible for the quality control of study selection. Y.H carried out quality control on the links of data extraction. KH.Y provided guidance in literature retrieval and data processing methodology and was responsible for the quality evaluation part. JR.N and PJ.Y have done the systematic review of the literature and extracted data. JR.N has conducted the meta-analyses, and two authors (JR.N, PJ.Y) have substantially contributed to interpretation of data and co-authored the article. All the authors have made repeated revisions to the article. The corresponding authors (JQ.L and B.S) and JR.N take responsibility of the integrity of the analyses.

Reference

1. Hoeper MM, Humbert M, Souza R, et al. A global view of pulmonary hypertension. Lancet Respir Med 2016;4(4):306-22.

2. Lau EM, Humbert M, Celermajer DS. Early detection of pulmonary arterial hypertension. Nat Rev Cardiol 2015;12(3):143-55.

3. Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Kardiol Pol 2015;73(12):1127-206.

4. Hoeper MM, Lee SH, Voswinckel R, et al. Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers. J Am Coll Cardiol 2006;48(12):2546-52.

5. Tian J, Zhang J, Ge L, et al. The methodological and reporting quality of systematic reviews from China and the USA are similar. J Clin Epidemiol 2017;85:50-58.

6. Xiu-xia L, Ya Z, Yao-long C, et al. The reporting characteristics and methodological quality of Cochrane reviews about health policy research. Health Policy 2015;119(4):503-10.

7. Janda S, Shahidi N, Gin K, et al. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. Heart 2011;97(8):612-22.

8. Taleb M, Khuder S, Tinkel J, et al. The diagnostic accuracy of Doppler echocardiography in assessment of pulmonary artery systolic pressure: a meta-analysis. Echocardiography 2013;30(3):258-65.

9. Zhang RF, Zhou L, Ma GF, et al. Diagnostic value of transthoracic Doppler

echocardiography in pulmonary hypertension: a meta-analysis. Am J Hypertens 2010;23(12):1261-64.

10. Ahmed M, Elshinawy O, Agmy G, et al. Tricuspid regurgitation velocity versus right ventricular systolic pressure in the echocardiographic evaluation of pulmonary hypertension. Egypt J Chest Dis Tuberc 2019;68(2):203-08.

11. Balcı MK, Arı E, Vayvada M, et al. Assessment of Pulmonary Hypertension in Lung Transplantation Candidates: Correlation of Doppler Echocardiography With Right Heart Catheterization. Transplant Proc 2016;48(8):2797-802.

12. Schneider M, Pistritto AM, Gerges C, et al. Multi-view approach for the diagnosis of pulmonary hypertension using transthoracic echocardiography. Int J Cardiovasc Imaging 2018;34(5):695-700.

13. Shujaat A, Bajwa AA, Al-Saffar F, et al. Diagnostic accuracy of echocardiography combined with chest CT in pulmonary hypertension. Clin Respir J 2018;12(3):948-52.

14. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.

15. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;350:g7647.

16. Ge L, Tian JH, Li YN, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. J Clin Epidemiol 2018;93:45-55.

17. Wang X, Chen Y, Yao L, et al. Reporting of declarations and conflicts of interest in

WHO guidelines can be further improved. J Clin Epidemiol 2018;98:1-8.

18. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004;8(25):iii, 1-234.

19. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

20. Lee J, Kim KW, Choi SH, et al. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean J Radiol 2015;16(6):1188-96.

21. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58(9):882-93.

22. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58(10):982-90.

23. Arcasoy SM, Christie JD, Ferrari VA, et al. Echocardiographic assessment of pulmonary hypertension in patients with advanced lung disease. Am J Respir Crit Care Med 2003;167(5):735-40.

24. Denton CP, Cailes JB, Phillips GD, et al. Comparison of Doppler echocardiography and right heart catheterization to assess pulmonary hypertension in systemic sclerosis. Br J Rheumatol 1997;36(2):239-43.

25. Fisher MR, Criner GJ, Fishman AP, et al. Estimating pulmonary artery pressures by echocardiography in patients with emphysema. Eur Respir J 2007;30(5):914-21.

26. Greiner S, Jud A, Aurich M, et al. Reliability of noninvasive assessment of systolic pulmonary artery pressure by Doppler echocardiography compared to right heart catheterization: analysis in a large patient population. J Am Heart Assoc 2014;3(4).

27. Habash F, Gurram P, Almomani A, et al. Correlation between echocardiographic pulmonary artery pressure estimates and right heart catheterization measurement in liver transplant candidates. J Cardiovasc Imaging 2018;26(2):75-84.

28. Hsu VM, Moreyra AE, Wilson AC, et al. Assessment of pulmonary arterial hypertension in patients with systemic sclerosis: comparison of noninvasive tests with results of right-heart catheterization. J Rheumatol 2008;35(3):458-65.

29. Hua R, Sun YW, Wu ZY, et al. Role of 2-dimensional Doppler echo-cardiography in screening portopulmonary hypertension in portal hypertension patients. Hepatobiliary Pancreat. Hepatobiliary Pancreat Dis Int 2009;8(2):157-61.

30. Keir GJ, Wort SJ, Kokosi M, et al. Pulmonary hypertension in interstitial lung disease: Limitations of echocardiography compared to cardiac catheterization. Respirology 2018;23(7):687-94.

31. Kim WR, Krowka MJ, Plevak DJ, et al. Accuracy of doppler echocardiography in the assessment of pulmonary hypertension in liver transplant candidates. Liver Transpl 2000;6(4):453-58.

32. Laaban JP, Diebold B, Zelinski R, et al. Noninvasive Estimation of Systolic Pulmonary-Artery Pressure Using Doppler Echocardiography in Patients with Chronic

Obstructive Pulmonary-Disease. Chest 1989;96(6):1258-62.

33. Lafitte S, Pillois X, Reant P, et al. Estimation of pulmonary pressures and diagnosis of pulmonary hypertension by doppler echocardiography: A retrospective comparison of routine echocardiography and invasive hemodynamics. J Am Soc Echocardiogr 2013;26(5):457-63.

34. Lange TJ, Baumgartner S, Arzt M, et al. Qualitative echocardiography parameters for prediction of pulmonary hypertension. Int J Clin Pract 2013;67(SUPPL. 179):5-12.

35. Lanzarini L, Fontana A, Campana C, et al. Two simple echo-doppler measurements can accurately identify pulmonary hypertension in the large majority of patients with chronic heart failure. J Heart Lung Transplant 2005;24(6):745-54.

36. Matsuyama W, Ohkubo R, Michizono K, et al. Usefulness of transcutaneous Doppler jugular venous echo to predict pulmonary hypertension in COPD patients. Eur Respir J 2001;17(6):1128-31.

37. Mogollon Jimenez MV, Escoresca Ortega AM, Cabeza Letran ML, et al. Correlation of Echocardiographic and Hemodynamic Parameters in Pulmonary Hypertension Assessment Prior to Heart Transplantation. Transplant Proc 2008;40(9):3023-24.

38. Mukerjee D, St George D, Knight C, et al. Echocardiography and pulmonary function as screening tests for pulmonary arterial hypertension in systemic sclerosis. Rheumatology (Oxford) 2004;43(4):461-6.

39. Nagel C, Henn P, Ehlken N, et al. Stress Doppler echocardiography for early detection of systemic sclerosis-associated pulmonary arterial hypertension. Arthritis Res Ther 2015;17(1):165.

40. Nathan SD, Shlobin OA, Barnett SD, et al. Right ventricular systolic pressure by echocardiography as a predictor of pulmonary hypertension in idiopathic pulmonary fibrosis. Respir Med 2008;102(9):1305-10.

41. Parsaee M, Amin A, Nematollahi MR, et al. Comparison of transthoracic echocardiography and right heart catheterization for assessing pulmonary arterial pressure in patients with congenital or valvular heart defects. Iran Heart J 2012;12(4):54-61.

42. Penning S, Robinson KD, Major CA, et al. A comparison of echocardiography and pulmonary artery catheterization for evaluation of pulmonary artery pressures in pregnant patients with suspected pulmonary hypertension. Am J Obstet Gynecol 2001;184(7):1568-70.

43. Raevens S, Colle I, Reyntjens K, et al. Echocardiography for the detection of portopulmonary hypertension in liver transplant candidates: An analysis of cutoff values. Liver Transpl 2013;19(6):602-10.

44. Rajaram S, Swift AJ, Capener D, et al. Comparison of the diagnostic utility of cardiac magnetic resonance imaging, computed tomography, and echocardiography in assessment of suspected pulmonary arterial hypertension in patients with connective tissue disease. J Rheumatol 2012;39(6):1265-74.

45. Sohrabi B, Kazemi B, Mehryar A, et al. Correlation between pulmonary artery pressure measured by echocardiography and right heart catheterization in patients with rheumatic mitral valve stenosis (A prospective study). Echocardiography 2016;33(1):7-13.

46. Galderisi M, Cosyns B, Edvardsen T, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2017;18(12):1301-10.

47. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. Eur Respir J 2019;53(1).

48. Gibbs JSR, Torbicki A. Proposed new pulmonary hypertension definition: is 4 mm(Hg) worth re-writing medical textbooks? Eur Respir J 2019;53(3).

49. Misra D, Kendes A, Sulica R, et al. Exercise-induced pulmonary hypertension by stress echocardiography: Prevalence and correlation with right heart hemodynamics. Int J Cardiol 2017;228:518-22.

50. Vitarelli A, Mangieri E, Terzano C, et al. Three-dimensional echocardiography and 2D-3D speckle-tracking imaging in chronic pulmonary hypertension: diagnostic accuracy in detecting hemodynamic signs of right ventricular (RV) failure. J Am Heart Assoc 2015;4(3):e001584.

Table 1 Characteristics of each study included in this meta-analysis.

Study	Year	Country	Design	N	Disease Composition of the Population	Mean Age (Years)	Male (%)	Time Interval	TTE Threshold (mmHg)	RHC Threshold (mmHg)	TTE Method
Ahmed	2019	USA	Retrospective	136	Multiple diseases	59±20	35	<3m	SPAP≥40	MPAP≥25	4TRVmax^2+RAP(IVC)
Keir	2018	Australia	Prospective	265	Interstitial lung disease	60.8±16.5	46	–	TRPG>46	MPAP≥25	4TRVmax^2
Habash-1	2018	USA	Retrospective	31	Liver transplantation candidates	57±11	42	36.8±13.4d	SPAP>47	MPAP≥25	4TRVmax^2+RAP(IVC)
Habash-2	2018	USA	Retrospective	49	Multiple diseases	59±15	31	16.0±11.6d	SPA>43	MPAP≥25	4TRVmax^2+RAP(IVC)
Schneider	2018	Austria	Prospective	65	Cardiac and lung diseases	67.2	43	<48h	TRPG>32	MPAP≥25	4TRVmax^2
Balcl	2016	Turkey	Prospective	103	Lung transplantation candidates	47.6±10.4	66	<72h	SPAP>35	MPAP≥25	4TRVmax^2+RAP (NR)
Shujaat	2016	USA	Retrospective	87	Multiple diseases	54.3±15.9	29	13d*	SPAP>40	MPAP>25	4TRVmax^2+RAP (NR)
Sohrabi	2016	Iran	Prospective	300	Rheumatic mitral stenosis	59.9	31	<24h	SPAP≥35	MPAP≥25	4TRVmax^2+RAP(IVC)
Nagel	2015	Germany	Prospective	76	Systemic sclerosis	58±14	16	–	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Greiner	2014	Germany	Retrospective	1695	Cardiac disease	63±15	67	<5d	SPAP≥36	MPAP≥25	4TRVmax^2+RAP(IVC)
Lafitte	2013	France	Retrospective	114	Cardiac and lung disease	64.8±15.9	52	<48h	SPAP≥38	MPAP>25	4TRVmax^2+RAP(IVC)
Lange	2013	Germany	Retrospective	231	Multiple diseases	62±13	43	5±4d	SPAP>50	MPAP≥25	4TRVmax^2+RAP (5)
Raevens	2013	Belgium	Retrospective	152	Liver transplantation candidates	58±11	66	–	SPAP>38	MPAP≥25	4TRVmax^2+RAP (NR)
Parsaee	2012	Iran	Prospective	103	Cardiac diseases	41.0±15.8	44	<4h	SPAP≥35	MPAP>25	4TRVmax^2+RAP(IVC)
Rajaram	2012	UK	Retrospective	81	Connective tissue disease	62±14	15	<48h	TRPG≥40	MPAP≥25	4TRVmax^2
Hua	2009	China	Prospective	105	Liver transplantation candidates	49.5±11.8	63	4.2±2.0d	SPAP≥30	MPAP≥25	4TRVmax^2+RAP(IVC)
Nathan	2008	USA	Retrospective	60	Idiopathic pulmonary fibrosis	62.9±8.6	55	32±78d	SPAP≥40	MPAP>25	4TRVmax^2+RAP(IVC)
Hsu	2008	USA	Prospective	49	Systemic Sclerosis	55	18	<4h	SPAP>47	MPAP≥25	4TRVmax^2+RAP (10)
Mogollon	2008	Spain.	Retrospective	67	Heart transplantation candidates	–	–	–	SPAP>40	MPAP>35	4TRVmax^2+RAP(IVC)
Fisher	2007	USA	Retrospective	63	Emphysema patients	65.6±6.6	60	23d	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Lanzarini	2005	Italy	Prospective	57	Heart failure	52±11	74	<24h	SPAP≥32	SPAP≥35	4TRVmax^2+RAP(IVC

											)
Mukerjee	2004	UK	Prospective	137	Systemic sclerosis	63	–	<3 m	TRPG>40	MPAP≥25	4TRVmax²
Arcasoy	2003	USA	Prospective	166	COPD 68%, ILD 28%, PVD 4%	51	43	<72h	SPAP≥45	SPAP≥45	4TRVmax²+RAP(IVC)
Penning	2001	USA	Retrospective	27	Pregnant women with cardiac diseases	28.6	0	25.8d	SPAP≥40	SPAP≥35	4TRVmax²+RAP(IVC)
Matsuyama	2001	Japan	Prospective	35	COPD	66	94	–	SPAP≥40	MPAP>25	4TRVmax²+RAP(IVC)
Kim	2000	USA	Prospective	74	Liver transplantation candidates	54	50	59d	SPAP>50	MPAP≥35	4TRVmax²+RAP(IVC)
Denton	1997	UK	Prospective	20	COPD	48.6±11.7	30	1.8±2.3m	SPAP≥30	SPAP≥30	4TRVmax²+RAP(JVP)
Laaban	1989	France	Prospective	27	COPD	63±9	78	<2d	SPAP≥35	SPAP≥35	4TRVmax²+RAP(5)

USA, United States of America; UK, United Kingdom of Great Britain and Northern Ireland; TTE, Transthoracic echocardiography; RHC, right heart catheterization; SPAP, systolic pulmonary artery pressure; MPAP, mean pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient; TRV, tricuspid regurgitation velocity; RAP, right atrial pressure;IVC, Inferior vena cava; JVP, jugular vein pressure; COPD, chronic obstructive pulmonary disease; ILD, Interstitial lung disease; PVD, peripheral vascular disease; NR, not report.

* The median time (other terms are mean time)

Table 2 Subgroup analysis

Group	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
All Studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
Time Interval								
≤24h	4	95(90–99)	0.94(0.92–0.96)	0.88(0.73–0.95)	0.90(0.53–0.99)	8.9(1.5–54.5)	0.13(0.06–0.29)	68(13–348)
≤48h*	7	95(90–99)	0.94(0.91–0.95)	0.88(0.81–0.93)	0.89(0.71–0.96)	7.8(2.8–21.3)	0.13(0.09–0.21)	59(23–148)
≤72h *	9	94(89–99)	0.91(0.89–0.93)	0.87(0.82–0.91)	0.83(0.65–0.93)	5.2(2.4–11.2)	0.15(0.11–0.21)	34(14–82)
≤1 week	13	93(87–99)	0.91(0.88–0.93)	0.87(0.84–0.90)	0.80(0.68–0.88)	4.3(2.7–6.9)	0.16(0.12–0.21)	27(15–48)
>1 week	10	97(95–99)	0.82(0.78–0.85)	0.85(0.73–0.92)	0.60(0.40–0.77)	2.1(1.3–3.4)	0.25(0.14–0.45)	9(4–21)
Unclear	5	82(63–100)	0.85(0.810.88)	0.79(0.63–0.99)	0.76(0.61–0.87)	3.4(1.9–5.9)	0.27(0.15–0.51)	12(5–33)
Population Disease								
Cardiac diseases	6	94(89–99)	0.90(0.87–0.92)	0.90(0.86–0.93)	0.67(0.29–0.91)	2.7(0.9–8.1)	0.15(0.08–0.30)	18(3–95)
Lung diseases	8	90(81–100)	0.73(0.69–9.77)	0.81(0.70–0.88)	0.61(0.53–0.69)	2.1(1.8–2.4)	0.32(0.21–0.48)	7(4–10)
Multiple diseases#	6	93(87–99)	0.90(0.87–0.92)	0.89(0.84–0.92)	0.70(0.40–0.89)	3.0(1.3–7.1)	0.16(0.11–0.23)	19(6–60)
Unclear&	8	88(77–100)	0.88(0/85–0.90)	0.80(0.64–0.90)	0.85(0.80–0.89)	5.3(4.0–7.0)	0.23(0.12–0.45)	23(10–51)
Published Year								
≥2010	15	97(95–99)	0.89(0.86–0.91)	0.87(0.81–0.91)	0.74(0.62–0.83)	3.3(2.3–4.9)	0.18(0.13–0.25)	19(11–13)
<2010	13	96(93–99)	0.86(0.83–0.89)	0.84(0.74–0.90)	0.73(0.56–0.85)	3.1(1.8–5.3)	0.22(0.14–0.37)	14(6–33)
Study Design								
Prospective	15	97(95–99)	0.90(0.87–0.92)	0.86(0.77–0.91)	0.79(0.69–0.87)	4.2(2.7–6.4)	0.18(0.11–0.28)	23(12–45)
Retrospective	13	96(92–99)	0.86(0.83–0.89)	0.86(0.80–0.90)	0.65(0.49–0.78)	2.5(1.6–3.7)	0.22(0.15–0.32)	11(6–22)
TTE Threshold								

SPAP≥40 mmHg	8	96(93–99)	0.80(0.76–0.83)	0.84(0.75–0.91)	0.52(0.31–0.71)	1.7(1.2–2.5)	0.30(0.21–0.44)	6(3–11)
SPAP≥35 mmHg	4	76(47–100)	0.92(0.890.94)	0.92(0.88–0.94)	0.65(0.43–0.83)	2.6(1.4–4.9)	0.13(0.08–0.22)	16(9–28)
TRPG	4	0(0–100)	0.85(0.82–0.88)	0.75(0.58–0.86)	0.81(0.70–0.89)	4.0(2.2–7.3)	0.31(0.17–0.57)	13(4–40)

AUC, Area under curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; TTE, Transthoracic echocardiography; SPAP, systolic pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient.
*Studies with time intervals less than or equal to 24 hours were included in this subgroup.

* Studies with time intervals less than or equal to 24 hours and 48 hours were included in this subgroup.

Studies included a variety of diseases, including heart disease and lung disease.
&Diseases were not specifically identified in the studies (transplant candidates).

Table 3 Sensitivity analysis for diagnostic accuracy meta-analysis.

Study characteristic	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
All included studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
Excluding study of Penning*	27	98(97–99)	0.88(0.85–0.91)	0,86(0.81–0.89)	0.75(0.66–0.82)	3.4(2.5–4.6)	0.19(0.14–0.26)	18(11–28)
RHC threshold MPAP≥25 mmHg	21	98(97–99)	0.87(0.84–0.90)	0.83(0.77–0.88)	0.76(0.67–0.83)	3.5(2.5–4.8)	0.22(0.16–0.30)	16(10–26)
RAP method(IVC)£	17	96(93–99)	0.89(0.86–0.91)	0.86(0.82–0.90)	0.73(0.59–0.84)	3.2(2.0–5.1)	0.19(0.13–0.27)	17(8–35)
Excluding high TTE threshold *	21	97(95–99)	0.90(0.87–0.92)	0.88(0.85–0.91)	0.72(0.59–0.82)	3.2(2.1–4.8)	0.16(0.12–0.22)	20(11–36)

AUC, Area under curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; MPAP, mean pulmonary artery pressure; TTE, Transthoracic echocardiography; RHC, right heart catheterization;RAP, right atrial pressure; IVC, Inferior vena cava.

* Study of Penning was excluded because only pregnant women withcardiac disease were included as subjects.

£The RAP was calculated through the diameter and collapse of IVC during spontaneous respiration. Habash's study was divided into two independent parts, so the results section showed 16 studies, but 17 sets of data were analyzed.

* High TTE threshold was defined as SPAP greater than 45 mmHg or tricuspid regurgitation pressure gradient (TRPG) greater than 40 mmHg.

Figure legend/caption

Figure 1 Flowchart for identification of studies.

*Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2). 27 studies were included, but 28 sets of data were analyzed.

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study (n=28).

Figure 3 Summary receiver operating characteristic (SROC) graph with 95% confidence region and 95% prediction region for TTE in the diagnosis of pulmonary hypertension (n=28).

Figure 4 Forrest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I^2 statistic for heterogeneity (n=28).

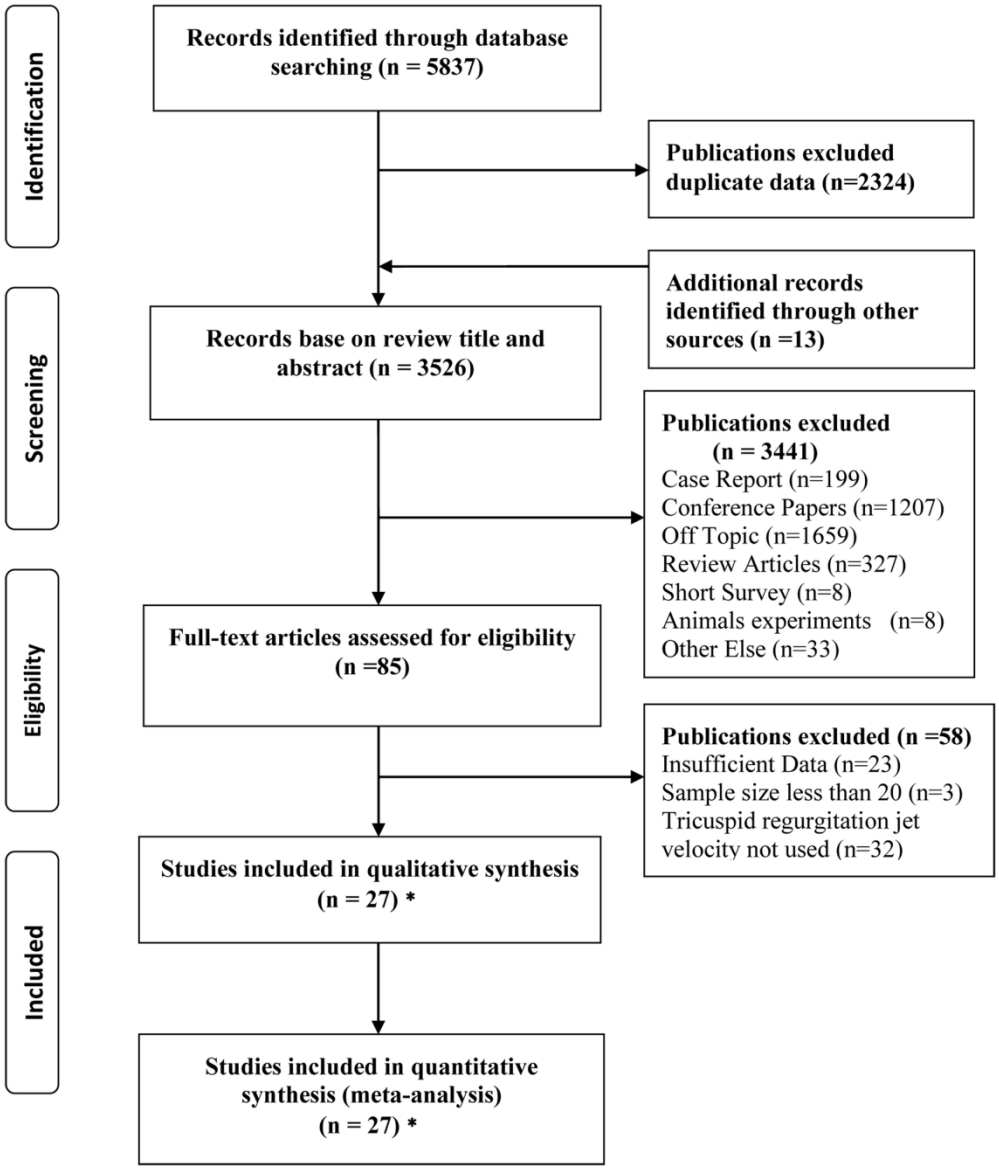

Figure 1 Flowchart for identification of studies.

·Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2). 27 studies were included, but 28 sets of data were analyzed.

157x188mm (300 x 300 DPI)

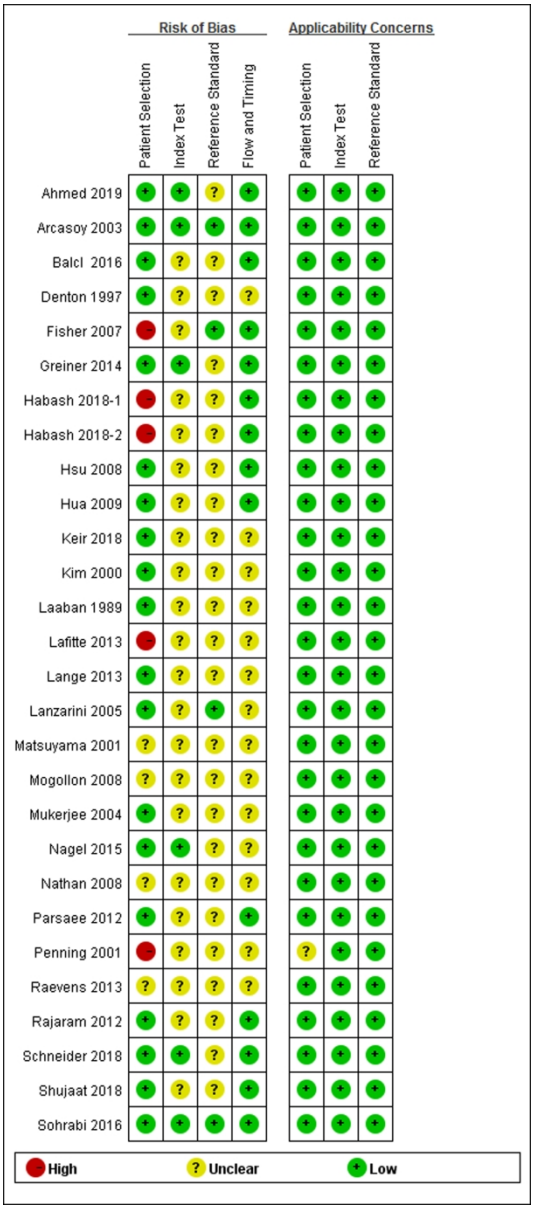

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study (n=28).

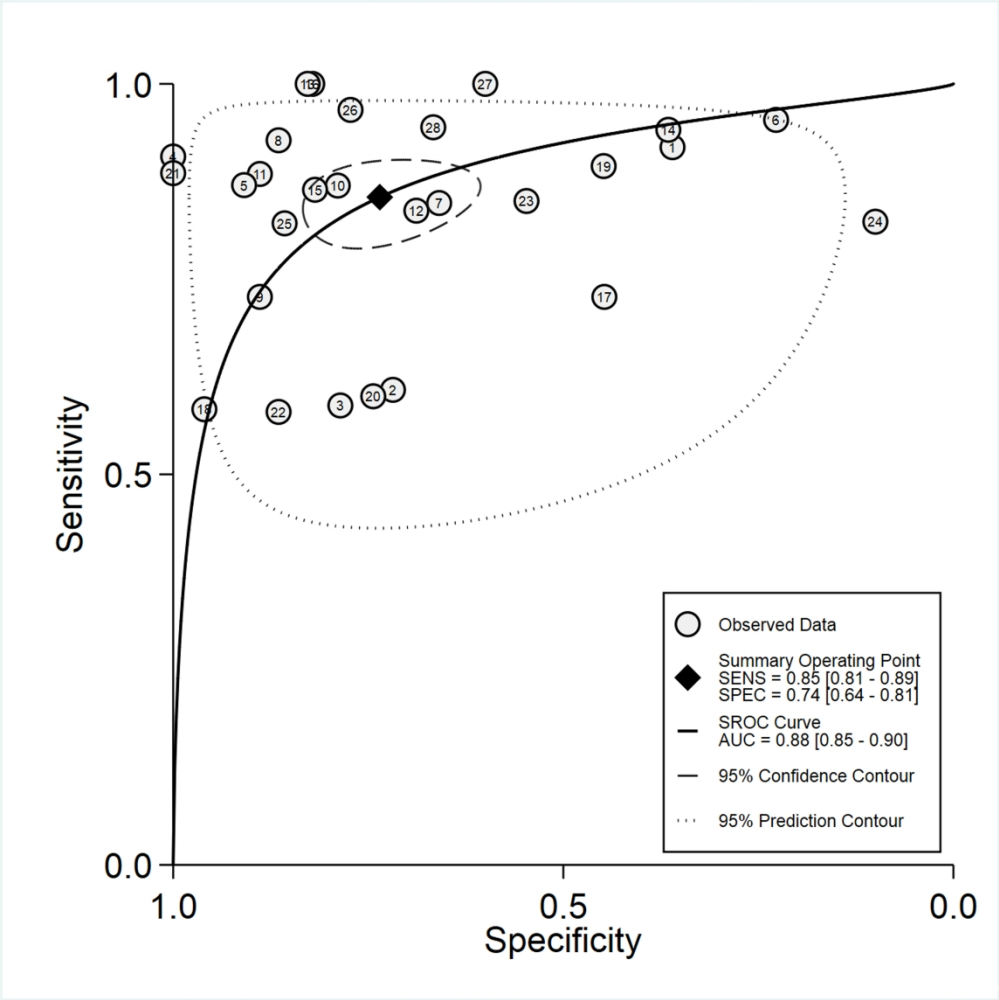

Figure 3 Summary receiver operating characteristic (SROC) graph with 95% confidence region and 95% prediction region for TTE in the diagnosis of pulmonary hypertension (n=28). The study of Habash et al was divided into two independent items, because their case group and control group provided different diagnostic data.

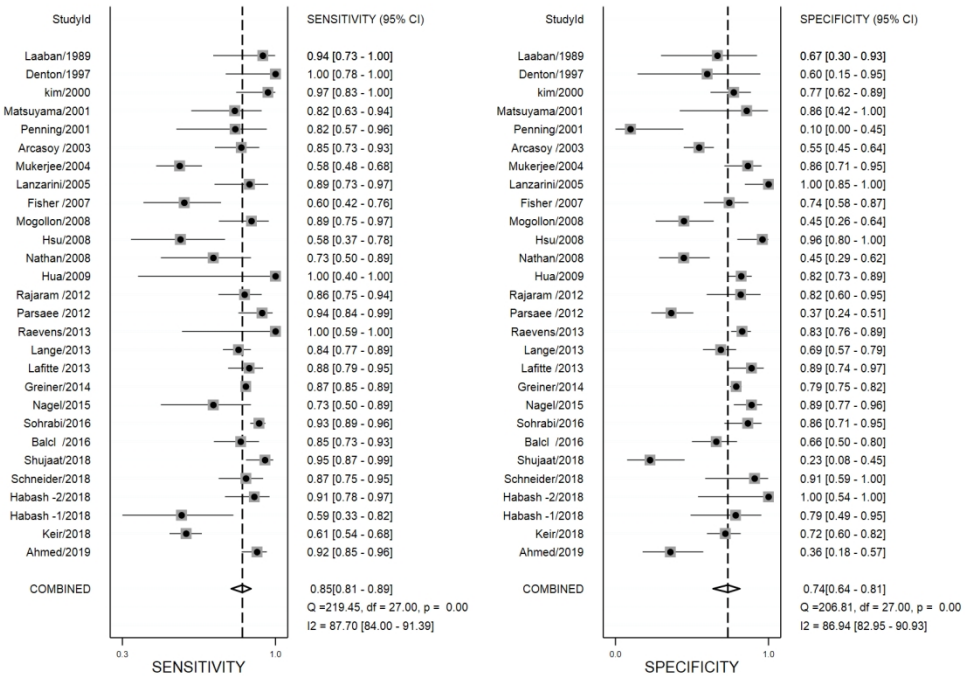

Figure 4 Forrest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I2 statistic for heterogeneity (n=28).

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	5
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Supplementary Data
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Figure 1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	6-7
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	7
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	7
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e)	7

PRISMA-DTA Checklist

		grouping and comparing tests, f) handling of different reference standards	

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	23 Table 1
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	Figure 2
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figure 3
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Figure 4
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	24-26 Table 2,3
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	11
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	12-13
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	2

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: **www.prisma-statement.org**.

BMJ Open

Diagnostic Accuracy of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-033084.R2
Article Type:	Original research
Date Submitted by the Author:	18-Nov-2019
Complete List of Authors:	Ni, Jin-Rong; The First Hospital of Lanzhou University Yan, Pei-Jing; Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital Liu, Shi-Dong; The First Hospital of Lanzhou University Hu, Yuan; the First Hospital of Lanzhou University Yang, Kehu; evidence based meidicine center, lanzhou university Song, Bing; The First Hospital of Lanzhou University Lei, Jun-Qiang; The First Hospital of Lanzhou University
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Radiology and imaging, Qualitative research, Global health, Cardiovascular medicine
Keywords:	Hypertension < CARDIOLOGY, Echocardiography < CARDIOLOGY, Diagnostic radiology < RADIOLOGY & IMAGING

Diagnostic Accuracy of Transthoracic Echocardiography for Pulmonary Hypertension: a Systematic Review and Meta-Analysis

Jin-Rong Ni[1,3,8,9], Pei-Jing Yan[2,5,6#], Shi-Dong Liu[1,3], Yuan Hu[3], Ke-Hu Yang[2,4,5,6], Bing Song[3*], Jun-Qiang Lei[1,7,8,9*]

[1]The First Hospital (the First Clinical Medical School) of Lanzhou University, Lanzhou 730000, China; [2]Institute of Clinical Research and Evidence Based Medicine, Gansu Provincial Hospital, Lanzhou, 730000, China; [3]Department of Cardiovascular Surgery, the First Hospital of Lanzhou University, Lanzhou 730000, China; [4]Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou 730000, China;[5]Evidence-Based Social Science Research Center, Lanzhou University, Lanzhou, 730000, China; [6]Key Laboratory of Evidence-based Medicine and Knowledge Translation of Gansu Province, Lanzhou, 730000, China.[7]Department of Radiology, the First Hospital of Lanzhou University, Lanzhou 730000, China. [8]Intelligent Imaging Medical Engineering Research Center of Gansu province, Lanzhou 730000, China; [9]Precision Image and Collaborative Innovation International Scientific and Technological Cooperation Base of Gansu province, Lanzhou 730000, China.

[#]Joint first author

[*]Joint corresponding authors

Corresponding Author: 1. Jun-Qiang Lei, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613919289040

E-mail: leijunqiangldyy@163.com

2. Bing Song, No.11 Donggang West Road, Chengguan District, Lanzhou, China.

Tel: +8613993118707

E-mail: songbingldyyxwk@163.com

Funding sources: This study was supported by the Key Laboratory of Evidence Based Medicine and Knowledge Translation Foundation of Gansu Province (Grant No. GSXZYZH2018006).

Statement on potential conflicts of interest: None declared.

Patient and Public Involvement: There is no patient or the public involved in our work.

Data Availability: All data relevant to the study are included in the article.

Key Words: Echocardiography; Pulmonary Hypertension; Catheterization; Diagnosis; Accuracy; Meta-analysis.

Abstract

Objective: To evaluate the diagnostic accuracy of transthoracic echocardiography (TTE) in patients with pulmonary hypertension (PH).

Design: Systematic review and meta-analysis.

Data sources and eligibility criteria: EMBASE, Cochrane Library for clinical trials, PubMed and Web of Science were used to search studies from inception to June 19, 2019. Studies using both TTE and right heart catheterization (RHC) to diagnose PH were included.

Main results: A total of 27 studies involving 4386 subjects were considered as eligible for analysis. TTE had a pooled sensitivity of 85%, a pooled specificity of 74%, a pooled positive likelihood ratio of 3.2, a pooled negative likelihood ratio of 0.20, a pooled diagnostic odds ratio of 16, and finally an area under the summary receiver operating characteristic (SROC) curve of 0.88. The subgroup with the shortest time interval between TTE and RHC had the best diagnostic effect, with sensitivity, specificity and AUC of 88%, 90% and 0.94, respectively. TTE had lower sensitivity (81%), specificity (61%) and AUC (0.73) in the subgroup of patients with definite lung diseases. Subgroup analysis also showed that different thresholds of TTE resulted in a different diagnostic performance in the diagnosis of PH.

Conclusion: TTE has a clinical value in diagnosing PH, although it cannot yet replace RHC considered as the gold standard. The accuracy of TTE may be improved by shortening the time interval between TTE and RHC and by developing an appropriate threshold. TTE may not be suitable to assess pulmonary arterial pressure in patients with pulmonary diseases.

Review registration number: PROSPERO CRD42019123289.

Strengths and limitations of this study

1. A comprehensive search was conducted in the main database, more studies were included, and a large sample size was obtained.

2. Detailed subgroup analysis and sensitivity analysis were performed.

3. The types of pulmonary hypertension included in the studies could not be distinguished.

4. Significant heterogeneity in our study limits the interpretation of the results.

Introduction

The prevalence of pulmonary hypertension (PH) is estimated at 1% in the general population, and as high as 10% in the 600 million people older than 65.[1] Early detection and accurate assessment are vital to obtain better outcomes for PH patients.[2] Right heart catheterization (RHC) is the gold standard in the diagnosis of PH,[3] but it is invasive and cannot be used frequently or repeatedly.[4] The latest guideline for PH recommends transthoracic echocardiography (TTE) as a noninvasive test for screening.[3]

High quality meta-analysis has been considered as one of the key tools for achieving evidence.[5 6] Three systematic reviews and meta-analysis regarding the diagnostic accuracy of TTE for PH were published between 2010 and 2013.[7-9] Studies included in these meta-analyses were all published before 2010. In addition, two of them included fewer studies and performed a simple diagnostic data synthesis.[8 9] The other included a relatively large number of studies, but did not assess a detailed subgroup analysis.[7] In recent years, TTE has still been used in the clinical diagnosis of PH, and many new original studies have been published.[10-13] Therefore, the purpose of our study was to undertake a comprehensive systematic review and quantitative meta-analysis on the accuracy of TTE in the diagnosis of PH.

Methods

The present study is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement and the published recommendations.[14 15] The detailed protocol is accessible in PROSPERO (CRD42019123289).[16 17]

Data sources and search

A systematic search in EMBASE, Cochrane Library for clinical trials, PubMed and Web of Science was performed to find the relevant literature from inception to June 19, 2019. Subject words were combined with free words, and the search strategy was developed and adapted for each database. ClinicalTrials.gov and the trials registers on the World Health Organization International Clinical Trials Registry Platform were used to search unpublished trails. The references of the included studies and other systematic reviews and meta-analysis were also reviewed to obtain a comprehensive list of included studies.

Study selection

Studies were selected based on the following inclusion criteria: studies that diagnosed PH by TTE; study population represented by patients with suspected PH; TTE measurement of systolic pulmonary artery pressure (SPAP) performed using tricuspid regurgitation; RHC as the gold standard for the diagnosis of PH.

The exclusion criteria were the following: insufficient data to construct a 2×2 table; studies with less than 20 subjects; duplicate data were used (in this case, the largest sample or the latest study was selected).

Two reviewers (JR.N and PJ.Y) independently screened the eligible studies for suitability. Disagreements were resolved by consensus. If consensus could not be reached, a third reviewer (SD.L) was deferred to arbitration and consensus. No language restriction was applied. If a study was not conducted in the authors' language, a professional translation software could be used.

Data extraction

The data were extracted independently by two reviewers (JR.N and PJ.Y) according to a predefined data extraction sheet. The following variables were extracted from the included studies: lead author, publication year, country of study, study design, study population demographics, sample size, mean age, male ratio, time interval between TTE and RHC, cut-off threshold levels for TTE and RHC, and number of true-positive (TP), false-negative (FN), true-negative (TN) and false-positive (FP) observations. Extracted data were cross-checked and disagreements were resolved via discussion or referral to a third reviewer (Y.H).

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies QUADAS-2 tool was used to assess the risk of bias and clinical applicability concerns of the included studies according to the Cochrane Collaboration recommendation.[18 19] Two reviewers (JR.N and PJ.Y) independently evaluated QUADAS-2 items, and all emerging conflicts were resolved by consensus.

Data synthesis and statistical analysis

Statistical analysis was performed using STATA/SE version 15.1 (Stata Corp, College Station, TX) and Review Manager Version 5.3 software (Copenhagen, Denmark, Nordic Cochrane Centre, Cochrane Collaboration, 2014). All tests were two-tailed. A p value< 0.05 was considered statistically significant.

The correlation coefficient between the logarithm of sensitivity and logarithm of one minus specificity was calculated to test whether the threshold effect was one of the sources of heterogeneity.[20] Deeks' test was used to test for publication bias.[21] The bivariate model for diagnostic meta-analysis was used to obtain pooled estimates of sensitivity and specificity.[22]

Statistical heterogeneity among studies was explored using the I^2 statistic.

Pooled sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), negative likelihood ratio (NLR), and area under the summary receiver operating characteristic (SROC) curve were calculated from the number of TPs, FNs, FPs, and TNs. The 95% confidence interval (CI) was estimated for each metric.

Subgroup analyses were performed based on the following variables: the time interval between TTE and RHC; disease classification of the study population; publication year of the study; study design (prospective or retrospective) and cut-off threshold of TTE to diagnose PH. Sensitivity analysis was undertaken by excluding low-quality studies (according to the QUADAS-2 quality assessment) or trials with characteristics different from the others.

Results

Studies selection and characteristics

Figure 1 shows the PRISMA flow chart of the literature screening. A total of 27 articles involving 4386 subjects met our inclusion criteria (Table 1).[10-13 23-45] Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2).[27]

Of the 27 eligible studies, fourteen (52%) were published between 2010–2019,[10-13 26 27 30 33 34 39 41 43-45] and thirteen (48%) were published before 2010.[23-25 28 29 31 32 35-38 40 42] Twelve (44%) studies were performed in Europe,[12 24 26 32-35 37-39 43 44] nine (30%) in the United States of America (USA),[10 13 23 25 27 28 31 40 42] two (8%) in East Asia,[29 36] three (12%) in the Middle East,[11 41 45] and one (4%) in Australia.[30] Most of the studies (15/27, 56%) [11 12 23 24 28-32 35 36 38 39 41 45]were of prospective design versus 44% (12/27)[10 13 25-27 33 34 37 40 42-44] retrospective.

All included studies used the tricuspid maximal regurgitation velocity (TRVmax) to estimate SPAP; the majority of these studies (23/27, 85%) used the classical method to calculate SPAP: 4TRVmax^2+ right atrial pressure (RAP).[10 11 13 23-28 31-37 39-45] The RAP was calculated through the diameter and collapse rate of the inferior vena cava (IVC) during spontaneous respiration in sixteen(59%) studies,[10 23 25-27 31 33 35-37 39-42 44 45] through the jugular vein pressure (JVP) in one study (4%), [24] and using a fixed value (5 or 10 mm Hg) in three studies (11%). [28 32 34] Three studies (11%) did not report their method for calculating RAP. [11 13 43] Four studies (15%) used a tricuspid gradient (4TRVmax ^{2}) instead of SPAP. [12 29 30 38]

The majority of the studies (22/27, 81%) reported the time interval (mean or maximum) between TTE and RHC, [10-13 23-29 31-35 38 40-42 44 45] while five (5/9, 19%) did not. [30 36 37 39 43] Nine studies (33%) considered time intervals greater than one week,[10 13 24 25 27 31 38 40 42] while thirteen studies (48%) considered time intervals of less than one week.[11 12 23 26 29 32-35 37 39 41 44] The time interval between TTE and RHC ranged from four hours to three months.

Quality Assessment

The quality assessment of the included studies according to the QUADAS-2 inventory is shown in Figure 2. Overall, the quality of the included studies was modest. The included studies were of good quality regarding the applicability concerns, but most of them were of low quality in the risk of bias. In twenty (74%) study protocols, [10-13 23 24 26 28-32 34 35 37-39 41 44 45] consecutive subjects were enrolled, with no inappropriate exclusions. The risk of bias during patient recruitment was unclear in the remaining seven (26%) studies, [25 27 33 36 40 42 43] as patient recruitment was not reported. In six (22%) studies investigators designed the single-blind methods for TTE.[10 12 23 26 39 45] Double blinding in imaging assessment was not mentioned in any

study. The risk of bias on flow and timing between the index test and reference standard was categorized as unclear in 14 (52%) study protocols that did not explicitly state the successful investigation with both index and reference tests in all included patients.[24 30-40 42 43]

Quantitative Analysis

The SROC curve for TTE is shown in Figure 3. Four studies fall within the 95% confidence interval.[11 26 34 44] The AUC was 0.88 (95%CI 0.85–0.90). The pooled sensitivity and specificity for TTE were 85% (95%CI 81%–90%) and 74% (95%CI 64%–81%), respectively (Figure 4). The pooled PLR and NLR were 3.2 (95%CI 2.3–4.4) and 0.20 (95%CI 0.15–0.26), respectively. The pooled DOR for TTE was 16 (95%CI 10–27).

The heterogeneity in our study was significant. The threshold test proved that the threshold effect was not the source of heterogeneity (r=-0.34, *P*=0.12). Deeks' test for funnel plot asymmetry suggested no publication bias (*P*=0.69).The results of the subgroup analysis are presented in Table 2. The sensitivity (87%, 95%CI 81%–91%), specificity (74%, 95%CI 62%–83%) and AUC (0.89, 95%CI 0.86–0.91) of TTE to diagnose PH were higher for studies published in 2010 and later compared to those published before 2010. Among the time interval subgroups, the group with the shortest time interval between TTE and RHC had the best diagnostic effect, with sensitivity, specificity and AUC of 88% (95%CI 73%–95%), 90% (95%CI 53%–99%) and 0.94(95%CI 0.92–0.96), respectively. The disease composition of the study population also affected the diagnostic accuracy of TTE. Compared with patients with other diseases, TTE had lower sensitivity (81%, 95%CI 70%–88%), specificity (61%, 95%CI 53%–69%) and AUC (0.73, 95%CI 0.69–0.77) in the subgroup of patients with definite lung diseases.

Subgroup analysis of different cut-off thresholds to diagnose PH based on TTE showed that the subgroup with a cut-off threshold of 35 mmHg had a higher diagnostic accuracy than that at 40 mmHg. The sensitivity, specificity and AUC of the former were respectively 92% (95%CI 88%–94%), 65% (95%CI 43%–83%) and 0.92 (95%CI 0.89–0.94), while the sensitivity, specificity and AUC at 40 mmHg were 84% (95%CI 75%–91%), 52% (95%CI 31%–71%) and 0.80 (95%CI 76%–83%), respectively.

The sensitivity analysis results are shown in Table 3. After excluding low-quality studies and studies with specific characteristics, the sensitivity analysis did not reveal a source for the heterogeneity in the diagnostic accuracy analysis. Overall, the pooled meta-analysis results for outcomes were in accordance to our sensitivity analyses.

Discussion

Our study found that TTE has a better sensitivity but moderate specificity for the detection of PH. In addition, shortening the time interval between TTE and RHC and developing an appropriate threshold could improve the accuracy of TTE. However, the accuracy of TTE to diagnose PH in patients with lung diseases was low.

Although PH is a chronic disease, we still believe that the shortest possible time interval between TTE and RHC is more favorable. Otherwise, changes in the patient's condition and the application of intervention measures would lead to an increase in the deviation of the results of the two examinations. A detailed subgroup analysis was performed according to the time interval between TTE and RHC. As expected, the diagnostic accuracy was the highest when the time interval was less than or equal to 24 hours. The results also showed that the efficacy of TTE in the diagnosis of PH was gradually reduced with the extension of the time interval.

Subgroup analysis based on the disease composition of the population suggested that the diagnostic accuracy of TTE was lower in patients with lung diseases. Changes associated with chronic pulmonary disease, including a marked increase in intrathoracic gas, consolidation of lung tissue, expansion of the thoracic cage, and alterations in the position of the heart, adversely affect the imaging quality and the parameter measurement of TTE.[46] Therefore, the use of TTE to measure pulmonary pressure in patients with lung diseases might not be an ideal choice.

The Guideline recommend the use of IVC width and collapse rate to estimate RAP, [3] which was not used in some of the included studies. The sensitivity analysis for this point showed that studies which calculated RAP through IVC do not seem to have a higher diagnostic performance. In order to avoid errors caused by RAP estimation, TRVmax was also considered as an indicator to evaluate the possibility of PH. Four studies using TRPG (4TRVmax2) instead of SPAP were grouped into a subgroup and showed that this subgroup had good diagnostic specificity but poor sensitivity.

The sensitivity analysis based on the mean pulmonary artery pressure (MPAP) threshold of 25 mmHg did not result in a higher diagnostic value than the whole, indicating that the overall results were stable. A previous work suggested that a MPAP threshold of 25 mmHg is arbitrarily chosen and lowering it to 20 mmHg (two standard deviations higher than MPAP for the population) is considered a scientific method.[47] However, some scientists insist that it is premature to reduce the MPAP threshold to 20 mmHg because of the risk of over-diagnosis, unclear treatment implications and additional psychological burden on patients.[48] Since none of the study we included used MPAP>20mmHg as the diagnostic threshold for RHC, subgroup analysis on the two thresholds of 20mmHg and 25mmHg could not be performed. Therefore,

we expect that more studies may be performed in the future to verify the appropriate threshold of RHC.

In our review, the cut-off thresholds of SPAP ranged from 30 to 50 mmHg. Subgroup analysis showed that the diagnostic accuracy of the group of 35 mmHg was higher. Sensitivity analysis results of studies that excluded high TTE cut-off value showed that a high cut-off value increased the specificity and reduced the sensitivity of TTE. Due to the small sample size of the subgroup in this study, the value of the cut-off threshold still needs to be determined by further prospective studies of multi-center and large samples.

Subgroup analysis according to the publication year confirmed that studies published after 2010 had only a slightly higher diagnostic accuracy than previous studies. With the improvement of TTE technology and instruments in the past ten years, the diagnostic efficiency of PH has not been significantly improved, which forces us to pay attention to other TTE parameters.[49 50] Perhaps, this could be a new direction for future studies on PH diagnosis.

Limitations:

Several limitations are present in our study. Firstly, the systematic review and meta-analysis is a secondary research method based on original research and the quality of the included study affects the results. In addition, the possibility of missing relevant articles objectively exists, and significant heterogeneity may limit the interpretation of the results. Secondly, the accuracy of echocardiography relies heavily on the operator's ability, experience, and operational discipline. In order to obtain more original studies, we did not consider this aspect as an exclusion criterion. Thirdly, the studies included in this review involve several different types of PH, and some of the included studies do not describe the basic disease and PH type in detail. It is clear that

pulmonary lesions can affect the quality of TTE imaging, leading to underestimated results.

Conclusion

TTE has clinical value in the diagnosis of PH thanks to its better sensitivity and moderate specificity, but it cannot yet replace RHC considered as the gold standard. Shortening the time interval between TTE and RHC and developing an appropriate threshold can improve the accuracy of TTE. TTE may not be suitable to assess pulmonary arterial pressure in patients with pulmonary disease. It may be necessary to combine multiple TTE parameters and conduct multi-center, large-sample studies to further improve the accuracy of TTE in the diagnosis of PH in future research.

Author's contribution

The joint corresponding authors (JQ.L and B.S) are responsible for the design and implementation of the study. SD.L is responsible for the quality control of study selection. Y.H performed the quality control on the links of data extraction. KH.Y provided guidance in literature retrieval and data processing methodology and was responsible for the quality evaluation part. JR.N and PJ.Y performed the systematic review of the literature and extracted the data. JR.N conducted the meta-analyses, and two authors (JR.N, PJ.Y) substantially contributed to the interpretation of the data and wrote the article. All authors repeatedly revised the article. The corresponding authors (JQ.L and B.S) and JR.N take responsibility for the integrity of the analyses.

Reference

1. Hoeper MM, Humbert M, Souza R, et al. A global view of pulmonary hypertension. Lancet Respir Med 2016;4:306-22.

2. Galie N, Rubin L, Hoeper M, et al. Treatment of patients with mildly symptomatic pulmonary arterial hypertension with bosentan (EARLY study): a double-blind, randomised controlled trial. Lancet 2008;371:2093-100.

3. Galie N, Humbert M, Vachiery JL, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. Kardiol Pol 2015;73:1127-206.

4. Hoeper MM, Lee SH, Voswinckel R, et al. Complications of right heart catheterization procedures in patients with pulmonary hypertension in experienced centers. J Am Coll Cardiol 2006;48:2546-52.

5. Tian J, Zhang J, Ge L, et al. The methodological and reporting quality of systematic reviews from China and the USA are similar. J Clin Epidemiol 2017;85:50-58.

6. Xiu-xia L, Ya Z, Yao-long C, et al. The reporting characteristics and methodological quality of Cochrane reviews about health policy research. Health Policy 2015;119:503-10.

7. Janda S, Shahidi N, Gin K, et al. Diagnostic accuracy of echocardiography for pulmonary hypertension: a systematic review and meta-analysis. Heart 2011;97:612-22.

8. Taleb M, Khuder S, Tinkel J, et al. The diagnostic accuracy of Doppler echocardiography in assessment of pulmonary artery systolic pressure: a meta-analysis. Echocardiography 2013;30:258-65.

9. Zhang RF, Zhou L, Ma GF, et al. Diagnostic value of transthoracic Doppler echocardiography in pulmonary hypertension: a meta-analysis. Am J Hypertens 2010;23:1261-64.

10. Ahmed M, Elshinawy O, Agmy G, et al. Tricuspid regurgitation velocity versus right ventricular systolic pressure in the echocardiographic evaluation of pulmonary hypertension. Egypt J Chest Dis Tuberc 2019;68:203-08.

11. Balcı MK, Arı E, Vayvada M, et al. Assessment of Pulmonary Hypertension in Lung Transplantation Candidates: Correlation of Doppler Echocardiography With Right Heart Catheterization. Transplant Proc 2016;48:2797-802.

12. Schneider M, Pistritto AM, Gerges C, et al. Multi-view approach for the diagnosis of pulmonary hypertension using transthoracic echocardiography. Int J Cardiovasc Imaging 2018;34:695-700.

13. Shujaat A, Bajwa AA, Al-Saffar F, et al. Diagnostic accuracy of echocardiography combined with chest CT in pulmonary hypertension. Clin Respir J 2018;12:948-52.

14. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.

15. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;350:g7647.

16. Ge L, Tian JH, Li YN, et al. Association between prospective registration and overall reporting and methodological quality of systematic reviews: a meta-epidemiological study. J Clin Epidemiol 2018;93:45-55.

17. Wang X, Chen Y, Yao L, et al. Reporting of declarations and conflicts of interest in WHO guidelines can be further improved. J Clin Epidemiol 2018;98:1-8.

18. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess 2004;8:iii, 1-234.

19. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.

20. Lee J, Kim KW, Choi SH, et al. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part II. Statistical Methods of Meta-Analysis. Korean J Radiol 2015;16:1188-96.

21. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 2005;58:882-93.

22. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 2005;58:982-90.

23. Arcasoy SM, Christie JD, Ferrari VA, et al. Echocardiographic assessment of pulmonary hypertension in patients with advanced lung disease. Am J Respir Crit Care Med 2003;167:735-40.

24. Denton CP, Cailes JB, Phillips GD, et al. Comparison of Doppler echocardiography and right heart catheterization to assess pulmonary hypertension in systemic sclerosis.

Br J Rheumatol 1997;36:239-43.

25. Fisher MR, Criner GJ, Fishman AP, et al. Estimating pulmonary artery pressures by echocardiography in patients with emphysema. Eur Respir J 2007;30:914-21.

26. Greiner S, Jud A, Aurich M, et al. Reliability of noninvasive assessment of systolic pulmonary artery pressure by Doppler echocardiography compared to right heart catheterization: analysis in a large patient population. J Am Heart Assoc 2014;3:1-8.

27. Habash F, Gurram P, Almomani A, et al. Correlation between echocardiographic pulmonary artery pressure estimates and right heart catheterization measurement in liver transplant candidates. J Cardiovasc Imaging 2018;26:75-84.

28. Hsu VM, Moreyra AE, Wilson AC, et al. Assessment of pulmonary arterial hypertension in patients with systemic sclerosis: comparison of noninvasive tests with results of right-heart catheterization. J Rheumatol 2008;35:458-65.

29. Hua R, Sun YW, Wu ZY, et al. Role of 2-dimensional Doppler echo-cardiography in screening portopulmonary hypertension in portal hypertension patients. Hepatobiliary Pancreat. Hepatobiliary Pancreat Dis Int 2009;8:157-61.

30. Keir GJ, Wort SJ, Kokosi M, et al. Pulmonary hypertension in interstitial lung disease: Limitations of echocardiography compared to cardiac catheterization. Respirology 2018;23:687-94.

31. Kim WR, Krowka MJ, Plevak DJ, et al. Accuracy of doppler echocardiography in the assessment of pulmonary hypertension in liver transplant candidates. Liver Transpl 2000;6:453-58.

32. Laaban JP, Diebold B, Zelinski R, et al. Noninvasive Estimation of Systolic

Pulmonary-Artery Pressure Using Doppler Echocardiography in Patients with Chronic Obstructive Pulmonary-Disease. Chest 1989;96:1258-62.

33. Lafitte S, Pillois X, Reant P, et al. Estimation of pulmonary pressures and diagnosis of pulmonary hypertension by doppler echocardiography: A retrospective comparison of routine echocardiography and invasive hemodynamics. J Am Soc Echocardiogr 2013;26:457-63.

34. Lange TJ, Baumgartner S, Arzt M, et al. Qualitative echocardiography parameters for prediction of pulmonary hypertension. Int J Clin Pract 2013;67:5-12.

35. Lanzarini L, Fontana A, Campana C, et al. Two simple echo-doppler measurements can accurately identify pulmonary hypertension in the large majority of patients with chronic heart failure. J Heart Lung Transplant 2005;24:745-54.

36. Matsuyama W, Ohkubo R, Michizono K, et al. Usefulness of transcutaneous Doppler jugular venous echo to predict pulmonary hypertension in COPD patients. Eur Respir J 2001;17:1128-31.

37. Mogollon Jimenez MV, Escoresca Ortega AM, Cabeza Letran ML, et al. Correlation of Echocardiographic and Hemodynamic Parameters in Pulmonary Hypertension Assessment Prior to Heart Transplantation. Transplant Proc 2008;40:3023-24.

38. Mukerjee D, St George D, Knight C, et al. Echocardiography and pulmonary function as screening tests for pulmonary arterial hypertension in systemic sclerosis. Rheumatology (Oxford) 2004;43:461-6.

39. Nagel C, Henn P, Ehlken N, et al. Stress Doppler echocardiography for early detection of systemic sclerosis-associated pulmonary arterial hypertension. Arthritis Res Ther

2015;17:165.

40. Nathan SD, Shlobin OA, Barnett SD, et al. Right ventricular systolic pressure by echocardiography as a predictor of pulmonary hypertension in idiopathic pulmonary fibrosis. Respir Med 2008;102:1305-10.

41. Parsaee M, Amin A, Nematollahi MR, et al. Comparison of transthoracic echocardiography and right heart catheterization for assessing pulmonary arterial pressure in patients with congenital or valvular heart defects. Iran Heart J 2012;12:54-61.

42. Penning S, Robinson KD, Major CA, et al. A comparison of echocardiography and pulmonary artery catheterization for evaluation of pulmonary artery pressures in pregnant patients with suspected pulmonary hypertension. Am J Obstet Gynecol 2001;184:1568-70.

43. Raevens S, Colle I, Reyntjens K, et al. Echocardiography for the detection of portopulmonary hypertension in liver transplant candidates: An analysis of cutoff values. Liver Transpl 2013;19:602-10.

44. Rajaram S, Swift AJ, Capener D, et al. Comparison of the diagnostic utility of cardiac magnetic resonance imaging, computed tomography, and echocardiography in assessment of suspected pulmonary arterial hypertension in patients with connective tissue disease. J Rheumatol 2012;39:1265-74.

45. Sohrabi B, Kazemi B, Mehryar A, et al. Correlation between pulmonary artery pressure measured by echocardiography and right heart catheterization in patients with rheumatic mitral valve stenosis (A prospective study). Echocardiography 2016;33:7-

13.

46. Galderisi M, Cosyns B, Edvardsen T, et al. Standardization of adult transthoracic echocardiography reporting in agreement with recent chamber quantification, diastolic function, and heart valve disease recommendations: an expert consensus document of the European Association of Cardiovascular Imaging. Eur Heart J Cardiovasc Imaging 2017;18:1301-10.

47. Simonneau G, Montani D, Celermajer DS, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. Eur Respir J 2019;53.

48. Gibbs JSR, Torbicki A. Proposed new pulmonary hypertension definition: is 4 mm(Hg) worth re-writing medical textbooks? Eur Respir J 2019;53.

49. Misra D, Kendes A, Sulica R, et al. Exercise-induced pulmonary hypertension by stress echocardiography: Prevalence and correlation with right heart hemodynamics. Int J Cardiol 2017;228:518-22.

50. Vitarelli A, Mangieri E, Terzano C, et al. Three-dimensional echocardiography and 2D-3D speckle-tracking imaging in chronic pulmonary hypertension: diagnostic accuracy in detecting hemodynamic signs of right ventricular (RV) failure. J Am Heart Assoc 2015;4:e001584.

Table 1 Characteristics of each study included in this meta-analysis.

Study	Year	Country	Design	N	Disease Composition of the Population	Mean Age (Years)	Male (%)	Time Interval	TTE Threshold (mmHg)	RHC Threshold (mmHg)	TTE Method
Ahmed	2019	USA	Retrospective	136	Multiple diseases	59±20	35	<3mo	SPAP≥40	MPAP≥25	4TRVmax^2+RAP(IVC)
Keir	2018	Australia	Prospective	265	Interstitial lung disease	60.8±16.5	46	–	TRPG>46	MPAP≥25	4TRVmax^2
Habash-1	2018	USA	Retrospective	31	Liver transplantation candidates	57±11	42	36.8±13.4d	SPAP>47	MPAP≥25	4TRVmax^2+RAP(IVC)
Habash-2	2018	USA	Retrospective	49	Multiple diseases	59±15	31	16.0±11.6d	SPA>43	MPAP≥25	4TRVmax^2+RAP(IVC)
Schneider	2018	Austria	Prospective	65	Cardiac and lung diseases	67.2	43	<48h	TRPG>32	MPAP≥25	4TRVmax^2
Balcl	2016	Turkey	Prospective	103	Lung transplantation candidates	47.6±10.4	66	<72h	SPAP>35	MPAP≥25	4TRVmax^2+RAP (NR)
Shujaat	2016	USA	Retrospective	87	Multiple diseases	54.3±15.9	29	13d*	SPAP>40	MPAP>25	4TRVmax^2+RAP (NR)
Sohrabi	2016	Iran	Prospective	300	Rheumatic mitral stenosis	59.9	31	<24h	SPAP≥35	MPAP≥25	4TRVmax^2+RAP(IVC)
Nagel	2015	Germany	Prospective	76	Systemic sclerosis	58±14	16	–	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Greiner	2014	Germany	Retrospective	1695	Cardiac disease	63±15	67	<5d	SPAP≥36	MPAP≥25	4TRVmax^2+RAP(IVC)
Lafitte	2013	France	Retrospective	114	Cardiac and lung disease	64.8±15.9	52	<48h	SPAP≥38	MPAP>25	4TRVmax^2+RAP(IVC)
Lange	2013	Germany	Retrospective	231	Multiple diseases	62±13	43	5±4d	SPAP>50	MPAP≥25	4TRVmax^2+RAP (5)
Raevens	2013	Belgium	Retrospective	152	Liver transplantation candidates	58±11	66	–	SPAP>38	MPAP≥25	4TRVmax^2+RAP (NR)
Parsaee	2012	Iran	Prospective	103	Cardiac diseases	41.0±15.8	44	<4h	SPAP≥35	MPAP>25	4TRVmax^2+RAP(IVC)
Rajaram	2012	UK	Retrospective	81	Connective tissue disease	62±14	15	<48h	TRPG≥40	MPAP≥25	4TRVmax^2
Hua	2009	China	Prospective	105	Liver transplantation candidates	49.5±11.8	63	4.2±2.0d	SPAP≥30	MPAP≥25	4TRVmax^2+RAP(IVC)
Nathan	2008	USA	Retrospective	60	Idiopathic pulmonary fibrosis	62.9±8.6	55	32±78d	SPAP≥40	MPAP>25	4TRVmax^2+RAP(IVC)
Hsu	2008	USA	Prospective	49	Systemic Sclerosis	55	18	<4h	SPAP>47	MPAP≥25	4TRVmax^2+RAP (10)
Mogollon	2008	Spain.	Retrospective	67	Heart transplantation candidates	–	–	–	SPAP>40	MPAP>35	4TRVmax^2+RAP(IVC)
Fisher	2007	USA	Retrospective	63	Emphysema patients	65.6±6.6	60	23d	SPAP>40	MPAP≥25	4TRVmax^2+RAP(IVC)
Lanzarini	2005	Italy	Prospective	57	Heart failure	52±11	74	<24h	SPAP≥32	SPAP≥35	4TRVmax^2+RAP(IVC

Mukerjee	2004	UK	Prospective	137	Systemic sclerosis	63	–	<3 mo	TRPG>40	MPAP≥25	4TRVmax²
Arcasoy	2003	USA	Prospective	166	COPD 68%, ILD 28%, PVD 4%	51	43	<72h	SPAP≥45	SPAP≥45	4TRVmax²+RAP(IVC)
Penning	2001	USA	Retrospective	27	Pregnant women with cardiac diseases	28.6	0	25.8d	SPAP≥40	SPAP≥35	4TRVmax²+RAP(IVC)
Matsuyama	2001	Japan	Prospective	35	COPD	66	94	–	SPAP≥40	MPAP>25	4TRVmax²+RAP(IVC)
Kim	2000	USA	Prospective	74	Liver transplantation candidates	54	50	59d	SPAP>50	MPAP≥35	4TRVmax²+RAP(IVC)
Denton	1997	UK	Prospective	20	COPD	48.6±11.7	30	1.8±2.3mo	SPAP≥30	SPAP≥30	4TRVmax²+RAP(JVP)
Laaban	1989	France	Prospective	27	COPD	63±9	78	<2d	SPAP≥35	SPAP≥35	4TRVmax²+RAP(5)

USA, United States of America; UK, United Kingdom of Great Britain and Northern Ireland; TTE, Transthoracic echocardiography; RHC, right heart catheterization; SPAP, systolic pulmonary artery pressure; MPAP, mean pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient; TRV, tricuspid regurgitation velocity; RAP, right atrial pressure;IVC, Inferior vena cava; JVP, jugular vein pressure; COPD, chronic obstructive pulmonary disease; ILD, Interstitial lung disease; PVD, peripheral vascular disease; NR, not reported.

* The median time (other terms are mean time)

Table 2 Subgroup analysis

Group	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
All Studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
Time Interval								
≤24h	4	95(90–99)	0.94(0.92–0.96)	0.88(0.73–0.95)	0.90(0.53–0.99)	8.9(1.5–54.5)	0.13(0.06–0.29)	68(13–348)
≤48h*	7	95(90–99)	0.94(0.91–0.95)	0.88(0.81–0.93)	0.89(0.71–0.96)	7.8(2.8–21.3)	0.13(0.09–0.21)	59(23–148)
≤72h *	9	94(89–99)	0.91(0.89–0.93)	0.87(0.82–0.91)	0.83(0.65–0.93)	5.2(2.4–11.2)	0.15(0.11–0.21)	34(14–82)
≤1 week	13	93(87–99)	0.91(0.88–0.93)	0.87(0.84–0.90)	0.80(0.68–0.88)	4.3(2.7–6.9)	0.16(0.12–0.21)	27(15–48)
>1 week	10	97(95–99)	0.82(0.78–0.85)	0.85(0.73–0.92)	0.60(0.40–0.77)	2.1(1.3–3.4)	0.25(0.14–0.45)	9(4–21)
Unclear	5	82(63–100)	0.85(0.810.88)	0.79(0.63–0.99)	0.76(0.61–0.87)	3.4(1.9–5.9)	0.27(0.15–0.51)	12(5–33)
Population Disease								
Cardiac diseases	6	94(89–99)	0.90(0.87–0.92)	0.90(0.86–0.93)	0.67(0.29–0.91)	2.7(0.9–8.1)	0.15(0.08–0.30)	18(3–95)
Lung diseases	8	90(81–100)	0.73(0.69–9.77)	0.81(0.70–0.88)	0.61(0.53–0.69)	2.1(1.8–2.4)	0.32(0.21–0.48)	7(4–10)
Multiple diseases#	6	93(87–99)	0.90(0.87–0.92)	0.89(0.84–0.92)	0.70(0.40–0.89)	3.0(1.3–7.1)	0.16(0.11–0.23)	19(6–60)
Unclear&	8	88(77–100)	0.88(0/85–0.90)	0.80(0.64–0.90)	0.85(0.80–0.89)	5.3(4.0–7.0)	0.23(0.12–0.45)	23(10–51)
Published Year								
≥2010	15	97(95–99)	0.89(0.86–0.91)	0.87(0.81–0.91)	0.74(0.62–0.83)	3.3(2.3–4.9)	0.18(0.13–0.25)	19(11–13)
<2010	13	96(93–99)	0.86(0.83–0.89)	0.84(0.74–0.90)	0.73(0.56–0.85)	3.1(1.8–5.3)	0.22(0.14–0.37)	14(6–33)
Study Design								
Prospective	15	97(95–99)	0.90(0.87–0.92)	0.86(0.77–0.91)	0.79(0.69–0.87)	4.2(2.7–6.4)	0.18(0.11–0.28)	23(12–45)
Retrospective	13	96(92–99)	0.86(0.83–0.89)	0.86(0.80–0.90)	0.65(0.49–0.78)	2.5(1.6–3.7)	0.22(0.15–0.32)	11(6–22)
TTE Threshold								

SPAP≥40 mmHg	8	96(93–99)	0.80(0.76–0.83)	0.84(0.75–0.91)	0.52(0.31–0.71)	1.7(1.2–2.5)	0.30(0.21–0.44)	6(3–11)
SPAP≥35 mmHg	4	76(47–100)	0.92(0.890.94)	0.92(0.88–0.94)	0.65(0.43–0.83)	2.6(1.4–4.9)	0.13(0.08–0.22)	16(9–28)
TRPG	4	0(0–100)	0.85(0.82–0.88)	0.75(0.58–0.86)	0.81(0.70–0.89)	4.0(2.2–7.3)	0.31(0.17–0.57)	13(4–40)

AUC, Area under the curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; TTE, transthoracic echocardiography; SPAP, systolic pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient.
*Studies with time intervals less than or equal to 24 hours were included in this subgroup.

* Studies with time intervals less than or equal to 24 hours and 48 hours were included in this subgroup.

Studies including a variety of diseases, including heart disease and lung disease.
&Diseases were not specifically identified in the studies (transplant candidates).

Table 3 Sensitivity analysis for diagnostic accuracy meta-analysis.

Study characteristic	N	I^2 (95%CI)	AUC (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PLR (95%CI)	NLR ((95%CI)	DOR (95%CI)
All included studies	28	98(97–99)	0.88(0.85–0.90)	0.85(0.81–0.90)	0.74(0.64–0.81)	3.2(2.3–4.4)	0.20(0.15–0.26)	16(10–27)
Excluding study of Penning*	27	98(97–99)	0.88(0.85–0.91)	0,86(0.81–0.89)	0.75(0.66–0.82)	3.4(2.5–4.6)	0.19(0.14–0.26)	18(11–28)
RHC threshold MPAP≥25 mmHg	21	98(97–99)	0.87(0.84–0.90)	0.83(0.77–0.88)	0.76(0.67–0.83)	3.5(2.5–4.8)	0.22(0.16–0.30)	16(10–26)
RAP method(IVC)£	17	96(93–99)	0.89(0.86–0.91)	0.86(0.82–0.90)	0.73(0.59–0.84)	3.2(2.0–5.1)	0.19(0.13–0.27)	17(8–35)
Excluding high TTE threshold *	21	97(95–99)	0.90(0.87–0.92)	0.88(0.85–0.91)	0.72(0.59–0.82)	3.2(2.1–4.8)	0.16(0.12–0.22)	20(11–36)

AUC, Area under the curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DOR, diagnostic odds ratio; MPAP, mean pulmonary artery pressure; TTE, transthoracic echocardiography; RHC, right heart catheterization;RAP, right atrial pressure; IVC, inferior vena cava.

* The study of Penning was excluded because only pregnant women withcardiac disease were included as subjects.

£The RAP was calculated through the diameter and collapse rate of IVC during spontaneous respiration. Habash's study was divided into two independent parts, thus the results section showed 16 studies, but 17 sets of data were analyzed.

* High TTE threshold was defined as SPAP greater than 45 mmHg or tricuspid regurgitation pressure gradient (TRPG) greater than 40 mmHg.

Figure legends/captions

Figure 1 Flowchart for identification of the studies.

*Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2). A total of 27 studies were included, but 28 sets of data were analyzed.

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements regarding each domain for each included study (n=28).

Figure 3 Summary receiver operating characteristic (SROC) graph with 95% confidence region and 95% prediction region for TTE in the diagnosis of pulmonary hypertension (n=28).

Figure 4 Forest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I^2 statistic for heterogeneity (n=28).

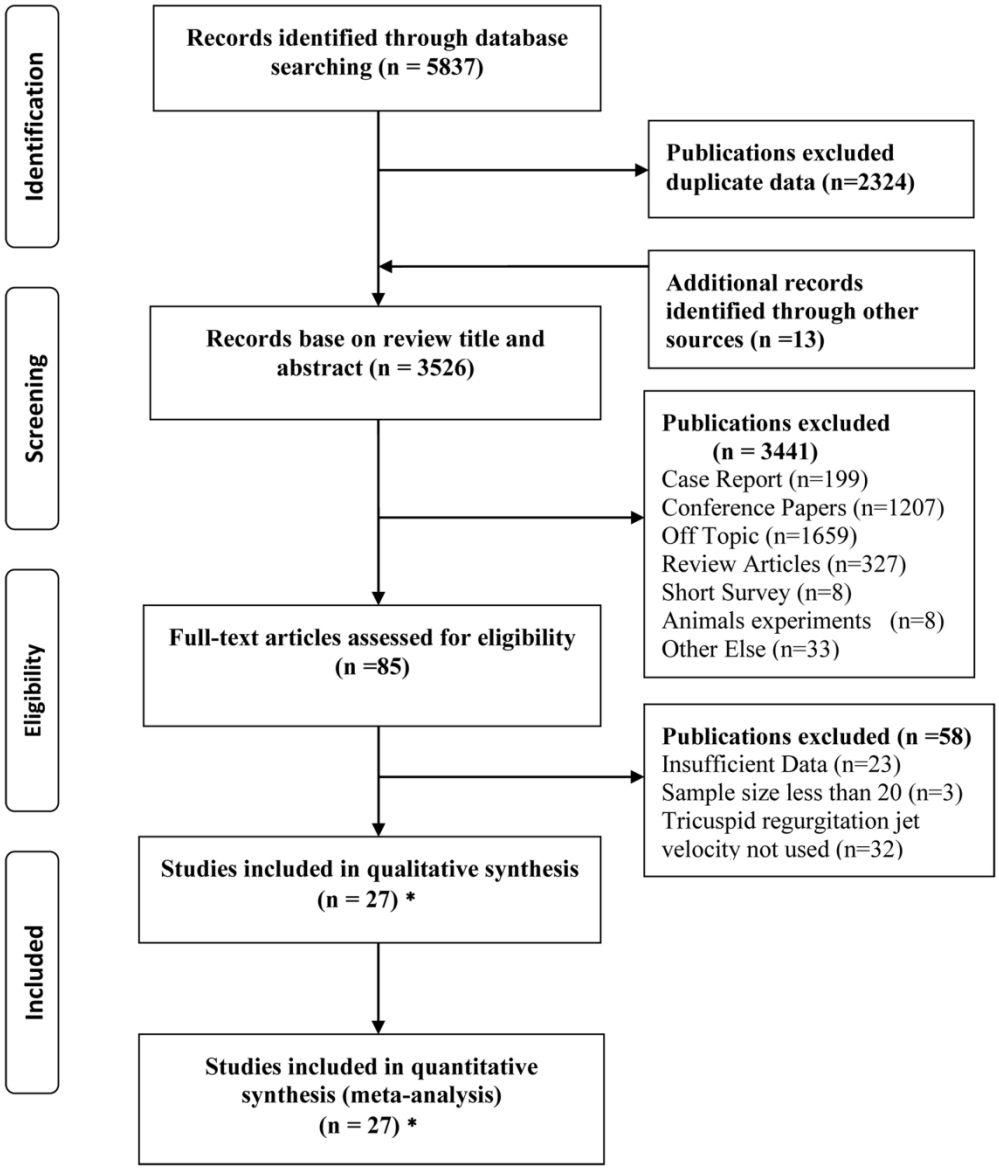

Figure 1 Flowchart for identification of the studies.

*Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2). A total of 27 studies were included, but 28 sets of data were analyzed.

157x188mm (300 x 300 DPI)

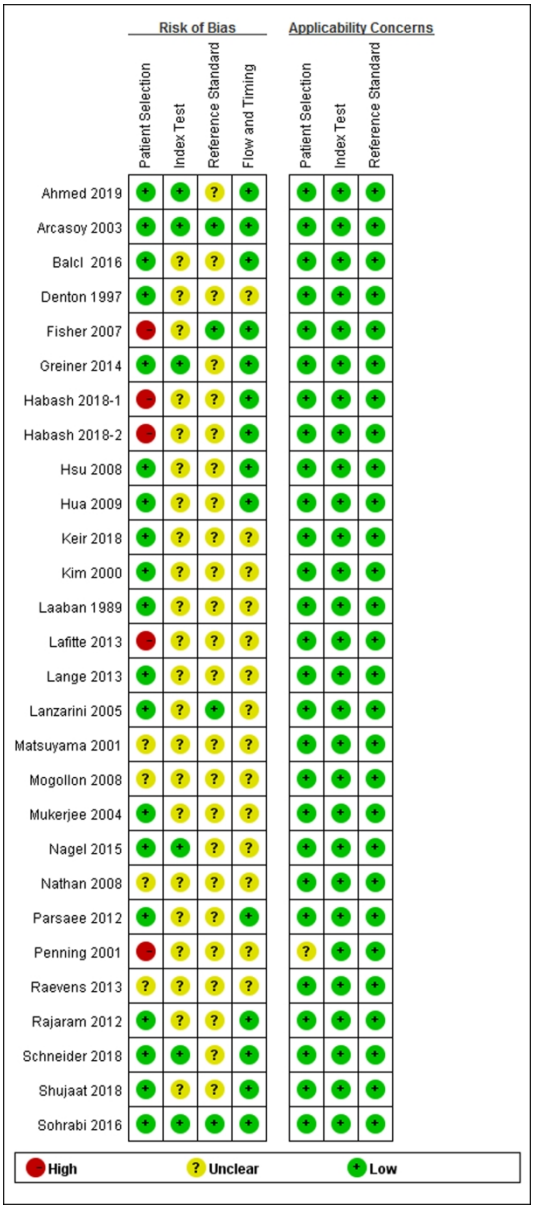

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements regarding each domain for each included study (n=28).

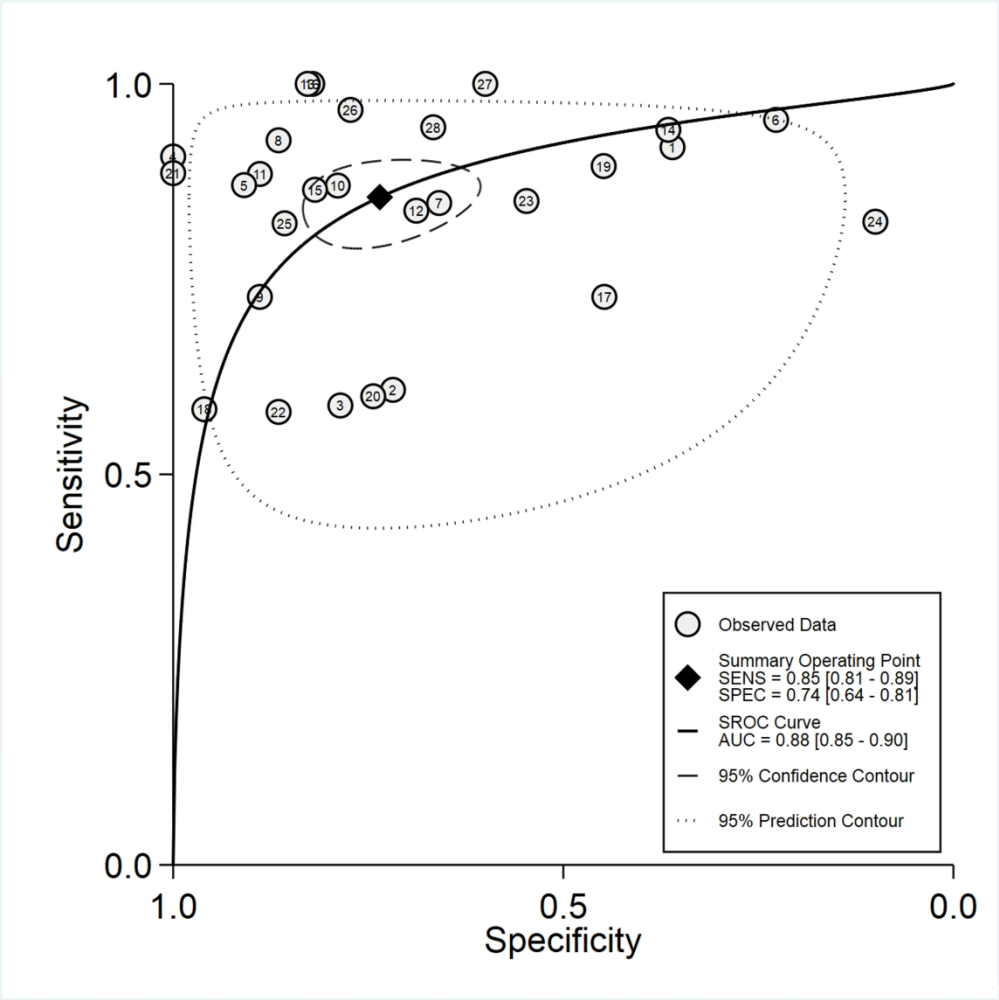

Figure 3 Summary receiver operating characteristic (SROC) graph with 95% confidence region and 95% prediction region for TTE in the diagnosis of pulmonary hypertension (n=28).

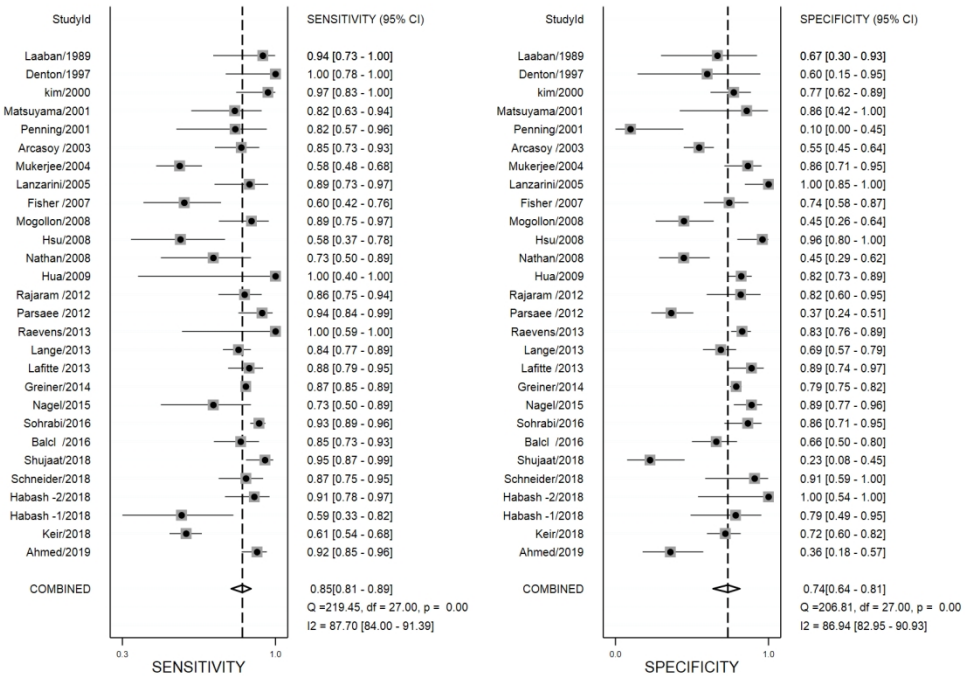

Figure 4 Forest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I2 statistic for heterogeneity (n=28).

PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
TITLE / ABSTRACT			
Title	1	Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies.	1
Abstract	2	Abstract: See PRISMA-DTA for abstracts.	3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Clinical role of index test	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design).	5
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).	5
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.	Supplementary Data
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Figure 1
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (e.g. study design, clinical setting).	6-7
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.	7
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity) and state the unit of assessment (e.g. per-patient, per-lesion).	7
Synthesis of results	14	Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to: a) handling of multiple definitions of target condition. b) handling of multiple thresholds of test positivity, c) handling multiple index test readers, d) handling of indeterminate test results, e)	7

PRISMA-DTA Checklist

		grouping and comparing tests, f) handling of different reference standards	

Section/topic	#	PRISMA-DTA Checklist Item	Reported on page #
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.	7
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	8
RESULTS			
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each included study provide citations and present key characteristics including: a) participant characteristics (presentation, prior testing), b) clinical setting, c) study design, d) target condition definition, e) index test, f) reference standard, g) sample size, h) funding sources	23 Table 1
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.	Figure 2
Results of individual studies	20	For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report 2x2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.	Figure 3
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.	Figure 4
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events).	24-26 Table 2,3
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence.	11
Limitations	25	Discuss limitations from included studies (e.g. risk of bias and concerns regarding applicability) and from the review process (e.g. incomplete retrieval of identified research).	13
Conclusions	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test).	12-13
FUNDING			
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.	2

Adapted From: McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, The PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. JAMA. 2018 Jan 23;319(4):388-396. doi: 10.1001/jama.2017.19163.

For more information, visit: **www.prisma-statement.org**.