

Contrast-enhanced endoscopic ultrasound for the differential diagnosis between benign and malignant lymph nodes: a meta-analysis



Authors

Andrea Lisotti¹, Claudio Ricci², Marta Serrani¹, Claudio Calvanese¹, Sandro Sferrazza¹, Nicole Brighi³, Riccardo Casadei², Pietro Fusaroli^{1,2}

Institutions

- 1 Gastroenterology Unit, Department of Medical and Surgical Sciences, University of Bologna, Hospital of Imola, Imola, Bologna, Italy
- 2 Department Medical and Surgical Sciences (DIMEC), University of Bologna, S.Orsola-Malpighi Hospital, Bologna, Italy
- 3 Department of Experimental, Diagnostic and Specialty Medicine (DIMES), University of Bologna, Oncology Unit, S. Orsola-Malpighi Hospital, Bologna, Italy

submitted 27.7.2018

accepted after revision 7.1.2019

Bibliography

DOI <https://doi.org/10.1055/a-0854-3785> |
Endoscopy International Open 2019; 07: E504–E513
© Georg Thieme Verlag KG Stuttgart · New York
ISSN 2364-3722

Corresponding author

Dr Andrea Lisotti, MD, Gastroenterology Unit, Hospital of Imola, Department of Medical and Surgical Science, University of Bologna, Via Montericco 4, 40026 Imola (BO), Italy
Fax: +39-51-6364120
lisotti.andrea@gmail.com

ABSTRACT

Background The differential diagnosis between benign and malignant lymph nodes (LNs) is crucial for patient management and clinical outcome. The use of contrast-enhanced endoscopic ultrasound (EUS) has been evaluat-

ed in several studies with diverse results. The aim of this meta-analysis was to evaluate the pooled diagnostic accuracy of contrast-enhanced EUS (CE-EUS) and contrast-enhanced harmonic EUS (CH-EUS) in this setting.

Methods A systematic electronic search was performed, including all original papers dealing with assessment of the nature of the LNs using CE-EUS or CH-EUS. A meta-analysis was performed to obtain pooled sensitivity, specificity, positive and negative likelihood ratio, and diagnostic odds ratio. The Summary Receiver Operating Characteristic (ROC) Curve method was used to calculate the area under the curve. Statistical analysis was carried out using Meta-Disc V.1.4, Stata V.12.0 and Review Manager V.5.2.

Results Among 210 pertinent studies, four (336 patients) were included in the analysis. The pooled sensitivity was 82.1% (75.1–87.7%) and pooled specificity was 90.7% (85.9–94.3%) with significant heterogeneity found in sensitivity; the positive-likelihood ratio (LR) was 7.77 (5.09–11.85) and the negative-LR was 0.15 (0.05–0.46); the pooled diagnostic odds ratio (DOR) was 54 (15–190). Subgroup analysis including studies performed using CH-EUS (two studies, 177 LNs) showed a pooled sensitivity of 87.7% (77.0–93.9%) and a pooled specificity of 91.8% (84.5–96.4%) with no significant heterogeneity; the pooled positive-LR was 9.51 (4.95–18.28) and the pooled negative-LR was 0.14 (0.06–0.35); pooled DOR was 68.42 (15.5–301.4).

Conclusions From these data, CE-EUS is not recommended due to inadequate sensitivity. On the other hand, CH-EUS studies showed optimal accuracy (pooled sensitivity 87.7% and specificity 91.8%), comparable to elastography and even EUS-guided fine needle aspiration (EUS-FNA), suggesting a role in the diagnostic algorithm.

Introduction

Mediastinal and abdominal lymph nodes (LNs) are commonly assessed for benign or malignant indications using endoscopic ultrasound (EUS). In some cases, the presence of malignant LNs

modifies subsequent clinical management; for this reason, an incorrect diagnosis may significantly affect patient outcome.

The differential diagnosis of LNs based on morphological characteristics is somewhat inaccurate. Different size cutoffs

have been proposed for each anatomical district [1–4]; however, it has been demonstrated that size alone has low sensitivity and specificity. In fact, while up to 30% of LNs smaller than 5mm could be malignant, benign LNs could be larger than 20mm in the case of acute or chronic inflammation [5]. In addition, shape, borders, architecture, echogenicity and echotexture, vascular pattern, and distance of LNs from the neoplasia have also been proposed to increase accuracy [6]. However, EUS evaluation of LN morphology in this respect allows a confident diagnosis in just 25% of cases [7].

To increase the accuracy of differential diagnosis, EUS-guided LN sampling has been advocated. A large meta-analysis reported good sensitivity (88.0%) but suboptimal specificity (96.4%) for the diagnosis of malignant LN with EUS-guided fine needle aspiration (EUS-FNA) [8].

EUS image enhancement techniques were developed to increase the negative predictive value (NPV) of EUS and EUS-FNA [5]. In particular, EUS elastography (EUS-E) is capable of displaying tissue stiffness with a color scale, blue LNs being stiffer and more likely malignant than green ones. In theory, EUS-E could be used as a targeting method for EUS-FNA to increase the accuracy and reduce the number of needle passes. A meta-analysis demonstrated that EUS-E shows the same sensitivity (88%) as EUS-FNA with a specificity of 85% [9].

Different imaging techniques have been used in an attempt to characterize LNs. In 2003, Kojima et al. described contrast-enhanced echolymphography, where they injected carbon-dioxide microbubbles with an FNA needle into an enlarged LN to assess its vascular architecture. Observed sensitivity and specificity were 95.8% and 90.3%, respectively [10]. Subsequently, different studies reported both Doppler imaging (CE-EUS) and contrast-harmonic EUS (CH-EUS), with mixed results (sensitivity in the range 60–100%, and specificity in the range 85–100%) [11].

The aim of this study was to evaluate the pooled sensitivity and specificity of CE-EUS and CH-EUS for the differential diagnosis between benign and malignant LNs.

Material and methods

Literature search

At the end of December 2017, we conducted a systematic review of English-language articles through MEDLINE using PubMed and Google Scholar interfaces. The following search terms were used: contrast, contrast-enhanced, contrast-enhanced harmonic, CE-EUS, CH-EUS, CEH-EUS, endoscopic ultrasound, EUS, endosonography, endoscopic ultrasonography. The references from the selected articles were then analyzed to retrieve any other study that eluded the primary search.

Study selection

Inclusion criteria were: (1) original studies [randomized controlled trials (RCT), prospective studies (PS) and retrospective studies (RS)] designed to evaluate the diagnostic accuracy of contrast-enhanced EUS for the differential diagnosis of benign and malignant LNs; (2) use of pathology (EUS-FNA samples or

surgical specimens) or at least 3 months' follow-up as reference standards; (3) sufficient data for identification of true positive (TP), false positive (FP), false negative (FN), and true negative (TN) cases.

Studies were excluded in the case of unavailable, incomplete, duplicated or updated data, or in the case of case reports or case series enrolling <10 patients.

Data extraction and assessment

Two different physicians independently recorded the following data: authors, affiliation, country of origin (east vs. west), year of publication, study design, ultrasound platform and echodoposcope equipment, contrast modality used, ultrasound contrast agent, diagnostic criteria used to identify malignant LNs, sample size of the study, mean patient age, patient gender, mean LN size. No disagreement emerged during data collection. Qualitative assessment and evaluation of the potential bias of each study were carried out according to Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) system, based on the following four domains: patient selection, index tests, reference standard, and flow-and-timing [12].

Statistical methods

All categorical variables were reported as frequencies and percentages while the continuous variables were described as means and, when possible, standard deviations, or medians and ranges. To calculate the mean and standard deviation in studies which presented medians and ranges, we used a dedicated statistical algorithm [13]. The diagnostic performance was evaluated using classic biostatistical measurements of accuracy: (1) discriminative measurements such as diagnostic odds ratio (DOR), sensitivity (SE), specificity (SP), area under the curve (AUC) and summary receiver operating characteristics (SROC) curve; (2) predictive measurements such as positive (PPV) and negative (NPV) predictive values or positive (LR+) and negative (LR-) likelihood ratios [14, 15]. Briefly, the diagnostic odds ratio represents the ratio of the odds of positivity in diseased patients relative to the odds of positivity in non-diseased patients [16]. The DOR was used as the main indicator of diagnostic performance. The values of each DOR were obtained starting from a 2 by 2 table for each parameter in each study included. The DOR was calculated starting from true positive (TP), false positive (FP), false negative (FN), and true negative (TN). When 0 counts occurred in any of the cells in the 2 by 2 table, 0.5 was added to all cell values as a correction. The value of a DOR ranges from 0 to infinity, with higher values indicating better discriminatory performance of the parameters (more positive in patients with disease), while a value lower than 1 was related to improper parametric interpretation (more negative in patients with disease). A value of 1 means that the parameter does not differentiate between patients with the disorder and those without it. The DORs were calculated as meta-analytic pooled data as points estimated with a 95% confidence interval (95%CI) using the DerSimonian-Liard random effect model [17]. The area under the curve (AUC) represents the accuracy, ranging from 0 to 1 and is classified as poor

($AUC < 0.5$), low ($0.5 \geq AUC < 0.7$), moderate ($0.7 \geq AUC < 0.9$), or high ($0.9 \geq AUC = 1$) [18].

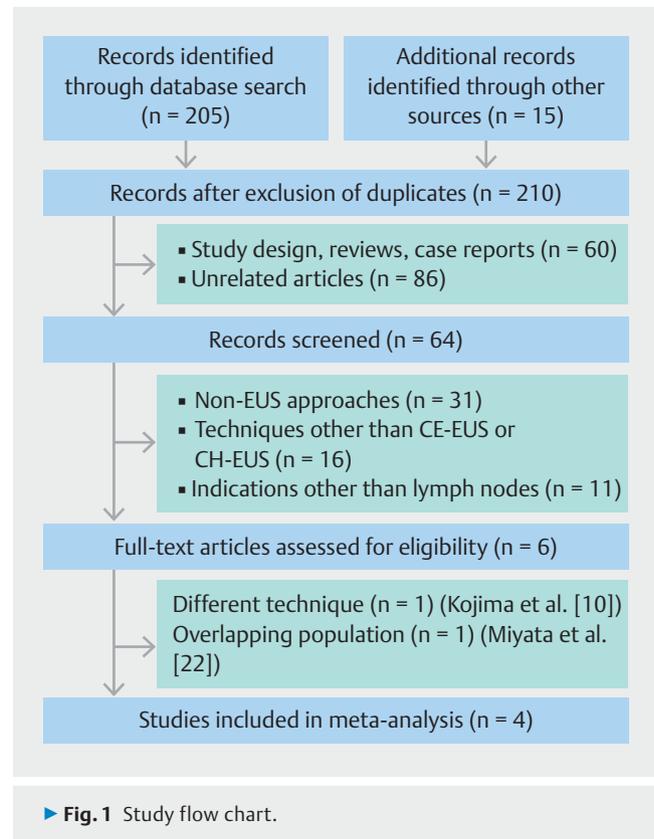
The positive likelihood ratio represents the ratio between patients with positive risk parameters and “diseased” as the same result in “healthy” patients; the negative likelihood ratio represents the ratio between patients with a negative risk parameter and “diseased” as the same result in “healthy” patients. The LR^+ and LR^- were calculated as meta-analytic pooled data as points estimated with a 95% confidence interval (95%CI) using the DerSimonian-Liard random effect model [17].

Finally, the presence of publication bias, diagnostic threshold variation, and heterogeneity “between-studies” was investigated to obtain robust conclusions. We explored the effect of the absence of small sample size “negative” studies and evaluated the asymmetry test described by Deeks et al. [19]. The presence of diagnostic threshold variation influences the values of the DOR and the symmetry of the SROC curve. When DOR variation is present, there is no linear relationship between the DOR and the AUC. Diagnostic threshold variation was studied using the Moses-Shapiro-Littenberg method [20]. A two-tailed P value < 0.10 indicated statistically significant asymmetry. A symmetrical SROC curve meant that the parameters had a constant DOR while an asymmetrical SROC curve resulted when the DOR changed among the various studies. Finally, the “between-study” statistical heterogeneity was assessed using both the Cochran Q statistic ($P < 0.10$) and I^2 statistics. In particular, the value of I^2 describes the percentage variability in point estimates which was due to heterogeneity rather than to sampling error. In fact, the I^2 statistic does not depend on the small sample size of the data, unlike the Cochran Q statistic. Heterogeneity was considered to be low when I^2 was $< 30\%$, moderate if between 30% and 50%, and high if $> 50\%$ [21]. Statistical analyses were carried out using Meta-Disc version 1.4 (Metadisc, Unit of Clinical Biostatistics of Ramon y Cajal Hospital, Madrid, Spain), Stata version 12.0 (StataCorp LP, College Station, TX, USA), and Review Manager Version 5.2 (The Cochrane Collaboration, Software Update, Oxford, UK).

Results

Included studies and quality assessment

In total, 205 studies were identified during the literature search and 15 additional manuscripts were found from evaluation of the references in those studies. After exclusion of duplicate publications, 210 articles constituted our study object. Of these, 146 manuscript were excluded after title and abstract evaluation. Sixty-four studies (56 full-text and 8 abstract) were retrieved; 58 out of 64 were excluded due to a non-EUS approach ($n = 31$), non-contrast-enhanced techniques ($n = 16$), or assessment not including LN ($n = 11$). Finally, the manuscript by Kojima et al. [10] was excluded because of investigation of a different technique (namely, “EUS echolymphography”); the abstract presented by Miyata et al. [22] was excluded because of data overlapping with a full-text study [23]. The study flow chart is shown in ► Fig. 1 and the main characteristics of the



► Fig. 1 Study flow chart.

four studies included (336 patients) are reported in ► Table 1 [23–26].

► Fig. 2 shows the quality assessment according to the QUADAS-2 tool. All studies showed high quality in terms of risk of bias and applicability; however, in the study by Xia et al., the risk of bias and applicability of patient selection and index test were unclear because data had been extrapolated from a larger study including intra-abdominal lesions of undetermined origin and seemed to be based only on FNA results [26].

Differential diagnosis between benign and malignant LNs

Meta-analysis results (random-effect model) are summarized in ► Table 2. Contrast-enhanced EUS had a pooled sensitivity of 82.1% (95%CI 75.1–87.7%) and a pooled specificity of 90.7% (95%CI 85.9–94.3%) (► Fig. 3a,b). Significant heterogeneity was found in sensitivity (Cochran Q test 31.28; d.f. 3; $P < 0.001$; $I^2 = 90.4\%$) but not in specificity (Cochran Q test 3.42; d.f. 3; $P = 0.331$; $I^2 = 12.4\%$).

Pooled positive likelihood ratio (LR^+) (random-effect model) was 7.77 (95%CI 5.09–11.85); pooled negative likelihood ratio (LR^-) was 0.15 (95%CI 0.05–0.46). No significant heterogeneity was found in LR^+ (Cochran Q test 1.21; d.f. 3; $P = 0.751$; $I^2 = 0.0\%$) while significant heterogeneity was found among LR^- (Cochran Q test 19.18; d.f. 3; $P < 0.001$; $I^2 = 84.4\%$) (► Fig. 3c, d). The pooled diagnostic odds ratio (DOR) was 54 (95%CI 15–190) with an estimated Prediction Interval of 0.47–6298 as shown in ► Supplementary Fig. 1. The symmetric SROC

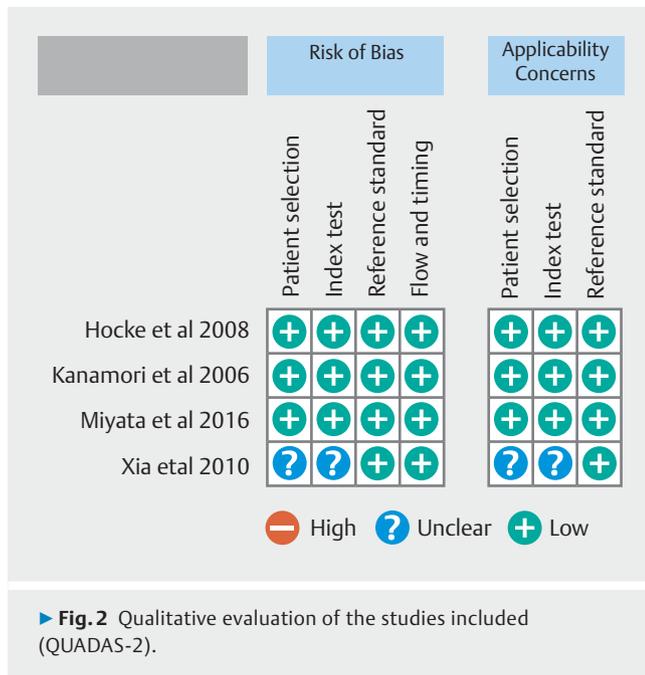
► **Table 1** Characteristics of the studies included.

Study	Affiliation	Country	Years	Study design	Equipment	Contrast mode	US contrast agent	Diagnostic criteria for malignant LN	Sample size	Mean age, years	M/F ratio	Benign LN size, mm	Malignant LN size, mm
Kanamori et al. [24]	Nagoya University School of Medicine, Nagoya, Japan	Japan (E)	2006	Retro-spective and validation set	Pentax Hitachi	Color Doppler (CE-EUS)	Levovist (Nihon Schering, Japan)	Presence of filling defects	71	63.7	0.9	17.2	26.1
Hocke et al. [25]	Friedrich-Schiller University Jena, Jena, Germany	Germany (W)	2008	Prospective, consecutive	Pentax Hitachi and Olympus Aloka	Power Doppler (CE-EUS)	SonoVue (Bracco, Italy)	Irregular vessel appearance; only artery visible	122	63	3.1	25.4	28.7
Xia et al. [26]	Kinki University School of Medicine, Osaka, Japan	Japan (E)	2010	Prospective, consecutive ¹	Olympus Aloka	Harmonic (CH-EUS)	Sonazoid (Daiichi-Sankyo, Japan)	Heterogeneous enhancement	34 ¹	NE ¹	NE ¹	NE ¹	NE ¹
Miyata et al. [23]	Kinki University School of Medicine, Osaka, Japan	Japan (E)	2016	Prospective, consecutive	Olympus Aloka	Harmonic (CH-EUS)	Sonazoid (Daiichi-Sankyo, Japan)	Heterogeneous enhancement	109 ²	61.7	1.7	NE	NE

Abbreviations: W, western; E, eastern; M, male; F, female; CE-EUS, contrast-enhanced EUS; CH-EUS, contrast-enhanced harmonic EUS; LN, lymph node; NE, data not extractable.

¹ Data have been extrapolated from a larger cohort of patients with intra-abdominal lesions of undetermined origin.

² Authors enrolled 109 patients and studied 143 lymph nodes with CH-EUS.



curve is shown in ► **Fig. 4**; the area under the curve was 0.958 (SE 0.02).

It was not possible to perform a meta-regression analysis due to the small number of studies included.

Pooled analysis of studies performed using CH-EUS

Two studies (143 patients with 177 LNs) were performed using dedicated contrast harmonic mode for evaluation of the nature of the LNs [23, 26]. Pooled sensitivity was 87.7% (77.0–93.9%); no significant heterogeneity was found (Cochran Q test 2.60; d. f. 1; $P=0.107$; $I^2=61.6\%$). Pooled specificity was 91.8% (84.5–96.4%); no significant heterogeneity was found (Cochran Q test 1.99; d. f. 1; $P=0.158$; $I^2=49.8\%$). Pooled LR+ was 9.51 (4.95–18.28); no significant heterogeneity was found (Cochran Q test 0.49; d. f. 1; $P=0.484$; $I^2=0.0\%$). Pooled LR–was 0.14 (0.06–0.35); no significant heterogeneity was found (Cochran

Q test 1.52; d. f. 1; $P=0.218$; $I^2=34.2\%$). Pooled DOR according to the use of Doppler imaging (CE-EUS) or dedicated contrast harmonic mode (CH-EUS) is shown in ► **Fig. 5**; no statistically significant difference was found between the two imaging techniques. In particular, DOR was 62.2 (2.7–1448) in CE-EUS studies and 68.4 (15.5–301.4) in CH-EUS studies (► **Fig. 5**).

Bias estimation

► **Fig. 6** shows the Deeks' funnel plot asymmetry test with $1/\text{root}$ (effective sample size) (ESS) plotted on the y axis and DOR on the x axis. The superimposed regression line weighted the effect of sample size on the DOR. The statistically non-significant P value (0.28) for the slope coefficient suggests symmetry in the data and a low likelihood of publication bias.

Discussion

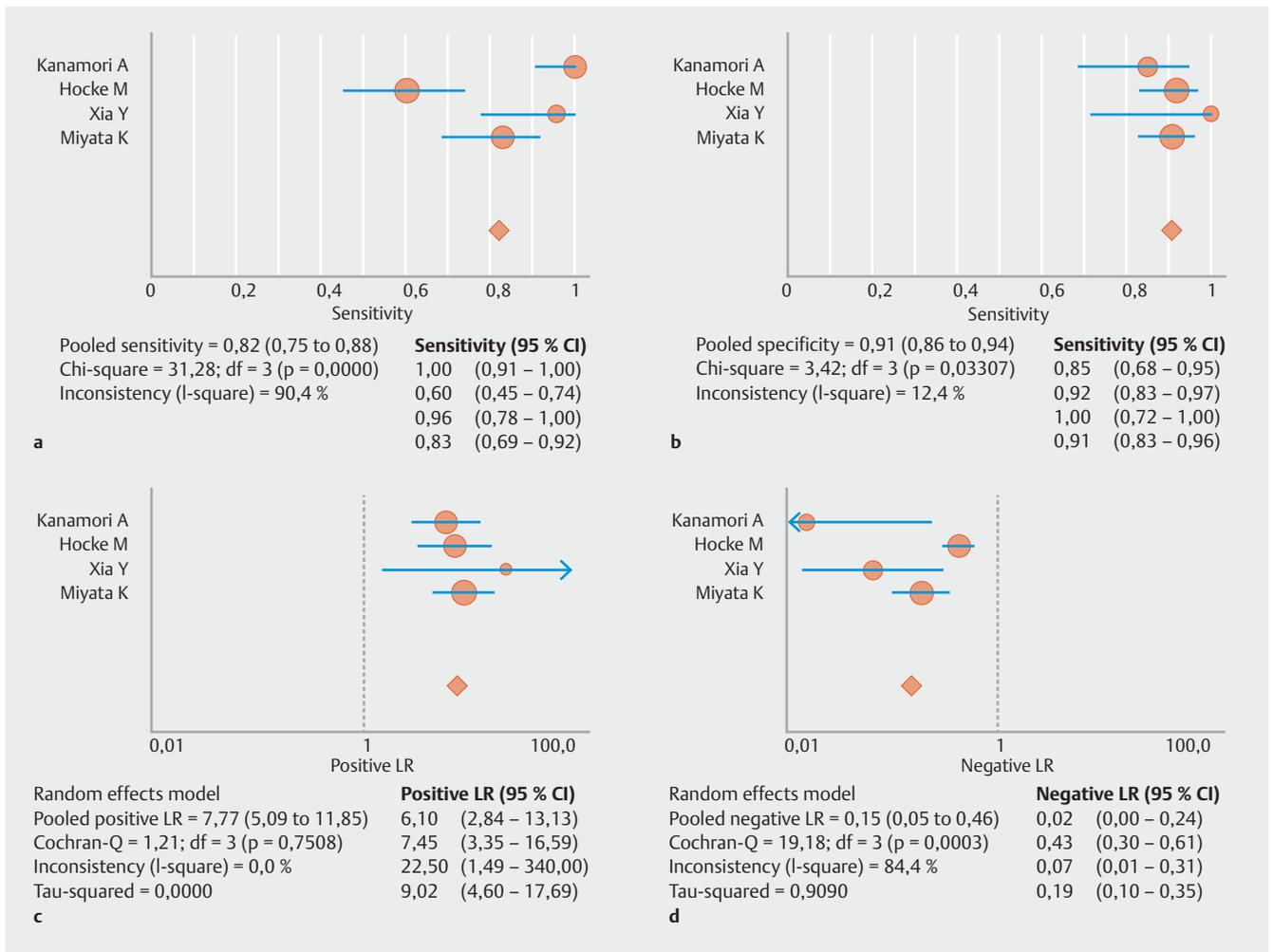
While the pooled diagnostic accuracy of B-mode EUS, EUS-E, and EUS-FNA has already been evaluated [8, 9], this is the first meta-analysis looking at the pooled sensitivity and specificity of contrast-enhanced EUS in the differential diagnosis of LNs. The correct differential diagnosis between benign and malignant LNs is crucial for patients undergoing EUS for either tumor staging or other indications. While B-mode EUS criteria (i. e. size, morphology, shape and echogenicity) are essentially inadequate to draw reliable conclusions about the nature of LNs, EUS tissue acquisition allows accurate pathological characterization. Since EUS-FNA presents 88% sensitivity in this setting, EUS image enhancement techniques (EUS-elastography and contrast-enhanced EUS) have been developed to increase the negative predictive value of EUS-FNA [5, 8].

Considering that diverse results were reported in the literature, we aimed at identifying the pooled sensitivity and specificity of contrast-enhanced EUS for the differential diagnosis of benign and malignant LNs [11]. The present meta-analysis showed a poor pooled sensitivity (82.1%) and an optimal pooled specificity (90.7%). We found significant heterogeneity among the four studies in terms of sensitivity but not specificity; this finding could be justified by the small number of studies

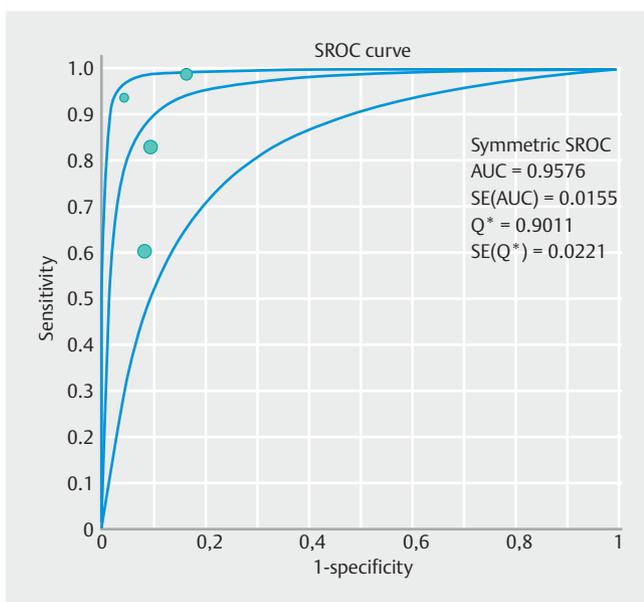
► **Table 2** Meta-analysis results.

Study	Malign/Total LNs ratio (%)	Sensitivity (95% CI)	Specificity (95% CI)	Pooled DOR (95% CI)	Pooled LR+ (95% CI)	Pooled LR– (95% CI)
Kanamori et al. 2006 [24]	38/71 (53.5)	1.000 (0.907–1.000)	0.848 (0.681–0.949)	399 (21–7512)	6.10 (2.84–13.13)	0.02 (0.01–0.24)
Hocke et al. 2008 [25]	48/122 (39.3)	0.604 (0.453–0.742)	0.919 (0.832–0.970)	17 (6–48)	7.45 (3.35–16.59)	0.43 (0.30–0.62)
Xia et al. 2010 [26]	23/34 (67.6)	0.957 (0.781–0.999)	1.000 (0.715–1.000)	345 (13–9155)	22.50 (1.49–340)	0.07 (0.01–0.31)
Miyata et al. 2016 [23]	47/143 (32.9)	0.830 (0.692–0.924)	0.908 (0.827–0.959)	48 (17–138)	9.02 (4.60–17.69)	0.19 (0.10–0.35)
Pooled	156/370 (42.2)	0.821 (0.751–0.877)	0.907 (0.859–0.943)	54 (15–190)	7.77 (5.09–11.85)	0.15 (0.05–0.46)

Abbreviations: CI, Confidence Interval; DOR, Diagnostic Odds Ratio; LR+, positive likelihood ratio; LR–, negative likelihood ratio.



► **Fig. 3** Forest plots: **a** pooled sensitivity; **b** pooled specificity; **c** pooled positive likelihood ratio; **d** pooled negative likelihood ratio.



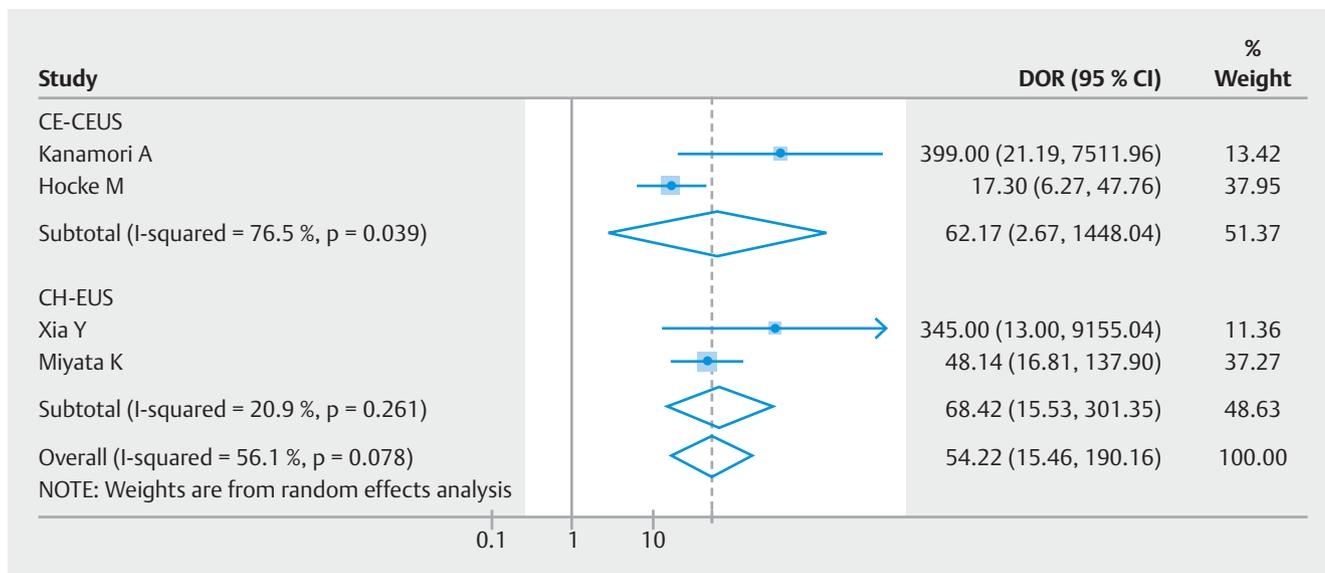
► **Fig. 4** Symmetric SROC curve.

included (n = 4) but also by significant differences in the characteristics of the patients and studies [27]. The number of cases in the four studies ranged from 34 to 143, without publication bias or other sample size related effects [19]. The disease prevalence ranged from 32.9% [23] to 67.6% [26]; however, pooled specificity tended to be lower in studies with higher disease prevalence, while no significant effect on pooled sensitivity was established [28].

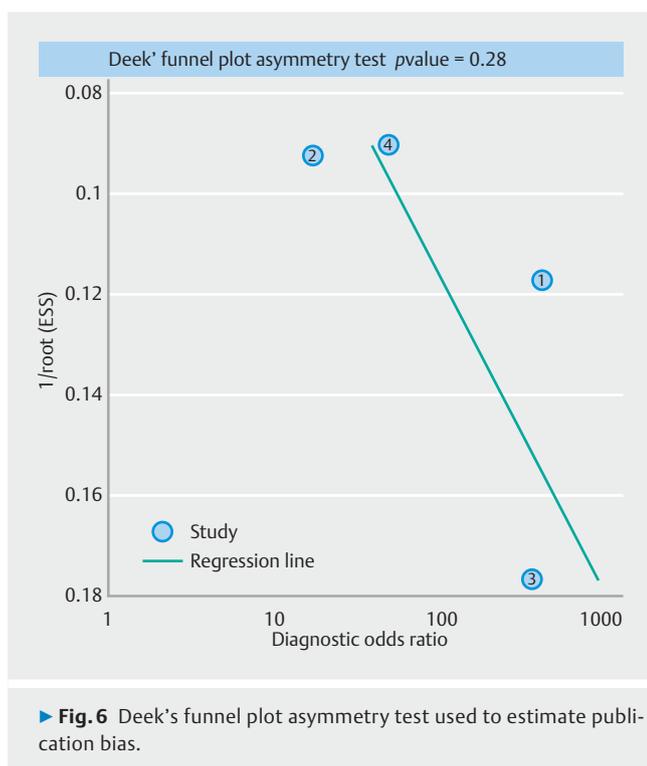
Although it was not possible to perform a meta-regression due to the small number of studies, we performed a subgroup analysis including only studies evaluating the performance of a dedicated contrast-harmonic mode (CH-EUS); pooled sensitivity increased significantly to 87.7% and pooled specificity to 91.8%. No significant heterogeneity was observed between these two studies [23, 26].

Recent guidelines recommended CH-EUS to distinguish benign from malignant pancreatic lesions but not to distinguish benign from malignant LNs [7, 29]. However, we believe that our meta-analysis has provided novel evidence in the field of CH-EUS for the differential diagnosis of LNs.

In fact, the two studies with CH-EUS reported increased pooled sensitivity and specificity compared to CE-EUS (sensitivity



► **Fig. 5** Forest plot of diagnostic odds ratio (DOR), according to the use of Doppler imaging mode (CE-EUS) and dedicated contrast-harmonic mode (CH-EUS).



ity 87.7% vs. 82.1%; specificity 91.8% vs. 90.7%). This difference can be attributed to the improved capability of detecting pathological alterations in the microvascular architecture of LNs. Color Doppler-enhanced CE-EUS is able to detect ultrasound contrast agents only at the level of arterioles and venules, while CH-EUS depicts the presence of microbubbles within the fine capillary network [30]. While a massive neoplastic involvement of LNs could be accurately detected even by CE-EUS, the involvement of smaller areas of LNs could not be iden-

tified by such an approximate technique. In contrast, tiny areas of capillary bed disruption appear clearly hypo-enhanced at CH-EUS while tumoral neoangiogenesis is shown as peripheral heterogeneous enhancement, with centripetal microvascularization and perfusion defects. Finally, the presence of neoplastic avascular necrosis could represent a limitation for CE-EUS and CH-EUS explaining the suboptimal sensitivity of both techniques.

EUS-FNA is the only method which can obtain pathological confirmation of the underlying disease. With mounting advances in cancer treatments and the advent of oncological target therapies, the characterization of a malignant LN is sometimes necessary to guide the subsequent clinical management. While the specificity of EUS-FNA for the characterization of extraluminal solid malignancies is estimated to be 100%, in the case of LN sampling, there is 1.1–5.3% of false positive results due to possible contamination as the needle passes through the neoplastic area or due to aspiration of luminal neoplastic cells [7, 8, 29]. In this setting, image enhancement techniques could be used to target malignant LNs with an increase in the diagnostic accuracy and a reduction in the number of needle passes and potential complications.

CH-EUS requires the injection of an ultrasound contrast agent (UCA) and the continuous observation of the target area for 2–3 minutes; this technique shows an optimal positive predictive value (>95%) in this setting. To date, no study has been designed to compare EUS-E and CH-EUS or investigate the diagnostic accuracy of the combination of the two techniques for the evaluation of LNs. Theoretically, the combination of the two techniques could overcome some particularly difficult cases such as benign necrotic LNs, hypo-enhanced on CH-EUS but red-green on EUS-E, or large malignant LNs, inhomogeneous on EUS-E with clear malignant portions on CH-EUS.

In this field, the presence of multiple LNs represents a well-known limitation of CH-EUS; indeed, while the crucial moment in the detection of lesions is the late venous phase, the characterization of LNs should be performed during the arterial and early venous phase. In cases where more than one LN is suspected, repeated UCA injections should be performed; we suggest starting the evaluation with B-mode and even EUS-elastography, and then using CH-EUS to study in detail the LNs with greater evidence of malignancy. Repeated injection of UCAs has been demonstrated to be a safe and reproducible technique; however, no study has used this combined approach to multiple LNs. Of course, this approach leads to an increase in length of procedures and increased costs; on the other hand, in the case of multiple suspected LNs, several EUS-FNAs need to be performed, changing the needle in any station, if possible, to reduce the risk of seeding.

The main limitation of this study is the small number of studies included and the relative number of cases evaluated. Although six studies were thoroughly evaluated [10,22–26], two had to be excluded due to differences between the techniques assessed (direct injection of contrast inside the LNs instead of a vascular perfusion technique) [10], and an overlap in population enrollment [22]. On the other hand, a strength of this study was in using pathology in all of the included studies as a reference standard for the diagnosis of malignant LNs. Quantitative analysis of these studies suggested significant heterogeneity among the results; in particular, the small number of included studies, differences among the techniques used (CE-EUS or CH-EUS), and population characteristics (known biliary-pancreatic cancers vs. abdominal masses of unknown origin) all have a bearing on the results. The prediction interval (Pr I) of the DOR was between 0.47 and 6298, confirming this hypothesis. On this basis, new evidence is required from original articles, using the data reported here as a starting point, and based on large homogeneous cohorts and different imaging techniques. Finally, none of the studies included here reported a calibration set in their analysis.

In our opinion, the inclusion of studies conducted with both first and second generation UCAs should not represent a significant limit; first generation UCAs present lower diagnostic accuracy compared to second generation ones (SonoVue and Sonazoid); however, in the setting of solid pancreatic masses, despite the larger number of studies conducted, some authors [31] have demonstrated with a meta-regression that the relative diagnostic odds ratio is not statistically significantly different between studies conducted with first and second generation UCAs. On this basis, the study by Kanamori et al. [24] was not excluded from the analysis, although it represents a further potential source of heterogeneity.

In the two studies conducted with CH-EUS, Sonazoid was used as UCA. No study is available in the literature directly comparing the diagnostic effect of the two second generation UCAs, in any setting. A recent meta-analysis [32] included studies with both first and second generation UCAs and identified a better diagnostic accuracy with Sonazoid for the characterization of focal liver lesions with trans-abdominal ultrasound (CEUS). The main difference between Sonazoid and SonoVue is

in the longer contrastographic effect of the former; however, in terms of differential diagnosis between benign and malignant LNs, the diagnostic advantage seems equivalent; indeed, the main differences appear in the arterial and early venous phases. On the other hand, the longer venous phase may give Sonazoid an advantage in other conditions, such as the detection of subtle lesions in large organs (such as the liver) or as a guide for EUS treatment.

In summary, these data provide interesting insights and new evidence in this field. The first conclusion is the recommendation against using contrast-enhanced EUS without a dedicated contrast harmonic mode; indeed, CE-EUS presents inadequate sensitivity (<85%). Second, this study recommends further larger studies evaluating the accuracy of CH-EUS, possibly in combination with elastography. Finally, despite the small number of available studies, this is the first level-1a evidence on the diagnostic accuracy of contrast-enhanced harmonic EUS for the characterization of LNs reporting good pooled sensitivity (87.7%) and optimal pooled specificity (91.8%), comparable with other image enhancement techniques (i.e. elastography) and even tissue sampling. Although new studies are required in this field, these findings indicate a role for CH-EUS in the diagnostic algorithm of suspected LNs.

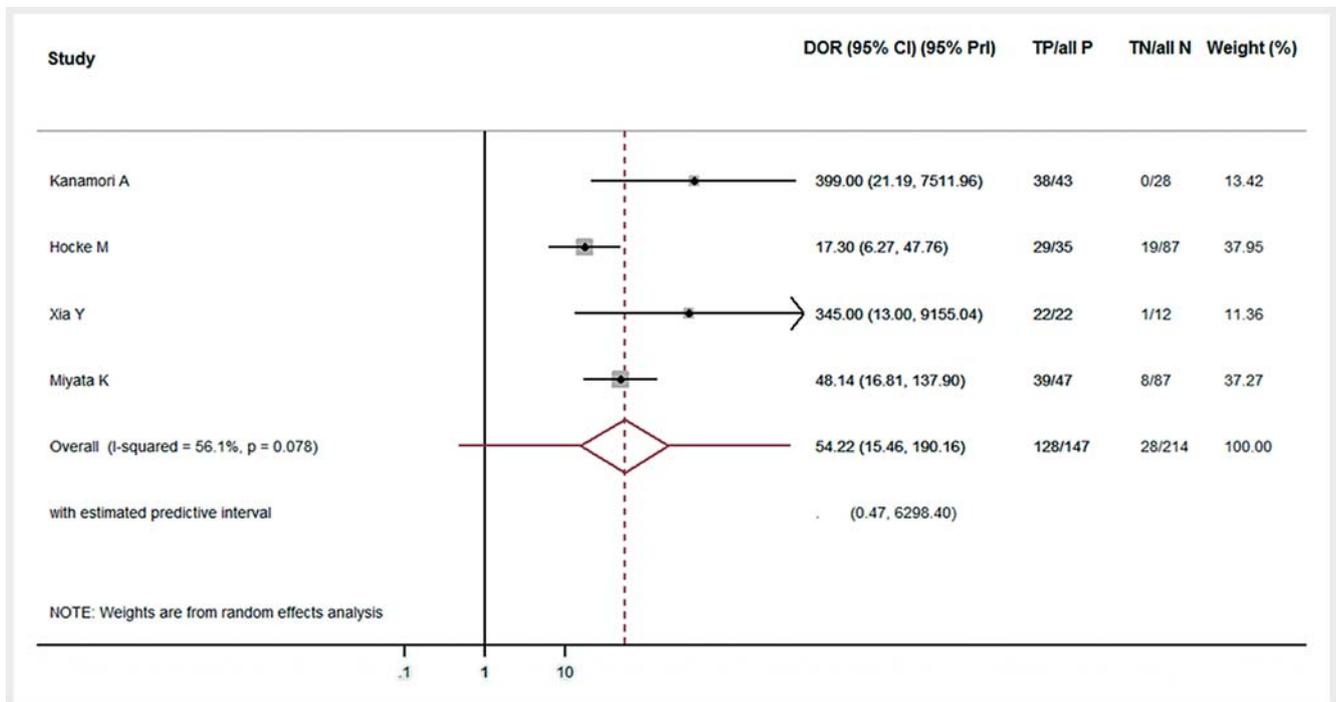
Competing interests

None

References

- [1] Catalano MF, Sivak MV Jr, Rice T et al. Endosonographic features predictive of lymph node metastasis. *Gastrointest Endosc* 1994; 40: 442–446
- [2] Bhutani MS, Hawes RH, Hoffman BJ. A comparison of the accuracy of echo features during endoscopic ultrasound (EUS) and EUS-guided fine-needle aspiration for diagnosis of malignant lymph node invasion. *Gastrointest Endosc* 1997; 45: 474–479
- [3] Faige DO. EUS in patients with benign and malignant lymphadenopathy. *Gastrointest Endosc* 2001; 53: 593–598
- [4] Cui XW, Hocke M, Jenssen C et al. Conventional ultrasound for lymph node evaluation, update 2013. *Z Gastroenterol* 2014; 52: 212–221
- [5] Dietrich CF, Jenssen C, Arcidiacono PG et al. Endoscopic ultrasound: Elastographic lymph node evaluation. *Endosc Ultrasound* 2015; 4: 176–190
- [6] Cui XW, Jenssen C, Saftoiu A et al. New ultrasound techniques for lymph node evaluation. *World J Gastroenterol* 2013; 19: 4850–4860
- [7] Jenssen C, Hocke M, Fusaroli P et al. EFSUMB Guidelines on Interventional Ultrasound (INVUS), Part IV – EUS-guided Interventions: General aspects and EUS-guided sampling (Long Version). *Ultraschall Med* 2016; 37: E33–76
- [8] Puli SR, Batapati Krishna Reddy J, Bechtold ML et al. Endoscopic ultrasound: it's accuracy in evaluating mediastinal lymphadenopathy? A meta-analysis and systematic review *World J Gastroenterol* 2008; 14: 3028–3037
- [9] Xu W, Shi J, Zeng X et al. EUS elastography for the differentiation of benign and malignant lymph nodes: a meta-analysis. *Gastrointest Endosc* 2011; 74: 1001–1009

- [10] Kojima S, Goto H, Hirooka Y et al. Differentiation of benign and malignant lymph nodes with contrast-enhanced echolymphography using endoscopic ultrasound-guided puncture. *Hepatogastroenterology* 2003; 50: 1285–1291
- [11] Fusaroli P, Napoleon B, Gincul R et al. The clinical impact of ultrasound contrast agents in EUS: a systematic review according to the levels of evidence. *Gastrointest Endosc* 2016; 84: 587–596.e10
- [12] University of Bristol. QUADAS-2. Available at: <http://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/>
- [13] Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5: 13
- [14] Irwig L, Bossuyt P, Glasziou P et al. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ* 2002; 324: 669–671
- [15] Raslich MA, Markert RJ, Stutes SA. Selecting and interpreting diagnostic tests. *Biochem Med* 2007; 17: 151–161
- [16] Glas AS, Lijmer JG, Prins MH et al. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003; 56: 1129–1135
- [17] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188
- [18] Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; 240: 1285–1293
- [19] 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: 882–893
- [20] Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. *Stat Med* 1993; 12: 1293–1316
- [21] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539–1558
- [22] Miyata T, Kitano M, Sakamoto H et al. 348 Role of contrast-enhanced harmonic EUS in differentiating malignant from benign lymphadenopathy. *Gastrointest Endosc* 2013; 77: AB142
- [23] Miyata T, Kitano M, Omoto S et al. Contrast-enhanced harmonic endoscopic ultrasonography for assessment of lymph node metastases in pancreaticobiliary carcinoma. *World J Gastroenterol* 2016; 22: 3381–3391
- [24] Kanamori A, Hirooka Y, Itoh A et al. Usefulness of contrast-enhanced endoscopic ultrasonography in the differentiation between malignant and benign lymphadenopathy. *Am J Gastroenterol* 2006; 101: 45–51
- [25] Hocke M, Menges M, Topalidis T et al. Contrast-enhanced endoscopic ultrasound in discrimination between benign and malignant mediastinal and abdominal lymph nodes. *J Cancer Res Clin Oncol* 2008; 134: 473–480
- [26] Xia Y, Kitano M, Kudo M et al. Characterization of intra-abdominal lesions of undetermined origin by contrast-enhanced harmonic EUS (with videos). *Gastrointest Endosc* 2010; 72: 637–642
- [27] Naaktgeboren CA, van Enst WA, Ochodo EA et al. Systematic overview finds variation in approaches to investigating and reporting on sources of heterogeneity in systematic reviews of diagnostic studies. *J Clin Epidemiol* 2014; 67: 1200–1209
- [28] Leeflang MM, Rutjes AW, Reitsma JB et al. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* 2013; 185: E537–544
- [29] Dumonceau JM, Deprez PH, Jenssen C et al. Indications, results, and clinical impact of endoscopic ultrasound (EUS)-guided sampling in gastroenterology: European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline – Updated January 2017. *Endoscopy* 2017; 49: 695–714
- [30] Hocke M, Ignee A, Dietrich C. Role of contrast-enhanced endoscopic ultrasound in lymph nodes. *Endosc Ultrasound* 2017; 6: 4–11
- [31] Gong TT, Hu DM, Zhu Q. Contrast-enhanced EUS for differential diagnosis of pancreatic mass lesions: a meta-analysis. *Gastrointest Endosc* 2012; 76: 301–309
- [32] Wu M, Li L, Wang J et al. Contrast-enhanced US for characterization of focal liver lesions: a comprehensive meta-analysis. *Eur Radiol* 2018; 28: 2077–2088



► **Supplementary Fig. 1** Forest plot of diagnostic odds ratio (DOR) with confidence interval (CI) and prediction interval (Pr I). DOR=Diagnostic Odds ratio; 95%CI=Confidence Interval at 95%; 95% Pr I=Prediction Interval at 95%; I-squared: between study heterogeneity according to the Higgins's test; P=P value referred to Q Cochrane test; gray squares: DOR of each study; Size of square: weight of each study in the analysis; Solid black line: 95% confidence interval for each study; Red diamond: the pooled DOR; Red line: sum of confidence interval plus prediction interval.