Pulmonary hypertension (PH), a syndrome characterized by increased pulmonary vascular resistance and remodeling, is associated with significant morbidity and mortality, which are directly related to cardiac function. Many patients are diagnosed at a late stage of the disease because of its nonspecific symptoms and signs; earlier diagnosis and more effective treatment of PH would be of great benefit. At present, right-heart catheterization is the gold standard for the definitive diagnosis of PH. Although right-heart catheterization could assess pulmonary arterial pressure and cardiac output accurately, it is not routinely used because of its invasiveness and complications. Thus, accurate noninvasive assessment of pulmonary arterial pressure is desirable both for diagnostic purposes and to assess response to therapy.

Among the current noninvasive methods, transthoracic Doppler echocardiography (TDE) is recommended as the initial noninvasive modality in the screening and evaluation of PH. Echocardiography can be used to evaluate right-sided chamber size and function, and the presence of pericardial effusion, which are known to influence survival. Frequently, TDE is used to estimate right ventricular systolic pressure by estimating the pressure gradient between the right ventricle and the right atrium using a modified Bernoulli equation. An estimated right atrial pressure is added to this number to approximate the right ventricular systolic pressure, which equals the pulmonary artery systolic pressure in the absence of pulmonic stenosis.

Using the above method, several studies have demonstrated an adequate correlation between the Doppler estimates and direct measurements with right-heart catheterization. As a noninvasive and convenient method, TDE has been used widely in the diagnosis of PH; however, many published studies on the potential usefulness of TDE as a diagnostic tool for PH have reported varying results. The aim of this meta-analysis was to systematically and quantitatively evaluate all published studies that assessed the accuracy of TDE for the diagnosis of PH.
to identify suitable studies prior to 24 March 2010; no limit was applied to the start date. The search terms were “PH,” “TDE,” “transthoracic Doppler echocardiogram,” “sensitivity and specificity,” and “right-heart catheterization.” Articles were also identified by use of the related-articles function in PubMed, and the references of identified articles were searched manually. We contacted the authors for further study details if needed. Although no language restrictions were imposed initially, for the full-text review and final analysis, only English-language articles were included. Conference abstracts and letters to journal editors were excluded because of the limited data they contained.

Studies were included in the meta-analysis if they provided both the sensitivity and specificity of levels of TDE for the diagnosis of PH. Two reviewers (R.F.Z. and F.C.S.) independently determined study eligibility, and differing decisions were resolved by consensus. Publications that may have been based on the same study (e.g., same authors, institutions, period of study) were discussed by R.F.Z., F.C.S., and K.J.Y., and only the best-quality study was used.

Data extraction and quality assessment. The final set of English-language articles was assessed independently by two reviewers (R.F.Z. and F.C.S.). The reviewers were blinded to publication details, and disagreements between them were resolved by consensus. Data retrieved from the reports included author, publication year, participant characteristics, test methods, sensitivity and specificity data, cut-off value, and methodological quality.

The methodological quality of the studies was assessed using guidelines published by the standards for reporting diagnostic accuracy (STARD) initiative (maximum score 25), and the quality assessment for studies of diagnostic accuracy (QUADAS) tool (maximum score 14).

Statistical analysis. We used standard methods for meta-analysis of diagnostic test evaluations. We computed the following measures of test accuracy for each study: sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and diagnostic odds ratio (DOR). The analysis was based on a summary receiver operator characteristic (SROC) curve. A random-effects model was used to calculate the average sensitivity, specificity, and other measures across studies.

The term “heterogeneity” when used in relation to meta-analysis refers to the degree of variability in results across studies. We used the $\chi^2$ and Fisher exact tests to detect statistically significant heterogeneity. To assess the effects of STARD and QUADAS scores on the diagnostic ability of TDE, we included them as covariates in univariate meta-regression analysis. The relative DOR (RDOR) was calculated according to standard methods to analyze the change in diagnostic precision in the study per unit in the covariate. Because publication bias is of concern for meta-analysis of diagnostic studies, we tested for the potential presence of this bias using Egger test.

RESULTS

After independent review, 15 publications concerned with TDE for the diagnosis of PH were considered to be eligible for inclusion in the analysis. After full-text review, nine studies were excluded: five articles were excluded because they did not have data available to generate a 2 × 2 table, and four studies were excluded because TDE was not used. In total, six studies including 188 patients with PH and 548 controls were available for analysis (Table 1). PH was confirmed by right-heart catheterization. The clinical characteristics, together with STARD and QUADAS scores of these studies, are outlined in Table 1.

Diagnostic accuracy

The sensitivity and specificity for six TDE assays in the diagnosis of PH were calculated. The sensitivity ranged from 0.58 to 1.00 (mean, 0.82; 95% confidence interval (CI), 0.76–0.88), whereas specificity ranged from 0.29 to 0.96 (mean, 0.68; 95% CI, 0.64–0.72). We also noted that PLR was 2.88 (95% CI, 0.64–0.72), FNR, false negative; FPR, false positive; PASP, pulmonary artery systolic pressure; PH, pulmonary hypertension; STARD, standards for reporting diagnostic accuracy; TDE, transthoracic Doppler echocardiography; RVSP, right ventricular systolic pressure; TDE, transesophageal Doppler echocardiography; TN, true negative; TP, true positive.

Table 1 Summary of included studies

<table>
<thead>
<tr>
<th>Study/year</th>
<th>Country</th>
<th>Patients no./ controls no.</th>
<th>Standard of PH by TDE</th>
<th>Reference standard</th>
<th>TP no.</th>
<th>FP no.</th>
<th>FN no.</th>
<th>TN no.</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Quality</th>
<th>STARD Score</th>
<th>QUADAS Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al.11/2000</td>
<td>United States</td>
<td>30/44</td>
<td>RVSP ≥50 mm Hg</td>
<td>RHC</td>
<td>29</td>
<td>10</td>
<td>1</td>
<td>34</td>
<td>96.7</td>
<td>77.3</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Arscosy et al.12/2003</td>
<td>United States</td>
<td>94/280</td>
<td>PASP ≥45 mm Hg or RV findings</td>
<td>RHC</td>
<td>80</td>
<td>107</td>
<td>14</td>
<td>173</td>
<td>85.1</td>
<td>61.8</td>
<td>16</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Saner et al.13/2006</td>
<td>Germany</td>
<td>14/60</td>
<td>PASP ≥40 mm Hg or RV findings</td>
<td>RHC</td>
<td>9</td>
<td>14</td>
<td>5</td>
<td>46</td>
<td>64.3</td>
<td>76.7</td>
<td>14</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Nathan et al.14/2008</td>
<td>United States</td>
<td>22/38</td>
<td>RVSP ≥35 mm Hg</td>
<td>RHC</td>
<td>19</td>
<td>27</td>
<td>3</td>
<td>11</td>
<td>86.4</td>
<td>28.9</td>
<td>12</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Hsu et al.14/2008</td>
<td>United States</td>
<td>24/25</td>
<td>RVSP ≥47 mm Hg</td>
<td>RHC</td>
<td>14</td>
<td>1</td>
<td>10</td>
<td>24</td>
<td>58.3</td>
<td>96.0</td>
<td>13</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Hua et al.10/2009</td>
<td>China</td>
<td>4/101</td>
<td>PASP ≥30 mm Hg</td>
<td>RHC</td>
<td>4</td>
<td>18</td>
<td>0</td>
<td>83</td>
<td>100.0</td>
<td>82.2</td>
<td>17</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

FN, false negative; FP, false positive; PASP, pulmonary artery systolic pressure; PH, pulmonary hypertension; QUADAS, quality assessment for studies of diagnostic accuracy; RHC, right-heart catheterization; RV, right ventricular; RVSP, right ventricular systolic pressure; STARD, standards for reporting diagnostic accuracy; TDE, transthoracic Doppler echocardiography; TN, true negative; TP, true positive.
1.77–4.70), NLR was 0.31 (95% CI, 0.18–0.53), and DOR was 11.36 (95% CI, 4.62–27.94). χ² values of sensitivity, specificity, PLR, NLR, and DOR were 18.40 (P < 0.01), 55.89 (P < 0.01), 42.32 (P < 0.01), 11.32 (P < 0.05), and 10.99 (P = 0.0515), respectively, indicating significant heterogeneity for sensitivity, specificity, PLR, and NLR among studies.

The SROC curve and its area under the curve (AUC) present an overall summary of best performance, and demonstrate the tradeoff between sensitivity and specificity. Unlike the traditional receiver operator characteristic plot that explores the effect of varying threshold on sensitivity and specificity in a single study, each data point in the SROC plot represents a separate study. In our study, AUC was 0.86 (weighted AUC, 0.86), indicating a high level of overall accuracy.

Multiple regression analysis and publication bias
By use of the STARD guidelines,6 a quality score for every study was compiled on the basis of title, introduction, methods, results, and discussion (Table 1). Quality scoring was also done by use of QUADAS5 (Table 1). These scores were used in the meta-regression analysis to assess the effect of study quality on the RDOR of TDE in the diagnosis of PH. In the six studies, the higher quality studies (QUADAS score ≥10) produced RDOR values that were not significantly higher than those produced by studies of lower quality (QUADAS score <10). The coefficient and RDOR were 0.019 and 1.02 (95% CI 0.00–297908.97) respectively, with P = 0.9953. Similarly, the studies with STARD score ≥13 did not have a better performance than did those studies with STARD score <13. Also, the coefficient and RDOR were −0.749 and 0.47 (95% CI 0.00–251.14), respectively, with P = 0.7292. These results indicated that the study quality did not substantially affect the estimate of the accuracy of TDE in the diagnosis of PH.

Publication bias was detected by using the Egger test. The result of the Egger test was not significant (P = 0.382). This showed no publication bias.

DISCUSSION
Distinguishing PH from non-PH is a critical clinical problem, for which conventional methods, such as chest X-ray, electrocardiogram, and right-heart catheterization, because of their limitations, are not always helpful in making the diagnosis. As a noninvasive method, TDE has been widely proposed to assist in the diagnosis of PH. Hsu and his colleagues found that TDE is a more sensitive and specific indicator when compared with magnetic resonance imaging and pulmonary function tests.14

The SROC curve and its AUC present an overall summary of test performance, and demonstrate the tradeoff between sensitivity and specificity. The present meta-analysis has shown that the mean sensitivity of the TDE assay is 0.82, whereas the specificity is 0.68, and that the maximum joint sensitivity and specificity (Q value) is 0.79, whereas the AUC is 0.86, indicating a high level of accuracy.

The DOR is a single indicator of test accuracy16 that combines the data from sensitivity and specificity into a single number. The DOR of a test is the ratio of the odds of positive test results in the patient with disease relative to the odds of positive test results in the patient without disease. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance. A DOR of 1.0 indicates that a test does not discriminate between patients with the disorder and those without it. In the present meta-analysis, we have found that the mean DOR was 11.36, also indicating a high level of overall accuracy.

Because the SROC curve and the DOR are not easy to interpret and use in clinical practice, and because likelihood ratios are considered to be more clinically meaningful,17 we also presented both PLR and NLR as our measures of diagnostic accuracy. In the present study, a PLR value of 2.88 suggests that patients with PH have an approximately threefold greater chance of having a positive TDE assay, compared with controls. This probability is not high enough to confirm the diagnosis of PH. On the other hand, NLR was found to be 0.31 in the present meta-analysis. Thus, if the TDE assay result is negative for any individual, the probability that this individual has PH is 31%. These data suggest that a negative TDE assay result should not be used alone as a justification to deny PH. The confirmation of PH should be based on the results of right-heart catheterization.

An exploration of the reasons for heterogeneity rather than the computation of a single summary measure is an important goal of meta-analysis.18 In our meta-analysis, both STARD and QUADAS scores were used in the meta-regression analysis to assess the effect of study quality on RDOR. We did not observe that the studies with higher quality (i.e., STARD score ≥13 or QUADAS score of ≥10) had better test performance than those with lower quality, although significant heterogeneity for sensitivity, specificity, PLR, and NLR was found among these studies.

Because PH has high morbidity and mortality, the early detection of PH is an important goal. Although the previous study showed that TDE was a more sensitive and specific indicator compared with other noninvasive methods,14,19 and it showed good diagnostic performance in the present meta-analysis, its power of discrimination is not great enough for the definitive diagnosis of PH. In spite of this, we still believe that TDE is a useful tool. Unlike other noninvasive methods, TDE can estimate the pulmonary artery systolic pressure indirectly. These characteristics make it useful to assess response to therapy, and to determine the next treatment plan, especially in those patients who could not tolerate right-heart catheterization. The combination of TDE with other noninvasive methods may increase the diagnostic accuracy, but evidence-based medical studies of such a combination are rare, and should be the next direction for future study of diagnostic modalities for TDE.

Taken together, to the best of our knowledge, this is the first meta-analysis addressing TDE for the diagnosis of PH. Current evidence suggests that as a noninvasive method, TDE is a test with acceptable mixed sensitivity, but in isolation, it has insufficient specificity for detecting PH. It may be useful for first-line surveillance in patients in whom PH is suspected, although the confirmation of PH should be based
Echocardiography for Pulmonary Hypertension

National Natural Science Foundation of China

Acknowledgment: This study was supported by research grant WKJ 2006-2-018 from the Ministry of Health in China and research grant 81000019 from National Natural Science Foundation of China.

Disclosure: The authors declared no conflict of interest.


