BMJ Open

Diagnostic accuracy of transthoracic echocardiography for pulmonary hypertension: a systematic review and meta-analysis

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

To cite: Ni J-R, Yan P-J, Liu S-D, *et al*. Diagnostic accuracy of transthoracic echocardiography for pulmonary hypertension: a systematic review and meta-analysis. *BMJ Open* 2019;**9**:e033084. doi:10.1136/bmjopen-2019-033084

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-033084).

J-RN and P-JY are joint first authors.

Received 20 July 2019
Revised 18 November 2019
Accepted 29 November 2019

For numbered affiliations see end of article.

Correspondence to
Dr Jun-Qiang Lei;
leijunqiangldyy@163.com

Bing Song;
songbingldyyxwk@163.com

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 19 June, 2019. Studies using both TTE and right heart catheterisation (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 OR of 16 and finally an area under the summary receiver operating characteristic curve of 0.88. The subgroup with the shortest time interval between TTE and RHC had the best diagnostic effect, with sensitivity, specificity and area under the curve (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.

PROSPERO registration number PROSPERO CRD42019123289.

Strengths and limitations of this study

- A comprehensive search was conducted in the main database, more studies were included and a large sample size was obtained.
- Detailed subgroup analysis and sensitivity analysis were performed.
- The types of pulmonary hypertension included in the studies could not be distinguished.
- 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 catheterisation (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 non-invasive 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

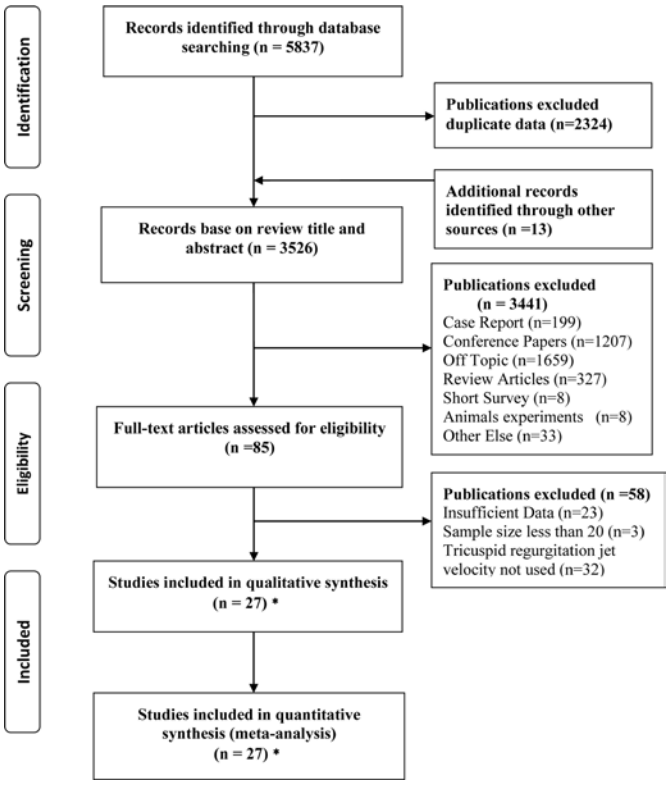

Figure 1 Flowchart for identification of the studies. *Habash's study was divided into two independent parts because of the differences between the case group (Habash-1) and the control group (Habash-2). A total of 27 studies were included, but 28 sets of data were analysed.

recommendations.[14 15] The detailed protocol is accessible in PROSPERO.[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 19 June, 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 WHO 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 (J-RN and P-JY) independently screened the eligible studies for suitability. Disagreements were resolved by consensus. If consensus could not be reached, a third reviewer (S-DL) 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 (J-RN and P-JY) 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 (YH).

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 (J-RN and P-JY) 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 V.15.1 (StataCorp, College Station, Texas) and Review Manager V.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 OR (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% 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

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 (mm Hg)	RHC threshold (mm Hg)	TTE method
Ahmed *et al*[10]	2019	USA	Retrospective	136	Multiple diseases	59±20	35	<3 months	SPAP ≥40	MPAP ≥25	4TRVmax^2+RAP (IVC)
Keir *et al*[30]	2018	Australia	Prospective	265	Interstitial lung disease	60.8±16.5	46	–	TRPG >46	MPAP ≥25	4TRVmax^2
Habash-1 et al[27]	2018	USA	Retrospective	31	Liver transplantation candidates	57±11	42	36.8±13.4 days	SPAP >47	MPAP ≥25	4TRVmax^2+RAP (IVC)
Habash-2 et al[27]	2018	USA	Retrospective	49	Multiple diseases	59±15	31	16.0±11.6 days	SPAP >43	MPAP ≥25	4TRVmax^2+RAP (IVC)
Schneider et al[12]	2018	Austria	Prospective	65	Cardiac and lung diseases	67.2	43	<48 hour	TRPG >32	MPAP ≥25	4TRVmax^2
Balcl et al[11]	2016	Turkey	Prospective	103	Lung transplantation candidates	47.6±10.4	66	<72 hour	SPAP >35	MPAP ≥25	4TRVmax^2+RAP (NR)
Shujaat et al[13]	2016	USA	Retrospective	87	Multiple diseases	54.3±15.9	29	13 days*	SPAP >40	MPAP >25	4TRVmax^2+RAP (NR)
Sohrabi *et al*[45]	2016	Iran	Prospective	300	Rheumatic mitral stenosis	59.9	31	<24 hour	SPAP ≥35	MPAP ≥25	4TRVmax^2+RAP (IVC)
Nagel[39]	2015	Germany	Prospective	76	Systemic sclerosis	58±14	16	–	SPAP >40	MPAP ≥25	4TRVmax^2+RAP (IVC)
Greiner *et al*[26]	2014	Germany	Retrospective	1695	Cardiac disease	63±15	67	<5 days	SPAP ≥36	MPAP ≥25	4TRVmax^2+RAP (IVC)
Lafitte *et al*[33]	2013	France	Retrospective	114	Cardiac and lung disease	64.8±15.9	52	<48 hour	SPAP ≥38	MPAP >25	4TRVmax^2+RAP (IVC)
Lange *et al*[34]	2013	Germany	Retrospective	231	Multiple diseases	62±13	43	5±4 days	SPAP >50	MPAP ≥25	4TRVmax^2+RAP (5)
Raevens *et al*[43]	2013	Belgium	Retrospective	152	Liver transplantation candidates	58±11	66	–	SPAP >38	MPAP ≥25	4TRVmax^2+RAP (NR)
Parsaee *et al*[41]	2012	Iran	Prospective	103	Cardiac diseases	41.0±15.8	44	<4 hour	SPAP ≥35	MPAP >25	4TRVmax^2+RAP (IVC)
Rajaram *et al*[44]	2012	UK	Retrospective	81	Connective tissue disease	62±14	15	<48 hour	TRPG ≥40	MPAP ≥25	4TRVmax^2
Hua *et al*[29]	2009	China	Prospective	105	Liver transplantation candidates	49.5±11.8	63	4.2±2.0 days	SPAP ≥30	MPAP ≥25	4TRVmax^2+RAP (IVC)
Nathan *et al*[40]	2008	USA	Retrospective	60	Idiopathic pulmonary fibrosis	62.9±8.6	55	32±78 days	SPAP ≥40	MPAP >25	4TRVmax^2+RAP (IVC)
Hsu *et al*[28]	2008	USA	Prospective	49	Systemic sclerosis	55	18	<4 hour	SPAP >47	MPAP ≥25	4TRVmax^2+RAP (10)
Mogollón *et al*[37]	2008	Spain.	Retrospective	67	Heart transplantation candidates	–	–	–	SPAP >40	MPAP >35	4TRVmax^2+RAP (IVC)
Fisher *et al*[25]	2007	USA	Retrospective	63	Emphysema patients	65.6±6.6	60	23 days	SPAP >40	MPAP ≥25	4TRVmax^2+RAP (IVC)
Lanzarini *et al*[35]	2005	Italy	Prospective	57	Heart failure	52±11	74	<24 hour	SPAP ≥32	SPAP ≥35	4TRVmax^2+RAP (IVC)
Mukerjee *et al*[38]	2004	UK	Prospective	137	Systemic sclerosis	63	–	<3 months	TRPG >40	MPAP ≥25	4TRVmax^2
Arcasoy *et al*[23]	2003	USA	Prospective	166	COPD 68%, ILD 28%, PVD 4%	51	43	<72 hour	SPAP ≥45	SPAP ≥45	4TRVmax^2+RAP (IVC)
Penning *et al*[42]	2001	USA	Retrospective	27	Pregnant women with cardiac diseases	28.6	0	25.8 days	SPAP ≥40	SPAP ≥35	4TRVmax^2+RAP (IVC)
Matsuyama *et al*[36]	2001	Japan	Prospective	35	COPD	66	94	–	SPAP ≥40	MPAP >25	4TRVmax^2+RAP (IVC)
Kim *et al*[31]	2000	USA	Prospective	74	Liver transplantation candidates	54	50	59 days	SPAP >50	MPAP ≥35	4TRVmax^2+RAP (IVC)
Denton *et al*[24]	1997	UK	Prospective	20	COPD	48.6±11.7	30	1.8±2.3 months	SPAP ≥30	SPAP ≥30	4TRVmax^2+RAP (JVP)
Laaban *et al*[32]	1989	France	Prospective	27	COPD	63±9	78	<2 days	SPAP ≥35	SPAP ≥35	4TRVmax^2+RAP (5)

*The median time (other terms are mean time).
COPD, chronic obstructive pulmonary disease; ILD, interstitial lung disease;IVC, inferior vena cava; JVP, jugular vein pressure; MPAP, mean pulmonary artery pressure; NR, not reported; PVD, peripheral vascular disease; RAP, right atrial pressure; RHC, right heart catheterisation; SPAP, systolic pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient; TRV, tricuspid regurgitation velocity; TTE, transthoracic echocardiography; UK, United Kingdom of Great Britain and Northern Ireland;USA, United States of America.

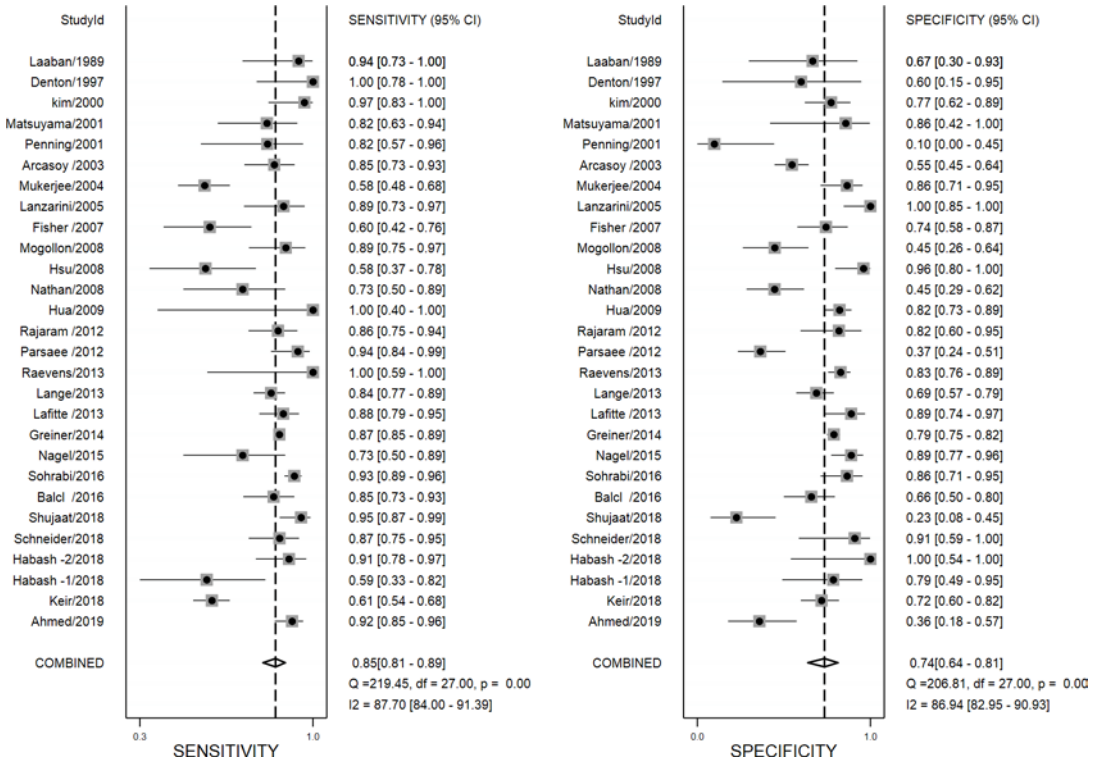

Figure 2 Risk of bias and applicability concerns summary: review authors' judgements regarding each domain for each included study (n=28).

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, 14 (52%) were published between 2010 to 2019,[10–13 26 27 30 33 34 39 41 43–45] and 13 (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 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 16 (59%) studies,[10 23 25–27 31 33 35–37 39–42 44 45] through the jugular vein pressure 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 1 week,[10 13 24 25 27 31 38 40 42] while 13 studies (48%) considered time intervals of less than 1 week.[11 12 23 26 29 32–35 37 39 41 44] The time interval between TTE and RHC ranged from 4 hours to 3 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 20 (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 categorised as unclear in 14 (52%) study protocols that did not explicitly state the successful

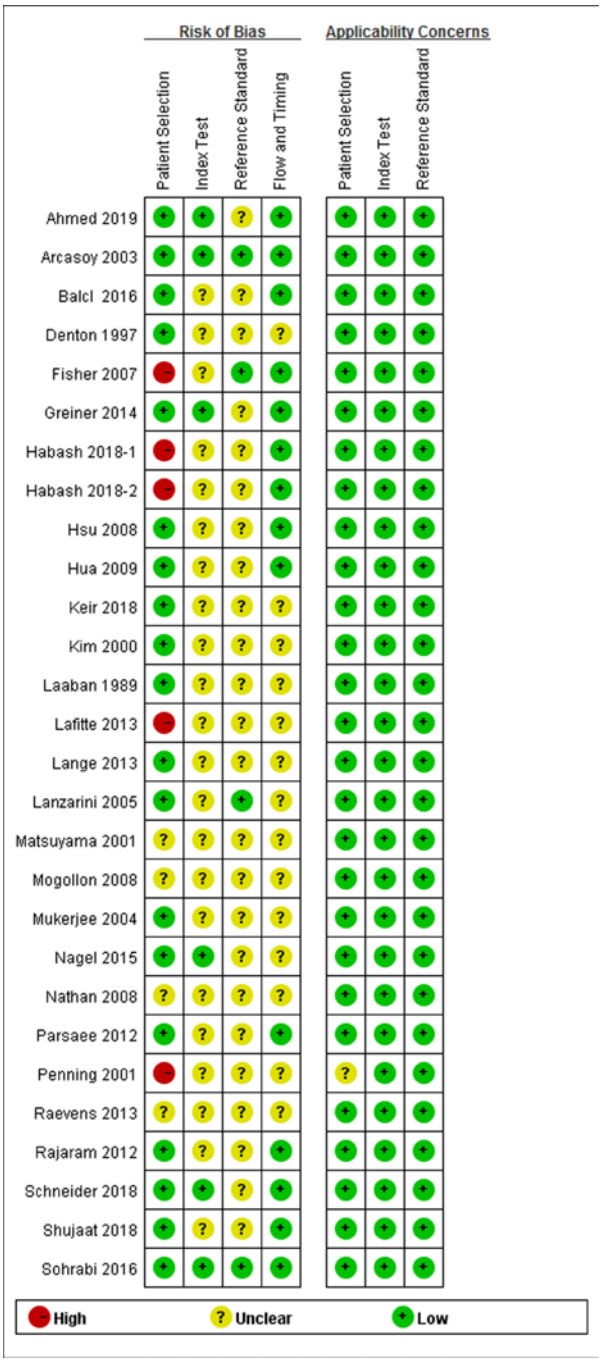

Figure 3 Summary receiver operating characteristic graph with 95% confidence region and 95% prediction region for transthoracicechocardiography in the diagnosis of pulmonary hypertension (n=28).

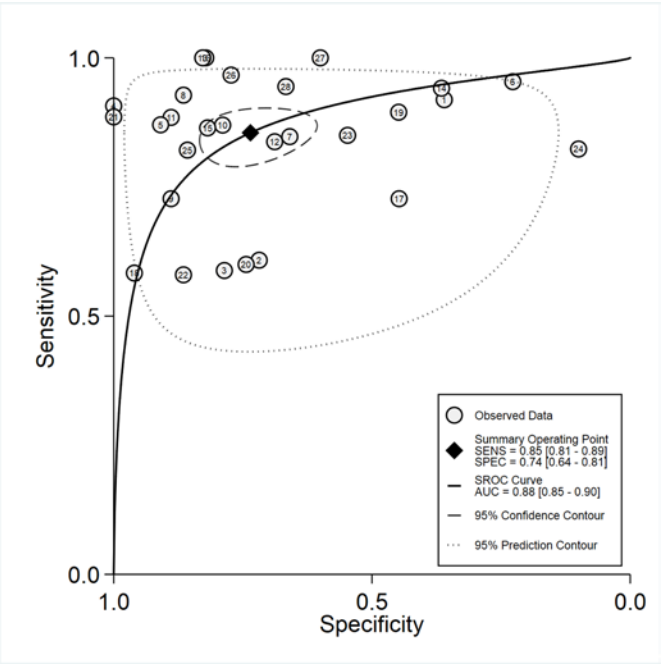

Figure 4 Forest plot of the sensitivity and specificity of each individual study, summary sensitivity and specificity and I^2 statistic for heterogeneity (n=28).

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% CI.[11 26 34 44] The areaunder the curve (AUC) was 0.88 (95% CI 0.85 to 0.90). The pooled sensitivity and specificity for TTE were 85% (95% CI 81% to 90%) and 74% (95% CI 64% to 81%), respectively (figure 4). The pooled PLR and NLR were 3.2 (95% CI 2.3 to 4.4) and 0.20 (95% CI 0.15 to 0.26), respectively. The pooled DOR for TTE was 16 (95% CI 10 to 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% to 91%), specificity (74%, 95% CI 62% to 83%) and AUC (0.89, 95% CI 0.86 to 0.91) of TTE to diagnose PH were higher for studies published in 2010 and later compared with 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% to 95%), 90% (95% CI 53% to 99%) and 0.94 (95% CI 0.92 to 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% to 88%), specificity (61%, 95% CI 53% to 69%) and AUC (0.73, 95% CI 0.69 to 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 mm Hg had a higher diagnostic accuracy than that at 40 mm Hg. The sensitivity, specificity and AUC of the former were respectively 92% (95% CI 88% to 94%), 65% (95% CI 43% to 83%) and 0.92 (95% CI 0.89 to 0.94), while the sensitivity, specificity and AUC at 40 mm Hg were 84% (95% CI 75% to 91%), 52% (95% CI 31% to 71%) and 0.80 (95% CI 76% to 83%), respectively.

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 to 99)	0.88 (0.85 to 0.90)	0.85 (0.81 to 0.90)	0.74 (0.64 to 0.81)	3.2 (2.3 to 4.4)	0.20 (0.15 to 0.26)	16 (10 to 27)
Time interval								
≤24 hour	4	95 (90 to 99)	0.94 (0.92 to 0.96)	0.88 (0.73 to 0.95)	0.90 (0.53 to 0.99)	8.9 (1.5 to 54.5)	0.13 (0.06 to 0.29)	68 (13 to 348)
≤48 hour*	7	95 (90 to 99)	0.94 (0.91 to 0.95)	0.88 (0.81 to 0.93)	0.89 (0.71 to 0.96)	7.8 (2.8 to 21.3)	0.13 (0.09 to 0.21)	59 (23 to 148)
≤72 hour†	9	94 (89 to 99)	0.91 (0.89 to 0.93)	0.87 (0.82 to 0.91)	0.83 (0.65 to 0.93)	5.2 (2.4 to 11.2)	0.15 (0.11 to 0.21)	34 (14 to 82)
≤1 week	13	93 (87 to 99)	0.91 (0.88 to 0.93)	0.87 (0.84 to 0.90)	0.80 (0.68 to 0.88)	4.3 (2.7 to 6.9)	0.16 (0.12 to 0.21)	27 (15 to 48)
>1 week	10	97 (95 to 99)	0.82 (0.78 to 0.85)	0.85 (0.73 to 0.92)	0.60 (0.40 to 0.77)	2.1 (1.3 to 3.4)	0.25 (0.14 to 0.45)	9 (4 to 21)
Unclear	5	82 (63 to 100)	0.85 (0.810.88)	0.79 (0.63 to 0.99)	0.76 (0.61 to 0.87)	3.4 (1.9 to 5.9)	0.27 (0.15 to 0.51)	12 (5 to 33)
Population disease								
Cardiac diseases	6	94 (89 to 99)	0.90 (0.87 to 0.92)	0.90 (0.86 to 0.93)	0.67 (0.29 to 0.91)	2.7 (0.9 to 8.1)	0.15 (0.08 to 0.30)	18 (3 to 95)
Lung diseases	8	90 (81 to 100)	0.73 (0.69 to 9.77)	0.81 (0.70 to 0.88)	0.61 (0.53 to 0.69)	2.1 (1.8 to 2.4)	0.32 (0.21 to 0.48)	7 (4 to 10)
Multiple diseases‡	6	93 (87 to 99)	0.90 (0.87 to 0.92)	0.89 (0.84 to 0.92)	0.70 (0.40 to 0.89)	3.0 (1.3 to 7.1)	0.16 (0.11 to 0.23)	19 (6 to 60)
Unclear§	8	88 (77 to 100)	0.88 (0/85 to 0.90)	0.80 (0.64 to 0.90)	0.85 (0.80 to 0.89)	5.3 (4.0 to 7.0)	0.23 (0.12 to 0.45)	23 (10 to 51)
Published year								
≥2010	15	97 (95 to 99)	0.89 (0.86 to 0.91)	0.87 (0.81 to 0.91)	0.74 (0.62 to 0.83)	3.3 (2.3 to 4.9)	0.18 (0.13 to 0.25)	19 (11 to 13)
<2010	13	96 (93 to 99)	0.86 (0.83 to 0.89)	0.84 (0.74 to 0.90)	0.73 (0.56 to 0.85)	3.1 (1.8 to 5.3)	0.22 (0.14 to 0.37)	14 (6 to 33)
Study design								
Prospective	15	97 (95 to 99)	0.90 (0.87 to 0.92)	0.86 (0.77 to 0.91)	0.79 (0.69 to 0.87)	4.2 (2.7 to 6.4)	0.18 (0.11 to 0.28)	23 (12 to 45)
Retrospective	13	96 (92 to 99)	0.86 (0.83 to 0.89)	0.86 (0.80 to 0.90)	0.65 (0.49 to 0.78)	2.5 (1.6 to 3.7)	0.22 (0.15 to 0.32)	11 (6 to 22)
TTE threshold								
SPAP ≥40 mm Hg	8	96 (93 to 99)	0.80 (0.76 to 0.83)	0.84 (0.75 to 0.91)	0.52 (0.31 to 0.71)	1.7 (1.2 to 2.5)	0.30 (0.21 to 0.44)	6 (3 to 11)
SPAP ≥35 mm Hg	4	76 (47 to 100)	0.92 (0.890.94)	0.92 (0.88 to 0.94)	0.65 (0.43 to 0.83)	2.6 (1.4 to 4.9)	0.13 (0.08 to 0.22)	16 (9 to 28)
TRPG	4	0 (0 to 100)	0.85 (0.82 to 0.88)	0.75 (0.58 to 0.86)	0.81 (0.70 to 0.89)	4.0 (2.2 to 7.3)	0.31 (0.17 to 0.57)	13 (4 to 40)

*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).
AUC, area under the curve; DOR, diagnostic OR; NLR, negative likelihood ratio; PLR, positive likelihood ratio; SPAP, systolic pulmonary artery pressure; TRPG, tricuspid regurgitation pressure gradient; TTE, transthoracic echocardiography.

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 to 99)	0.88 (0.85 to 0.90)	0.85 (0.81 to 0.90)	0.74 (0.64 to 0.81)	3.2 (2.3 to 4.4)	0.20 (0.15 to 0.26)	16 (10 to 27)
Excluding study of Penning*	27	98 (97 to 99)	0.88 (0.85 to 0.91)	0,86 (0.81 to 0.89)	0.75 (0.66 to 0.82)	3.4 (2.5 to 4.6)	0.19 (0.14 to 0.26)	18 (11 to 28)
RHC threshold MPAP ≥25 mm Hg	21	98 (97 to 99)	0.87 (0.84 to 0.90)	0.83 (0.77 to 0.88)	0.76 (0.67 to 0.83)	3.5 (2.5 to 4.8)	0.22 (0.16 to 0.30)	16 (10 to 26)
RAP method (IVC)†	17	96 (93 to 99)	0.89 (0.86 to 0.91)	0.86 (0.82 to 0.90)	0.73 (0.59 to 0.84)	3.2 (2.0 to 5.1)	0.19 (0.13 to 0.27)	17 (8 to 35)
Excluding high TTE threshold*	21	97 (95 to 99)	0.90 (0.87 to 0.92)	0.88 (0.85 to 0.91)	0.72 (0.59 to 0.82)	3.2 (2.1 to 4.8)	0.16 (0.12 to 0.22)	20 (11 to 36)

*High TTE threshold was defined as SPAP greater than 45 mm Hg or tricuspid regurgitation pressure gradient (TRPG) greater than 40 mm Hg.
†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 analysed.
‡The study of Penning was excluded because only pregnant women with cardiac disease were included as subjects.
AUC, area under the curve; DOR, diagnostic OR; IVC, inferior vena cava; MPAP, mean pulmonary artery pressure; NLR, negative likelihood ratio; PLR, positive likelihood ratio;RAP, right atrial pressure; RHC, right heart catheterisation; TTE, transthoracic echocardiography.

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 favourable. 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 tricuspid regurgitation pressure gradient (TRPG) (4TRVmax^2) 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 mm Hg 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 mm Hg is arbitrarily chosen and lowering it to 20 mm Hg (two SDs 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 mm Hg 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 >20 mm Hg as the diagnostic threshold for RHC, subgroup analysis on the two thresholds of 20 mm Hg and 25 mm Hg 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 mm Hg. Subgroup analysis showed that the diagnostic accuracy of the group of 35 mm Hg 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 multicentre 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 10 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. First, 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. Second, 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. Third, 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 multicentre, large-sample studies to further improve the accuracy of TTE in the diagnosis of PH in future research.

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

Contributors The joint corresponding authors (J-QL and BS) are responsible for the design and implementation of the study. S-DL is responsible for the quality control of study selection. YH performed the quality control on the links of data extraction. K-HY provided guidance in literature retrieval and data processing methodology and was responsible for the quality evaluation part. J-RN and P-JY performed the systematic review of the literature and extracted the data. J-RN conducted the meta-analyses, and two authors (J-RN, P-JY) substantially contributed to the interpretation of the data and wrote the article. All authors repeatedly revised the article. The corresponding authors (J-QL and BS) and J-RN take responsibility for the integrity of the analyses.

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

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD
Jun-Qiang Lei http://orcid.org/0000-0002-2636-9389

REFERENCES

1 Hoeper MM, Humbert M, Souza R, *et al*. A global view of pulmonary hypertension. *Lancet Respir Med* 2016;4:306–22.
2 Galiè 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 Galiè N, Humbert M, Vachiery J-L, *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;85:50–8.
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;23:1261–4.
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;68:203–8.
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;349:g7647.
16 Ge L, Tian J-H, Li Y-N, *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 A, 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 AWS, 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 AWS, *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 Y-W, Wu Z-Y, *et al*. Role of 2-dimensional Doppler echo-cardiography in screening portopulmonary hypertension in portal hypertension patients. *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;6:453–8.
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 Suppl* 2013:5–12.
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 Mogollón Jiménez MV, Escoresca Ortega AM, Cabeza Letrán ML, *et al*. Correlation of echocardiographic and hemodynamic parameters in pulmonary hypertension assessment prior to heart transplantation. *Transplant Proc* 2008;40:3023–4.
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.
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.