Pediatric Surgery International (2018) 34:363–380
https://doi.org/10.1007/s00383-018-4229-1

REVIEW ARTICLE

Early differential diagnosis methods of biliary atresia: a meta-analysis

Li Wang[1] · Yuan Yang[1] · Yang Chen[1] · Jianghua Zhan[1,2]

Accepted: 11 January 2018 / Published online: 3 February 2018

Abstract

Purpose To evaluate the accuracy of early differential diagnosis methods of biliary atresia in patients with infantile cholestasis.

Methods We searched PubMed, EMBASE and the Web of Science databases for articles evaluated the early differential diagnosis methods of biliary atresia. The methodological quality of each study was assessed with version 2 of the Quality Assessment of Diagnostic Accuracy Studies tool. Two reviewers extracted data independently. Pooled sensitivity, specificity, positive likelihood ratio (LR +), negative likelihood ratio (LR −), diagnostic odds ratio (DOR) with 95% CIs were calculated to assess each diagnosis method.

Results A total of 38 articles were included. Summary sensitivity and specificity were 77% (95% CI 74–80%) and 93% (95% CI 91–94%), respectively, for B-US in 23 studies; 96% (95% CI 92–98%) and 58% (95% CI 51–65%), respectively, for MRCP in five studies; 87% (95% CI 82–91%) and 78% (95% CI 74–82%), respectively, for acholic stool in seven studies; 84% (95% CI 78–89%) and 97% (95% CI 97–98%), respectively, for serum liver function test in seven studies; 96% (95% CI 94–97%) and 73% (95% CI 70–76%), respectively, for hepatobiliary scintigraphy in 18 studies; 98% (95% CI 96–99%) and 93% (95% CI 89–95%), respectively, for percutaneous liver biopsy in 11 studies.

Conclusion The accuracy rate of percutaneous liver biopsy is better than all of the noninvasive methods. Take into consideration the advantages and disadvantages of the six methods, combination of multidisciplinary noninvasive diagnosis methods is the first choice for differential diagnosis of BA from other causes of neonatal cholestasis.

Keywords Biliary atresia · Early differential diagnosis · Noninvasive diagnosis methods

Introduction

Biliary atresia (BA) is a disease of unknown etiology that affects both the extrahepatic and the intrahepatic bile ducts, leading to progressive obliteration of the biliary tree [1], causing severe cholestasis and biliary cirrhosis, that leads finally to death in the first years of life. The recommended treatment of BA is sequential: in the first and second month of life, the Kasai portoenterostomy, or its technical variants, aims to restore the biliary flow to the intestine; in the case of failure of the operation and/or life-threatening complications of the biliary cirrhosis, liver transplantation (LT) may eventually be needed [2]. Current general conclusion is that the earlier the Kasai portoenterostomy performed, the better the effect. So early diagnosis of BA is very important for the BA infants' long-term free-transplant survival. The objective of our study is to analyze the accuracy of different diagnosis methods for diagnosing BA.

Methods

Literature search

We searched PubMed, EMBASE and the Web of Science databases for articles published up to July 2017, with searching ((((((diagnosis[Title/Abstract]) OR diagnose[Title/Abstract]) OR diagnostic[Title/Abstract]) OR screening[Title/Abstract])) AND ((((((((((((((((((((Ultrasonograph[Title/Abstract]) OR Echography[Title/Abstract]) OR Ultrasound Imaging[Title/Abstract]) OR ultrasound[Title/Abstract]) OR Imaging, Ultrasound[Title/Abstract]) OR

✉ Jianghua Zhan
zhanjianghuatj@163.com

[1] Department of Pediatric Surgery, Tianjin Children's Hospital, Tianjin 300134, China

[2] Tianjin Pediatrics Research Institute, Tianjin 300134, China

Ultrasound Imagings[Title/Abstract]) OR Sonography, Medical[Title/Abstract]) OR Medical Sonography[Title/Abstract]) OR Diagnostic Ultrasound[Title/Abstract]) OR Ultrasound, Diagnostic[Title/Abstract]) OR Echotomography[Title/Abstract]) OR Diagnosis, Ultrasonic[Title/Abstract]) OR Diagnosis, Ultrasonic[Title/Abstract]) OR Ultrasonic Tomography[Title/Abstract]) OR "Ultrasonography"[Mesh])) OR (((((((Cholangiopancreatography, Magnetic Resonance[Title/Abstract]) OR Magnetic Resonance Cholangiopancreatography[Title/Abstract]) OR MRCP[Title/Abstract]) OR MR Cholangiopancreatography[Title/Abstract]) OR Magnetic Resonance Cholangiography[Title/Abstract]) OR MR Cholangiography[Title/Abstract])) OR "Cholangiopancreatography, Magnetic Resonance"[Mesh])) OR (((((acholic stool[Title/Abstract]) OR pale stool[Title/Abstract]) OR clay stool[Title/Abstract]) OR stool color card[Title/Abstract]) OR stool colour card[Title/Abstract])) OR (((((((Liver Function Tests[Title/Abstract]) OR Function Test, Liver[Title/Abstract]) OR Function Tests, Liver[Title/Abstract]) OR Liver Function Test[Title/Abstract]) OR Test, Liver Function[Title/Abstract]) OR Tests, Liver Function[Title/Abstract]) OR "Liver Function Tests"[Mesh])) OR ((((Hepatobiliary scintigraphy[Title/Abstract]) OR Technetium Tc 99 m Lidofenin[Title/Abstract]) OR HIDA[Title/Abstract]) OR hepatobiliary scintiscanning[Title/Abstract])) OR (((((liver[Title/Abstract]) OR hepatic[Title/Abstract]) OR hepatology[Title/Abstract])) AND ((((biopsy[Title/Abstract]) OR pathology[Title/Abstract]) OR pathological[Title/Abstract]) OR histopathology[Title/Abstract]))))) AND ((((((((((Biliary Atresia[Title/Abstract]) OR Biliary Atresia, Extrahepatic[Title/Abstract]) OR Atresia, Extrahepatic Biliary[Title/Abstract]) OR Atresias, Extrahepatic Biliary[Title/Abstract]) OR Biliary Atresias, Extrahepatic[Title/Abstract]) OR Extrahepatic Biliary Atresia[Title/Abstract]) OR Extrahepatic Biliary Atresias[Title/Abstract]) OR Atresia, Biliary[Title/Abstract]) OR Familial Extrahepatic Biliary Atresia[Title/Abstract]) OR Idiopathic Extrahepatic Biliary Atresia[Title/Abstract]) OR "Biliary Atresia"[Mesh]).

Inclusion criteria

The inclusion criteria for the identified articles were as follows: (1) diagnostic test accuracy (DTA) studies evaluating sensitivity and specificity of at least one of B-US, MRCP, acholic stool, serum liver function test, hepatobiliary scintigraphy and percutaneous liver biopsy, (2) articles were published in full texts in English and (3) studies with sufficient information for analysis.

Exclusion criteria

The exclusion criteria for the identified articles were as follows: (1) letters, reviews, case reports, conference abstracts, editorials, expert opinion reviews and abstracts, (2) data of sensitivity, specificity is incorrect or insufficient for analysis or evaluated by more than one researcher without a consensus, (3) screening studies with a large population without cholestasis and (4) studies with overlapping cases and data. If the cases of two or more studies overlap each other, give priority to the study with more diagnosis methods evaluated and whose cases are more if diagnosis methods are the same.

Screening

Screening was performed in duplicates, independently, by two researchers at all stages. Disagreements in study selection between the two reviewers were resolved through consensus.

Data extraction

Data were extracted on study characteristics (e.g. study period, design, sample size, and location of the study), study sample characteristics (e.g. age at diagnosis), and diagnostic data (e.g. true positives, true negatives, false positives, false negatives, sensitivity, specificity). Extract the data of the commonest criteria if a study evaluates two or more criteria of a diagnosis method.

Quality assessment

Using the version 2 of the Quality Assessment of Diagnostic Test Accuracy Studies (QUADAS-2) tool [3], quality of studies included in our study was assessed by two researchers. All disagreements were discussed and consensus was reached.

Data analysis

Heterogeneity was assessed using the I^2 statistic index, with a value $> 50\%$ considered to represent substantial heterogeneity. When a great heterogeneity was noted, heterogeneity by a "threshold effect" was analyzed using Spearman correlation coefficients ($p < 0.05$ represents threshold effect). We used a random effects model for the primary meta-analysis to obtain a summary estimate for sensitivity, specificity, positive likelihood ratio (LR +), negative likelihood ratio (LR −), diagnostic odds ratio (DOR) with 95% CIs, positive predictive value (PPV) and negative predictive value (NPV) of each diagnosis method. If there is not substantial heterogeneity among studies, pool data by fixed effects model are

done. Then, we constructed a summary receiver operator characteristic curve (SROC) and calculated the area under curve (AUC).

Subgroup analyses are performed by following covariates: (1) study design (prospective versus retrospective), (2) cases (≤ 60 versus > 60) and (3) final diagnosis method (intraoperative cholangiography with/without surgery or histology versus surgery and/or histology). In addition, publication bias is assessed by a Deeks funnel plot ($p < 0.05$ was considered representative of significant statistical publication bias). We used the Meta-DiSc 14.0 and Stata 14 to perform the statistical analyses.

Results

Study selection

Initial search of PubMed, EMBASE and the Web of Science databases yielded 1489 studies. Figure 1 shows the flow diagram of the study selection. Of the 80 full-text articles assessed for final eligibility, 42 are excluded (4 without full text, 3 non-English, 6 without sufficient data, 1 evaluated by two or more researchers without a consensus, 10 incorrect data, 13 with overlapping cases, 5 screening study).

Study characteristics

A total of 3053 patients were included within in the 38 studies [4–41] (Table 1) included for analysis, 25 studies were prospective, 10 were retrospective and 3 could not be clearly identified. Studies were published between 1985 and 2016. Studies most commonly originated from the China (8/38 studies), followed by Korea (7/38 studies) and USA (6/38 studies). The overall quality of the included studies assessed by the QUADAS-2 (Table 2), was moderate, and all of the studies was low risk of bias on 5 or more of the 7 items.

There were 21 articles that final diagnosis methods of BA explicitly included intraoperative cholangiography (IC). Of the 21 articles, 6 were diagnosed only by IC, 6 are diagnosed by IC and surgery, 8 were diagnosed by IC and histology, 1 was diagnosed by IC, surgery and histology. Besides, 1 article that final diagnosis method did not include IC, 8 articles that are final diagnosed

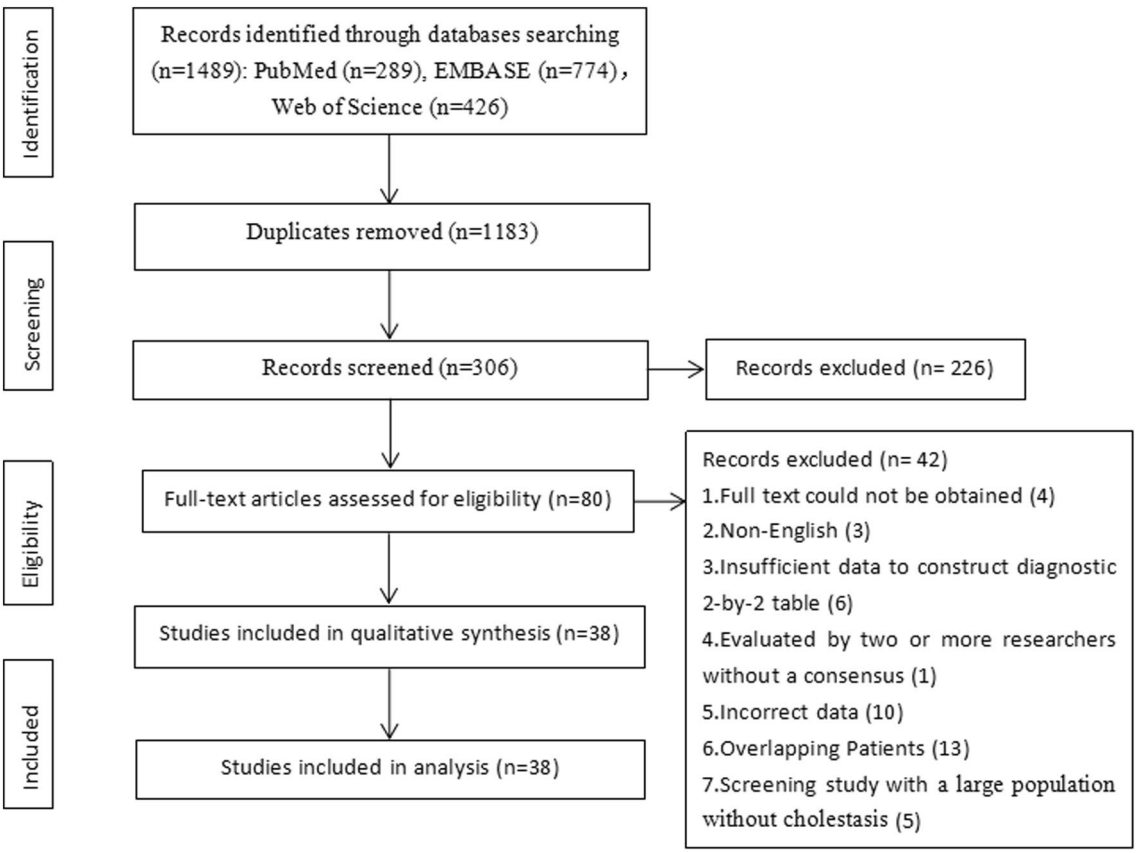

Fig. 1 Flow diagram of the study selection process

Table 1 The characteristics of the studies included in this study

References	Year	Country	Total patients	Patients with BA	Age at diagnosis (day)	Study design
Spivak [4]	1985	USA	28	7	59	R
Tolia [5]	1986	USA	28	10	NA	U
Cox [6]	1987	USA	33	9	NA	P
Park [7]	1997	Korea	73	25	12–120	P
Lee [8]	2000	China	152	49	NA	P
Tan [9]	2000	Singapore	60	12	NA	P
Farrant [10]	2001	UK	158	38	NA	P
Sun [11]	2001	China	182	151	NA	P
Han [12]	2002	Korea	47	23	65.9 (15–210)	P
Azuma [13]	2003	Japan	30	23	62	U
Lee [14]	2003	Korea	86	20	53.72	P
Visrutaratna [15]	2003	Thailand	46	23	NA	P
Ryeom [16]	2005	Korea	23	4	69 (24–139)	P
Dehghani [17]	2006	Iran	65	19	62 ± 17	P
Hu [18]	2006	China	52	18	NA	P
Humphrey [19]	2007	UK	90	30	51.1	P
Kim [20]	2007	Korea	68	38	61	P
Takamizawa [21]	2007	Japan	85	48	47*	P
Wongsawasdi [22]	2008	Thailand	61	31	88.6	P
Lee [23]	2009	Korea	64	29	51 ± 24(3–91)	P
Poddar [24]	2009	India	101	60	2.8 ± 1.7 mon	P
Rouzrokh [25]	2009	Iran	42	18	39	R
Yang [26]	2009	China	69	34	62 ± 14 (31–121)	R
Liu [27]	2010	China	84	29	NA	P
Aziz [28]	2011	USA	35	15	2.1 mon	P
Jensen [29]	2012	USA	68	19	63	R
El-Guindi [30]	2013	Egypt	76	27	68.52	P
Jiang [31]	2013	China	51	23	2.90 mon	P
Kwatra [32]	2013	USA	186	43	48*	R
Boskovic [33]	2014	Serbia	109	72	NA	R
El-Guindi [34]	2014	Egypt	60	30	63.67 ± 12.73	P
Liu [35]	2014	China	190	104	69*	P
Guan [36]	2015	China	197	107	63.9	R
Jancelewicz [37]	2015	Canada	212	45	NA	R
Lee [38]	2015	Korea	100	46	55	R
Brittain [39]	2016	Denmark	47	14	NA	R
Rafeey [40]	2016	Iran	30	18	54.66	U
Zhen [41]	2016	China	80	40	NA	P

NA not available, *R* retrospective, *P* prospective, *U* unclear

*Median

by surgery with/without histology and 8 articles did not mention how to final diagnose BA. Of the 38 articles, 25 articles performed the diagnostic test when the reference test results were unknown, 10 articles knew the reference test results in advance and 3 articles did not mention.

Diagnostic values

B-US

Data on the diagnostic performance of the B-US were collected from 23 studies with 1774 patients (Table 3). The Spearman correlation coefficient was 0.033, p value was 0.883, indicating no threshold effect. The diagnostic odds

Table 2 Risk of Bias assessed by QUADAS-2

Author	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Spivak [4]	Low	High	Low	Low	Low	Low	Low
Tolia [5]	Unclear	High	Low	Low	Low	Low	Low
Cox [6]	Unclear	Low	Low	Low	Low	Low	Low
Park [7]	Low	Low	Low	Low	Low	Low	Low
Lee [8]	Low	Low	Low	Low	Low	Low	Low
Tan [9]	Low	Low	Low	Low	Low	Low	Low
Farrant [10]	Unclear	Low	Unclear	Low	Low	Low	Low
Sun [11]	Low	High	Low	Low	Low	Low	Low
Han [12]	Low	Low	Low	Low	Low	Low	Low
Azuma [13]	Low	Low	Low	Low	Low	Low	Low
Lee [14]	Low	High	Low	Low	Low	Low	Low
Visrutaratna [15]	Low	Low	Low	High	Low	Low	Low
Ryeom [16]	Low	Low	Low	Low	Low	Low	Low
Dehghani [17]	Low	Low	Low	Low	Low	Low	Low
Hu [18]	Low	Low	Low	Low	Low	Low	Low
Humphrey [19]	Low	High	Low	Low	Low	Low	Low
Kim [20]	Low	High	Low	Low	Low	Low	Low
Takamizawa [21]	Low	Low	Low	Low	Low	Low	Low
Wongsawasdi [22]	Unclear	Low	Low	High	Low	Low	Low
Lee [23]	Low	Low	Low	Low	Low	Low	Low
Poddar [24]	Low	Low	Low	Low	Low	Low	Low
Rouzrokh [25]	Unclear	High	Low	Low	Low	Low	Low
Yang [26]	Low	High	Low	Low	Low	Low	Low
Liu [27]	Low	Low	Low	Low	Low	Low	Low
Aziz [28]	Low	Low	Low	Low	Low	Low	Low
Jensen [29]	Unclear	High	Low	Unclear	Low	Low	Low
El-Guindi [30]	Unclear	Low	Low	Low	Low	Low	Low
Jiang [31]	Low	Low	Low	Low	Low	Low	Low
Kwatra [32]	Low	High	Low	Low	Low	Low	Low
Boskovic [33]	Unclear	High	Low	Unclear	Low	Low	Low
El-Guindi [34]	Low	Low	Low	Low	Low	Low	Low
Liu [35]	Low	Low	Low	Low	Low	Low	Low
Guan [36]	Low	High	Low	Low	Low	Low	Low
Jancelewicz [37]	Unclear	High	Low	Low	Low	Low	Low
Lee [38]	Unclear	High	Low	Low	Low	Low	Low
Brittain [39]	Low	High	Low	Low	Low	Low	Low
Rafeey [40]	Low	Low	Low	Low	Low	Low	Low
Zhen [41]	Low	High	Low	Low	Low	Low	Low

ratio was 46.02 (95% CI 22.71–93.27), I^2 was 71.4%, showing high heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of B-US is shown in Figs. 2, 3. The sensitivities and specificities of individual studies varied from 31 to 99% and from 71 to 100%, respectively. The B-US showed pooled sensitivity of 77% (95% CI 74–80%), specificity of 93% (95% CI 91–94%), LR + of 8.48 (95% CI 5.52–13.02) and LR − of 0.28 (95% CI 0.20–0.39). The summary ROC curves of B-US for the diagnosis of biliary atresia are illustrated in Fig. 4. The summary ROC curve was symmetric, and the AUC was 0.9396, Q was 0.8770. The PPV is 88.6% and the NPV is 85.3%.

Table 3 Diagnostic profile of various diagnostic methods

Studies	Cases	Sensitivity (%)	Specificity (%)	TP	FP	FN	TN
B-US							
Cox [6]	33	67	83	6	4	3	20
Park [7]	73	85	100	17	0	3	43
Tan [9]	60	83.3	100	10	0	2	48
Farrant [10]	158	91.90	96.70	34	4	3	117
Sun [11]	182	99.3	83.9	150	5	1	26
Azuma [13]	30	83	71	19	2	4	5
Lee [14]	86	80	98	16	1	4	65
Visrutaratna [15]	46	95.70	73.90	22	6	1	17
Ryeom [16]	23	75	89	3	2	1	17
Dehghani [17]	65	52.6	76.1	10	11	9	35
Humphrey [19]	90	73	100	22	0	8	60
Kim [20]	68	58	96	22	2	16	45
Takamizawa [21]	85	85	95	41	2	7	35
Lee [23]	85	62	100	18	0	11	35
Poddar [24]	101	71	82	25	12	10	54
Rouzrokh [25]	42	72	92	13	2	5	22
Yang [26]	69	50	82.86	17	6	17	29
Aziz [28]	35	60	95	9	1	6	19
El-Guindi [29]	54	59.30	88.90	16	3	11	24
Jiang [31]	51	91	93	21	2	2	26
El-Guindi [34]	60	63.3	86.7	19	4	11	26
Jancelewicz [37]	192	31	99	14	1	31	146
Lee [38]	100	100	94.40	46	3	0	51
MRCP							
Han [12]	47	100	96	23	1	0	23
Ryeom [16]	23	100	58	4	8	0	11
Hu [18]	52	94.4	88.24	17	4	1	30
Yang [26]	69	85.29	57.14	29	15	5	20
Liu [35]	190	99.04	36.05	103	55	1	31
Acholic stool							
Wongsawasdi [22]	61	58	100	18	0	13	30
Poddar [24]	101	86	76	30	16	5	50
Rouzrokh [25]	42	100	83	18	4	0	20
El-Guindi [30]	54	92.60	55.60	25	12	2	15
El-Guindi [34]	60	93.3	56.7	28	13	2	17
Jancelewicz [37]	212	89	81	40	32	5	135
Zhen [41]	80	95	85	38	6	2	34
Serum liver function test							
Dehghani [17]	65	68.4	43.5	13	26	6	20
Wongsawasdi [22]	58	65.50	96.60	19	1	10	28
Poddar [24]	101	91	32	32	45	3	21
El-Guindi [30]	54	74.10	77.80	20	6	7	21
El-Guindi [34]	60	76.7	80	23	6	7	24
Lee [38]	96	87.0	76	40	12	6	38
Rafeey [40]	60	76.70	80	23	6	7	24
Hepatobiliary scintigraphy							
Spivak [4]	28	100	43	7	12	0	9
Tolia [5]	32	100	54.4	10	10	0	12
Cox [6]	33	100	67	9	8	0	16
Park [7]	71	96	35	24	30	1	16

Table 3 (continued)

Studies	Cases	Sensitivity (%)	Specificity (%)	TP	FP	FN	TN
Lee [8]	152	100	86	49	14	0	89
Tan [9]	38	91.7	76.9	11	6	1	20
Ryeom [16]	21	100	65	4	6	0	11
Dehghani [17]	65	84.2	47.8	16	24	3	22
Wongsawasdi [22]	54	100	92	29	2	0	23
Poddar [24]	34	100	86	13	3	0	18
Rouzrokh [25]	42	100	87.5	18	3	0	21
Yang [26]	69	88.24	45.71	30	19	4	16
Liu [27]	84	100	74.5	29	14	0	41
Jensen [29]	68	95	57	18	21	1	28
Kwatra [32]	186	100	93	43	10	0	133
Guan [36]	197	90.65	78.89	97	19	10	71
Jancelewicz [37]	202	100	75	41	41	0	120
Brittain [39]	47	100	63.6	14	12	0	21
Percutaneous liver biopsy							
Tolia [5]	33	96	90	22	1	1	9
Cox [6]	24	100	87	9	2	0	13
Park [7]	44	90	96	18	1	2	23
Dehghani [17]	65	100	95.7	19	2	0	44
Wongsawasdi [22]	25	93.80	100	15	0	1	9
Poddar [24]	69	100	100	35	0	0	34
Yang [26]	69	100	94.29	34	2	0	33
El-Guindi [30]	52	100	88	27	3	0	22
Boskovic [33]	109	98.6	100	71	0	1	37
El-Guindi [34]	60	96.7	86.7	29	4	1	26
Jancelewicz [37]	96	98.0	84	39	9	1	47

MRCP

Data on the diagnostic performance of the MRCP were collected from five studies with 381 patients (Table 3). The Spearman correlation coefficient was 0.000, p value was 1.000, indicating no threshold effect. The diagnostic odds ratio was 43.49 (95% CI 8.53–221.83), I^2 was 64.3%, showing high heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of MRCP is shown in Figs. 5, 6. The sensitivities and specificities of individual studies varied from 85 to 100% and from 36 to 96%, respectively. The MRCP showed summary sensitivity of 96% (95% CI 92–98%), specificity of 58% (95% CI 51–65%), LR + of 2.96 (95% CI 1.58–5.55) and LR − of 0.08 (95% CI 0.02–0.30). The summary ROC curves of MRCP for the diagnosis of biliary atresia are illustrated in Fig. 7. The summary ROC curve was symmetric, and the AUC was 0.9409, Q was 0.8788. The PPV is 68.0% and the NPV is 94.3%.

Acholic stool

Data on the diagnostic performance of the acholic stool were collected from seven studies with 610 patients (Table 3). The Spearman correlation coefficient was 0.071, p value was 0.879, indicating no threshold effect. The diagnostic odds ratio was 30.66 (95% CI 17.48–53.76), I^2 was 0.0%, showing low heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of acholic stool is shown in Figs. 8, 9. The sensitivities and specificities of individual studies varied from 58 to 100% and from 56 to 100%, respectively. The acholic stool showed pooled sensitivity of 87% (95% CI 82–91%), specificity of 78% (95% CI 74–82%), LR + of 3.87 (95% CI 3.17–4.72) and LR − of 0.17 (95% CI 0.12–0.23). The summary ROC curves of acholic stool for the diagnosis of biliary atresia are illustrated in Fig. 10. The summary ROC curve was symmetric, and the AUC was 0.9238, Q was 0.8578. The PPV is 70.4% and the NPV is 91.2%.

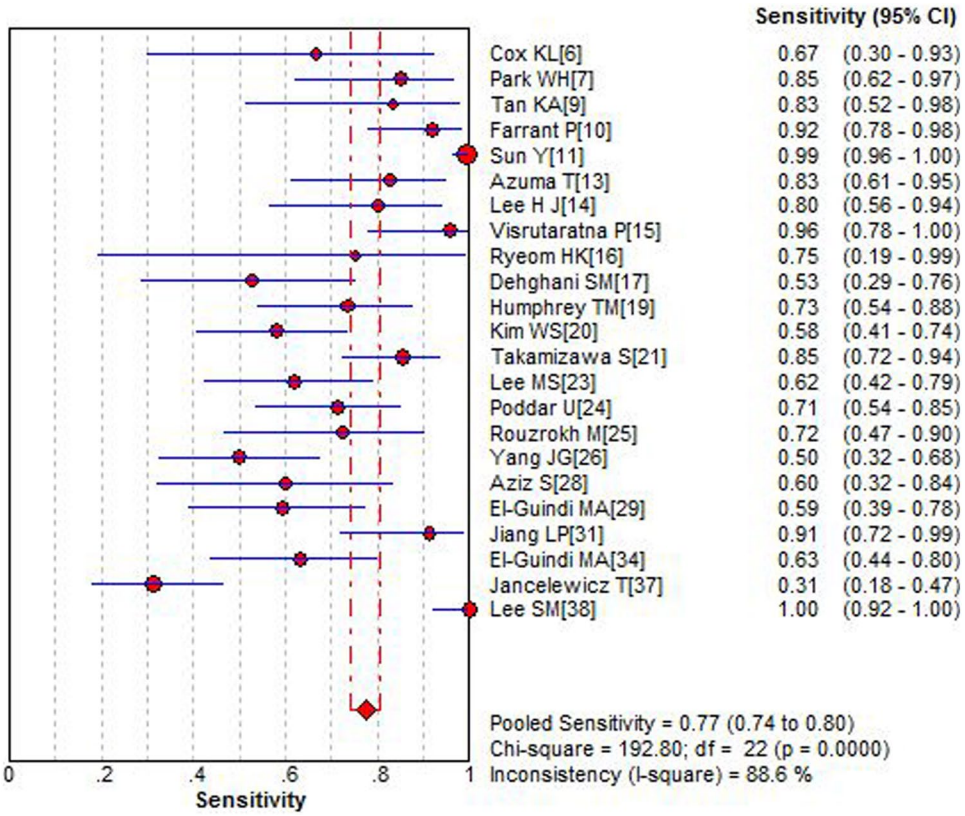

Fig. 2 The forest plots of pooled sensitivity for B-US

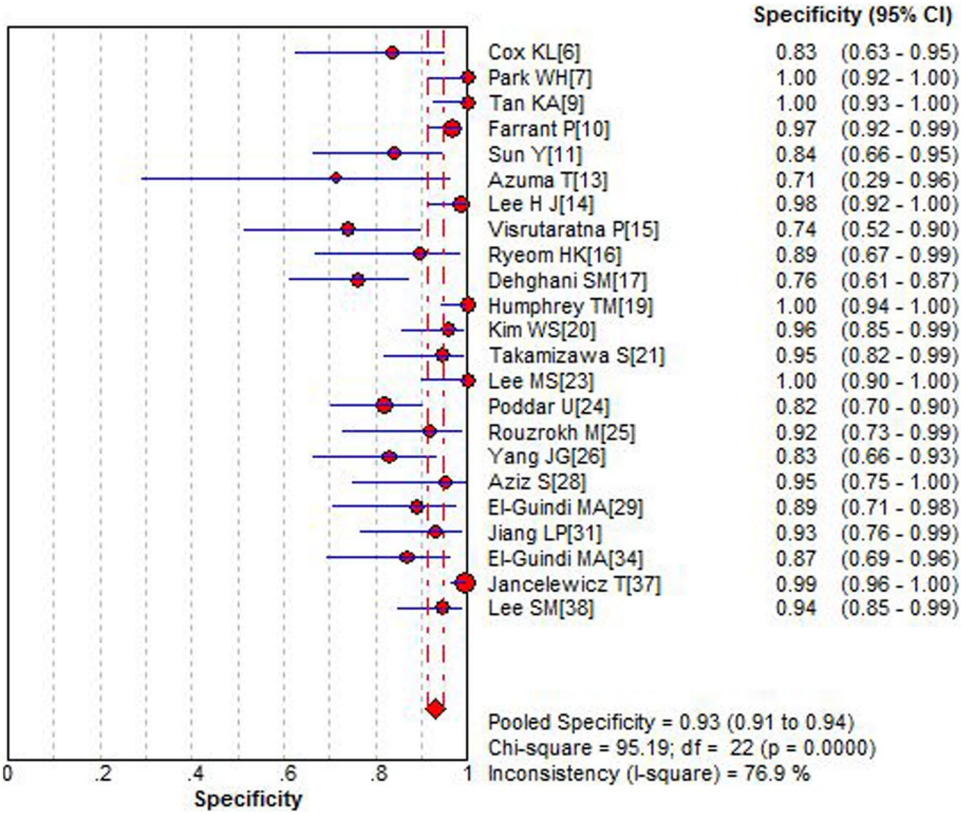

Fig. 3 The forest plots of pooled specificity for B-US

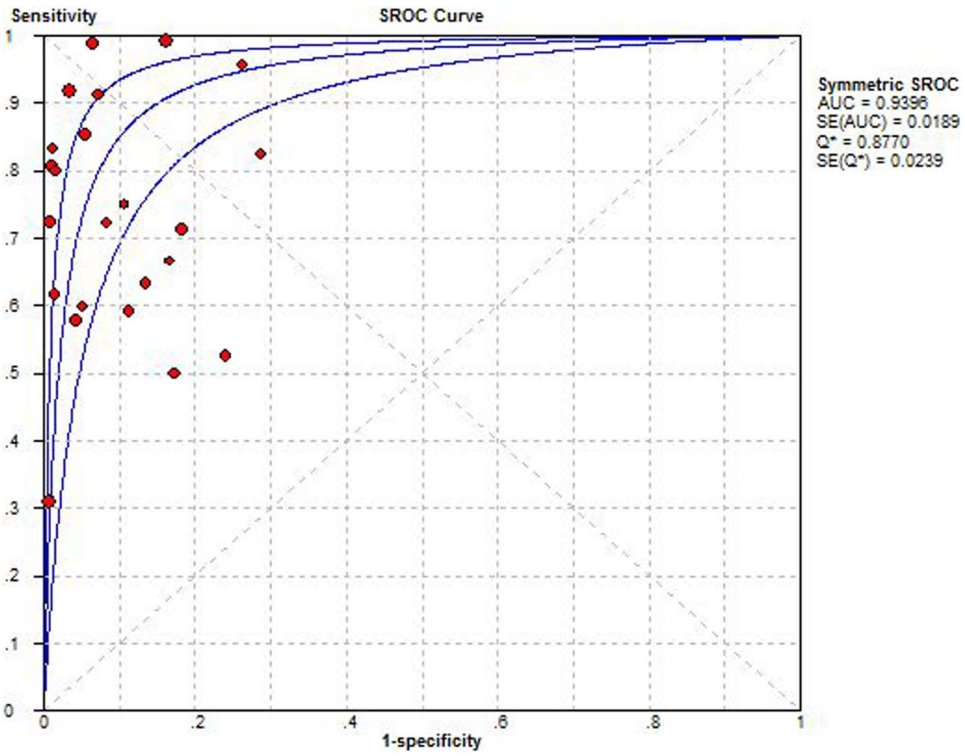

Fig. 4 Summary receiver operating characteristics (SROC) curve of the B-US

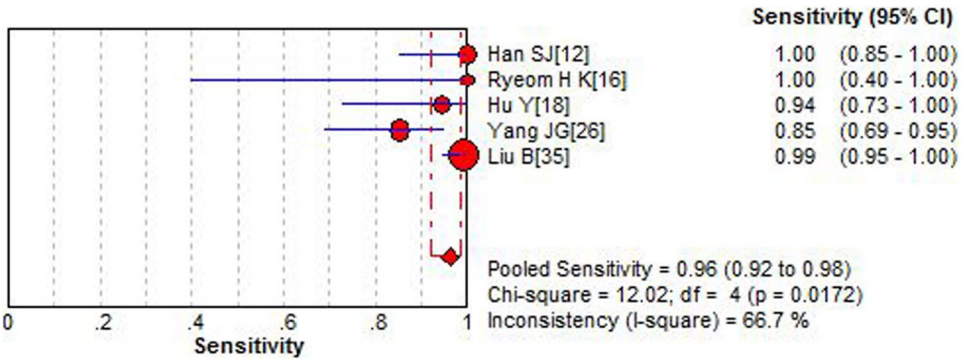

Fig. 5 The forest plots of pooled sensitivity for MRCP

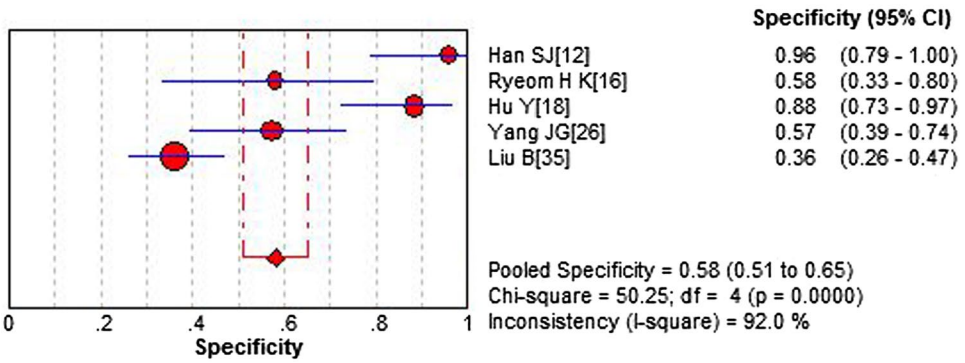

Fig. 6 The forest plots of pooled specificity for MRCP

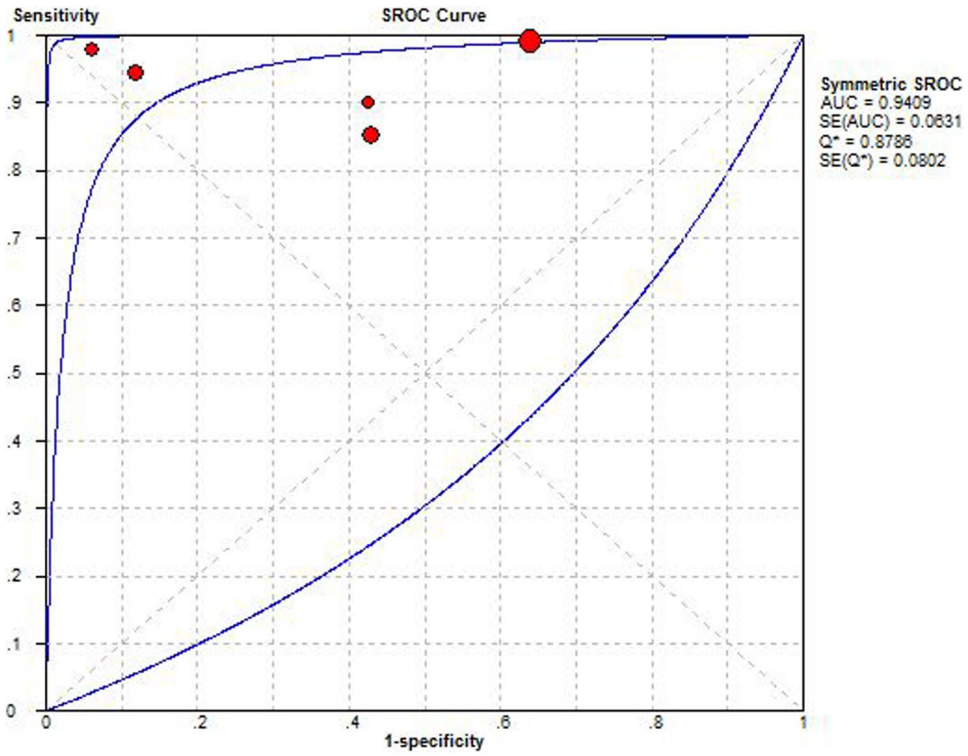

Fig. 7 Summary receiver operating characteristics (SROC) curve of the MRCP

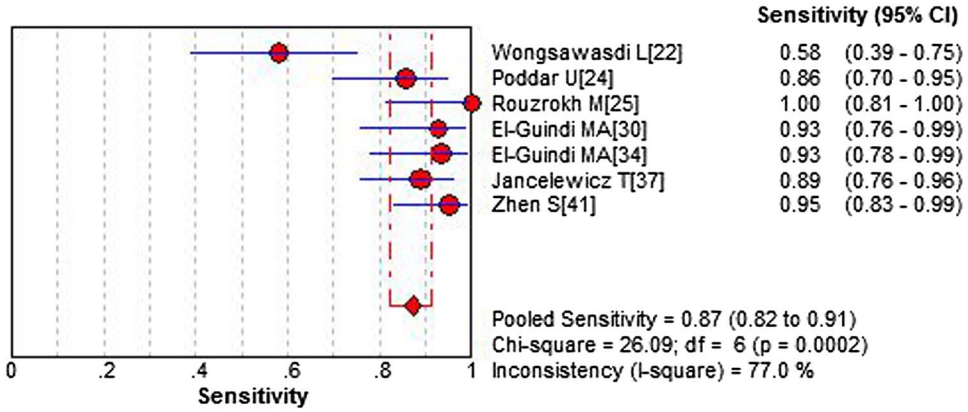

Fig. 8 The forest plots of pooled sensitivity for acholic stool

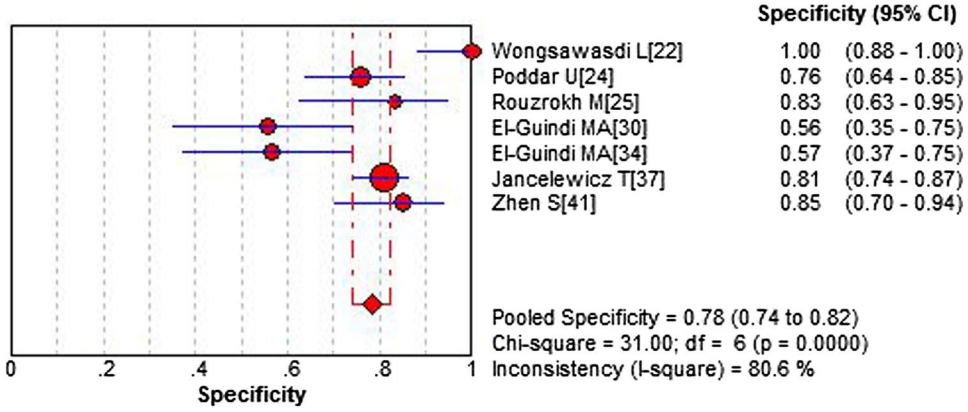

Fig. 9 The forest plots of pooled specificity for acholic stool

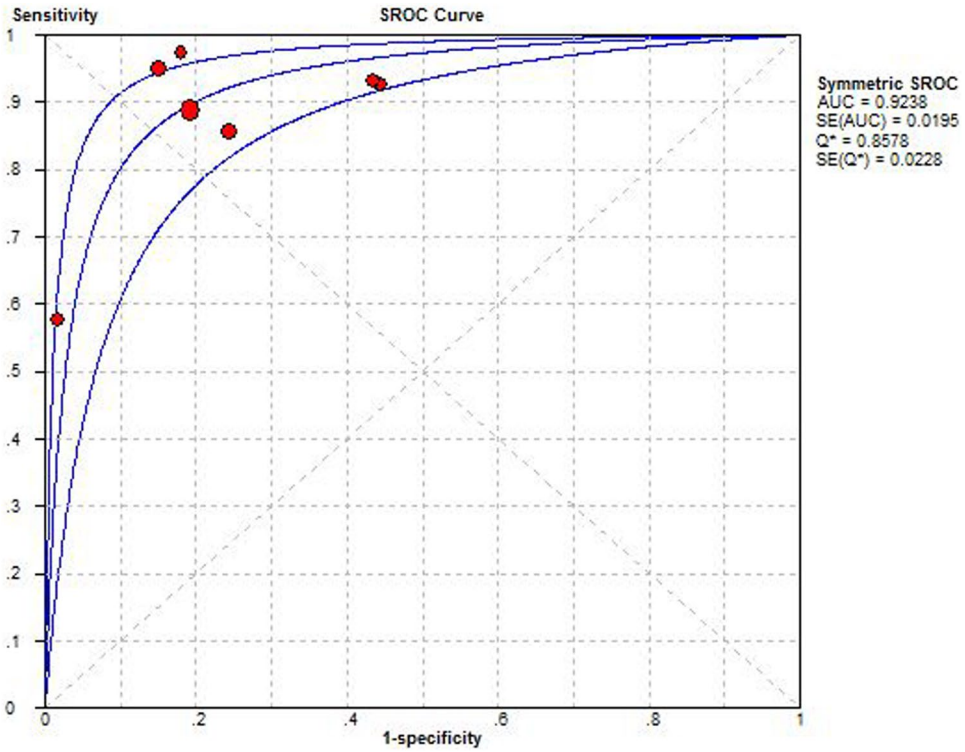

Fig. 10 Summary receiver operating characteristics (SROC) curve of the acholic stool

Serum liver function test

Data on the diagnostic performance of the serum liver function test were collected from seven studies with 494 patients (Table 3). The Spearman correlation coefficient was 0.036, *p* value was 0.939, indicating no threshold effect. The diagnostic odds ratio was 19.00 (95% CI 4.99–72.30), I^2 was 82.4%, showing high heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of serum liver function test is shown in Figs. 11, 12. The sensitivities and specificities of individual studies varied from 66 to 100% and from 32 to 98%, respectively. The serum liver function test showed pooled sensitivity of 84% (95% CI 78–89%), specificity of 97% (95% CI 97–98%), LR + of 4.73 (95% CI 0.66–34.02) and LR − of 0.26 (95% CI 0.14–0.51). The summary ROC curves of serum liver function test for the diagnosis of biliary atresia are illustrated in Fig. 13. The summary ROC curve was symmetric, and the AUC was 0.9080, *Q* was 0.8399. The PPV is 62.5% and the NPV is 79.3%.

Hepatobiliary scintigraphy

Data on the diagnostic performance of the hepatobiliary scintigraphy were collected from 18 studies with 1423 patients (Table 3). The Spearman correlation coefficient was −0.613, *p* value was 0.007, indicating threshold effect. The

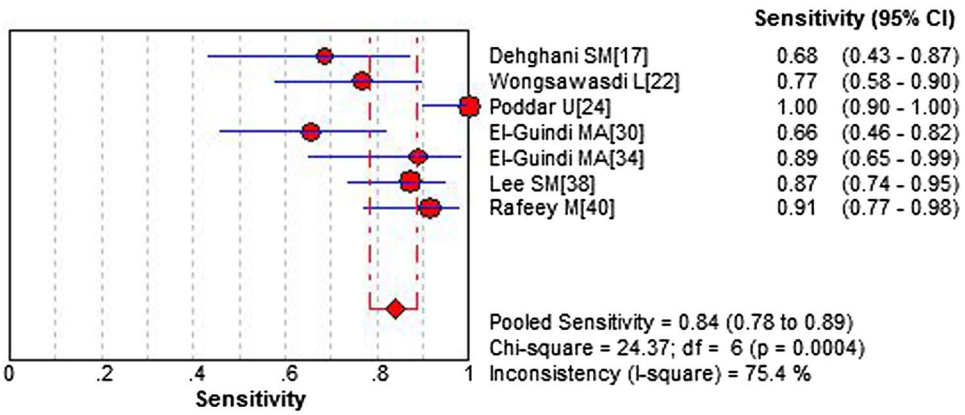

Fig. 11 The forest plots of pooled sensitivity for serum liver function test

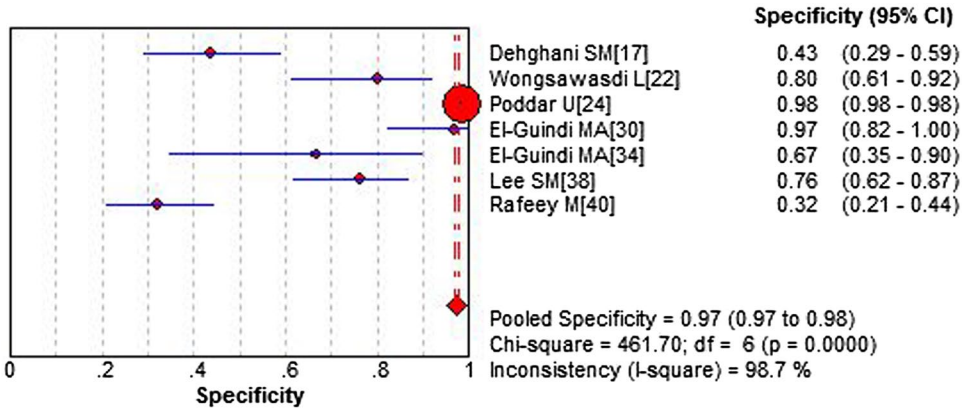

Fig. 12 The forest plots of pooled specificity for serum liver function test

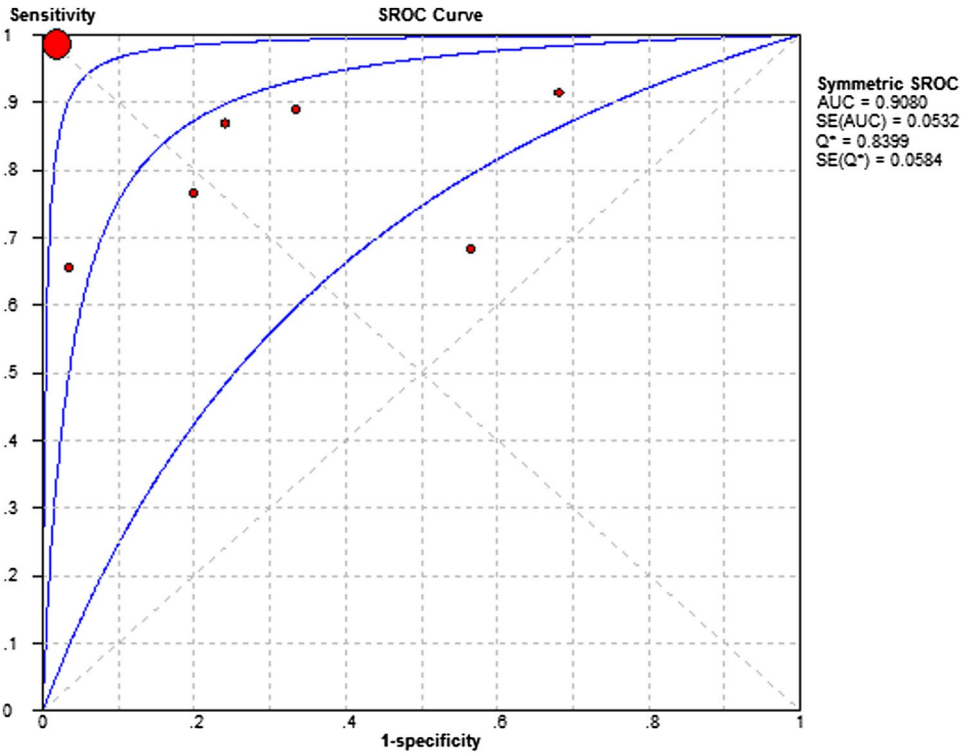

Fig. 13 Summary receiver operating characteristics (SROC) curve of the serum liver function test

diagnostic odds ratio was 43.11 (95% CI 19.98–93.00), I^2 was 53.4%, showing high heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of hepatobiliary scintigraphy is shown in Figs. 14, 15. The sensitivities and specificities of individual studies varied from 84 to 100% and from 35 to 93%, respectively. The hepatobiliary scintigraphy showed pooled sensitivity of 96% (95% CI 94–97%), specificity of 73% (95% CI 70–76%), LR + of 3.26 (95% CI 2.38–4.48) and LR − of 0.09 (95% CI 0.05–0.16). The summary ROC curves of hepatobiliary scintigraphy for the diagnosis of biliary atresia are illustrated in Fig. 16. The summary ROC curve was symmetric, and the AUC was 0.9300, Q was 0.8651. The PPV is 64.5% and the NPV is 97.2%.

Percutaneous liver biopsy

Data on the diagnostic performance of the percutaneous liver biopsy were collected from 11 studies with 646 patients (Table 3). The Spearman correlation coefficient was − 0.109, p value was 0.749, indicating no threshold effect. The diagnostic odds ratio was 348.51 (95% CI 148.74–816.63), I^2 was 0.0%, showing low heterogeneity among the studies.

The forest plot of the sensitivity and specificity of the diagnostic performance of percutaneous liver biopsy is shown in Figs. 17, 18. The sensitivities and specificities of individual studies varied from 90 to 100% and from 84 to 100%, respectively. The percutaneous liver biopsy showed pooled sensitivity of 98% (95% CI 96–99%), specificity of

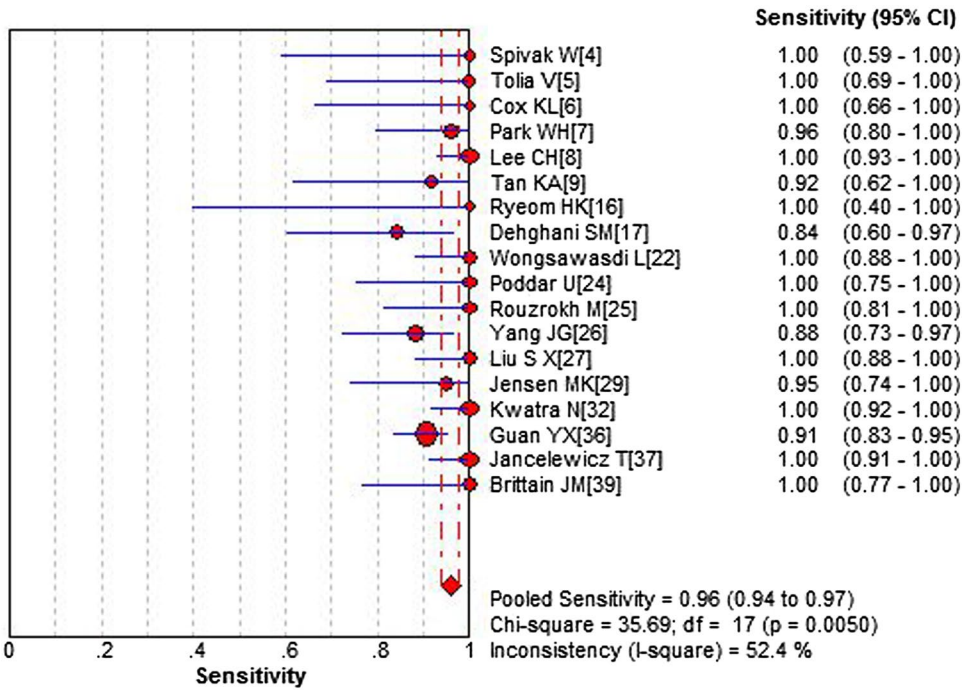

Fig. 14 The forest plots of pooled sensitivity for hepatobiliary scintigraphy

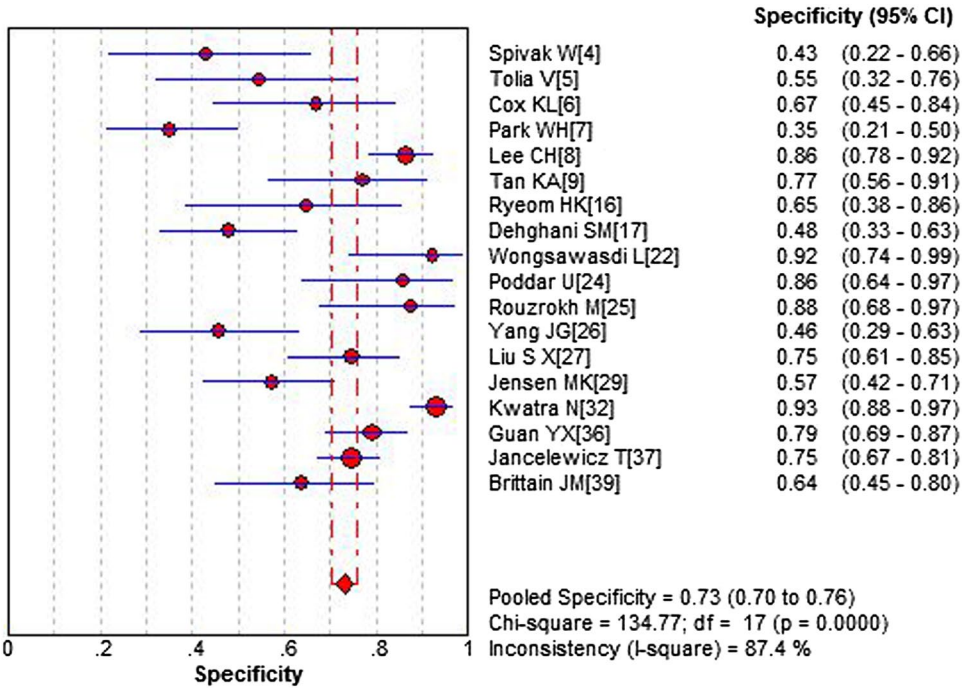

Fig. 15 The forest plots of pooled specificity for hepatobiliary scintigraphy

93% (95% CI 89–95%), LR + of 12.09 (95% CI 8.28–17.63) and LR − of 0.03 (95% CI 0.02–0.06). The summary ROC curves of percutaneous liver biopsy for the diagnosis of biliary atresia are illustrated in Fig. 19. The summary ROC curve was symmetric, and the AUC was 0.9882, Q was 0.9543. The PPV is 93.0% and the NPV is 97.7%.

Subgroup analyses

We performed subgroup analyses for B-US, MRCP and serum liver function test and the results are present in the Table 4. The heterogeneity of articles evaluated MRCP is caused by study design according to the results.

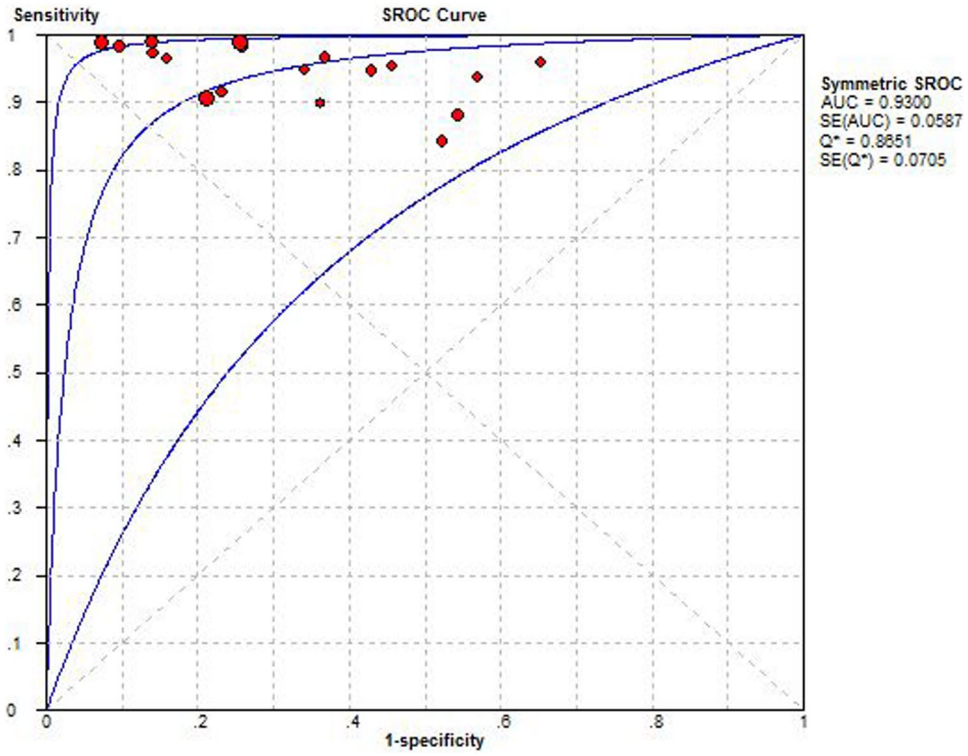

Fig. 16 Summary receiver operating characteristics (SROC) curve of the hepatobiliary scintigraphy

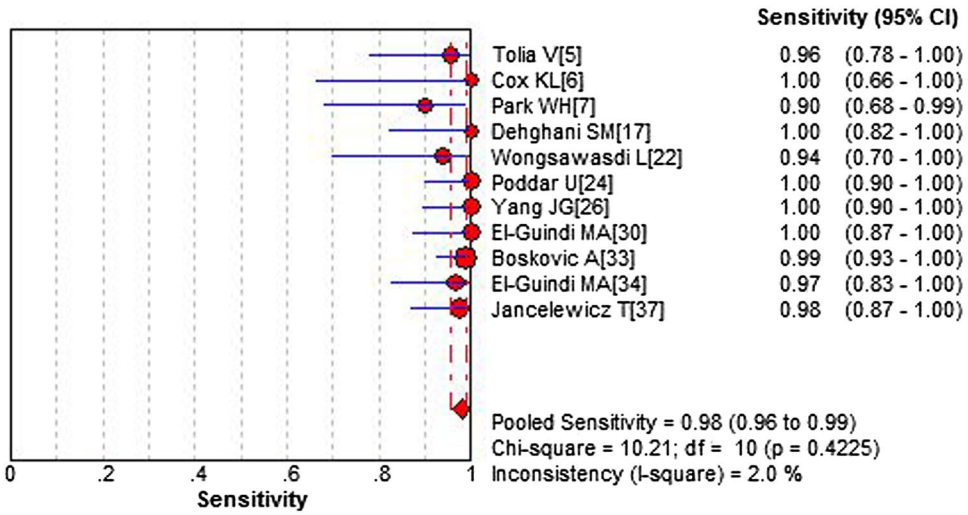

Fig. 17 The forest plots of pooled sensitivity for percutaneous liver biopsy

Publication bias

We constructed Deeks funnel plot to assess publication bias of the studies of B-US, MRCP, acholic stool, serum liver function test, hepatobiliary scintigraphy and percutaneous liver biopsy, there are no bias in all methods (the *p* values are 0.10, 0.97, 0.59, 0.87, 0.11, 0.09, respectively).

Discussion

We know that a good prognosis of Kasai portoenterostomy depends on early diagnosis and early Kasai operation. However, BA and other diseases causing cholestasis jaundice share a great deal of common ground on symptom and laboratory examination. None of early diagnosis method of BA is with accuracy of 100%, which leads to difficulty diagnosing BA within 2 months. Therefore in

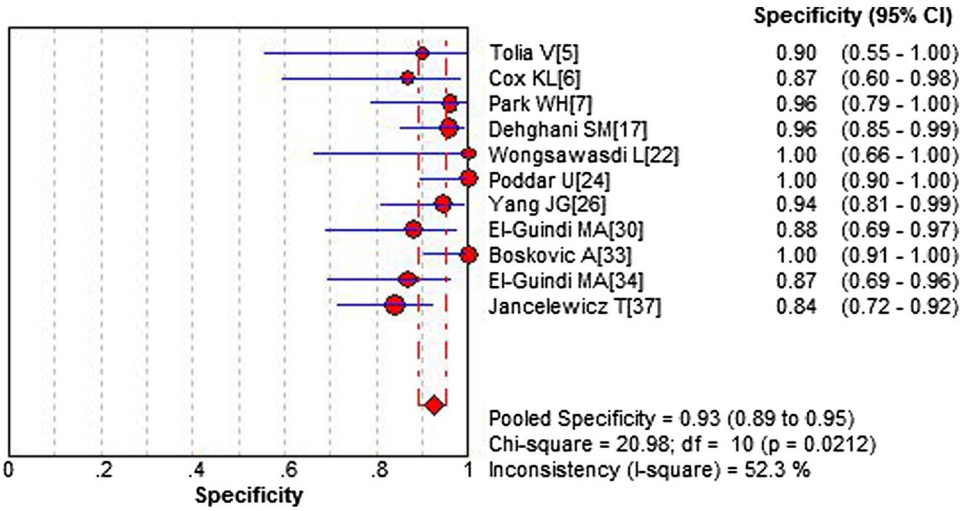

Fig. 18 The forest plots of pooled specificity for percutaneous liver biopsy

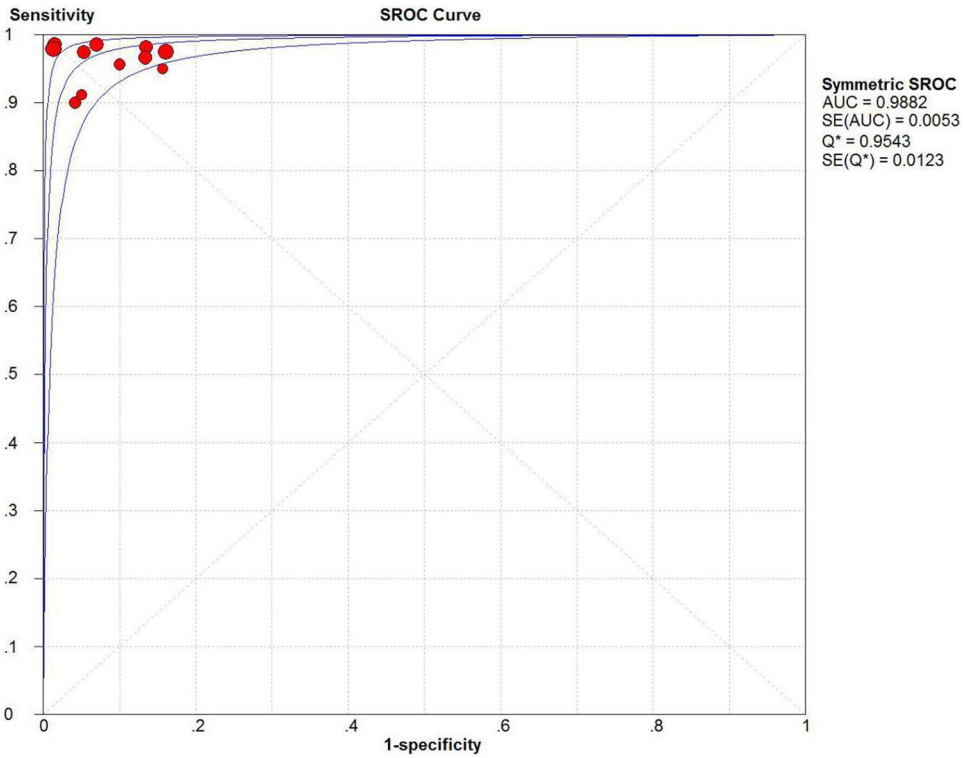

Fig. 19 Summary receiver operating characteristics (SROC) curve of the percutaneous liver biopsy

this meta-analysis, the studies evaluate several diagnosis methods are given precedence.

BA is diagnosed by intraoperative cholangiography with/without intraoperative liver biopsy finally in clinical practice. So even though the preoperative liver biopsy is the most accurate based on AUC, but it is not the method for final diagnosis of BA, just because it is not 100% accurate. In addition, it is invasive, leading to many complications. So in clinical practice, surgeons prefer to use noninvasive method for early diagnosis. Now that none of noninvasive method is with high sensitivity and specificity at the same time, maybe combination of a method with high sensitivity and another method with specificity is a good idea. So combination of MRCP/hepatobiliary scintigraphy (high sensitivity) and B-US/serum liver function (high specificity) is the best according to our data. But hepatobiliary scintigraphy is radioactive. Considering acholic stool is convenient and its sensitivity is acceptable, combination of MRCP/acholic stool and B-US/serum liver function test could be the first choice. But Ağın [42] reported that combination of B-US, acholic stool and GGT for diagnosis BA is with sensitivity of 55.9% and

Table 4 Subgroup analyses of B-US, MRCP and serum liver function test

Covariate	Heterogeneity (I^2)	Threshold effect (P)	Sensitivity (%)	Specificity (%)	DOR	AUC
B-US	Yes (71.4%)	No (0.883)	77	93	46.02	0.9396
Study design						
Prospective	Yes (69.9%)	No (0.997)	81	92	47.56	0.9361
Retrospective	Yes (81.6%)	No (0.800)	63	95	44.71	0.9584
Cases						
≤ 60	Yes (71.4%)	No (0.627)	77	93	46.02	0.9398
> 60	Yes (81.8%)	No (0.869)	78	94	75.15	0.9630
Final diagnosis method						
IC	Yes (76.4%)	No (0.897)	83	89	38.23	0.9378
Surgery and/or histology	No (0%)	No (0.397)	79	97	121.31	0.9552
Did not mention	Yes (73.7%)	No (0.505)	60	95	29.90	0.8950
MRCP	Yes (64.3%)	No (1.000)	96	58	43.49	0.9409
Study design						
Prospective	No (15.5%)	No (0.800)	99	58	83.44	0.9725
Retrospective[a]	–	–	85.3	57.1	–	–
Cases[b]						
≤ 60	No (40.3%)	Yes (0.000)	98	83	102.25	0.9968
Final diagnosis method[b]						
IC	Yes (80.1%)	Yes (0.000)	92	78	66.93	0.9991
Serum liver function test	Yes (82.4%)	No (0.939)	84	97	19.00	0.9080
Study design[b]						
Prospective	Yes (87.5%)	No (0.624)	81	98	29.19	0.9161
Cases						
≤ 60	No (20%)	Yes (0.000)	80	59	12.12	0.8687
> 60	Yes (93.5%)	Yes (0.000)	88	98	38.96	0.9993
Final diagnosis method[b]						
IC	Yes (82.4%)	No (0.200)	84	97	19.00	0.9080

IC diagnosed by intraoperative cholangiography with/without surgery or histology

[a]Only one article in this subgroup

[b]Not all of subgroups have sufficient data to be analyzed, only the data of the analyzable subgroups are listed

specificity of 95%, which is disappointing because of its low sensitivity.

Although sensitivity and specificity are direct index, they could be influenced by cutoff value. We can also use predictive value (PV) to find the best method. PV is an index that use test results to estimate the possibility of sick or health. So we can use a method with high PPV to make a definite diagnosis of BA firstly, and then a method with high NPV should be performed to exclude BA if cannot confirm. According to the criteria, combination with B-US (high PPV) and MRCP/acholic stool/ hepatobiliary scintigraphy (high NPV) is the best. Because of reason as above, maybe combination of B-US and MRCP/acholic stool is the first choice.

Besides, prevalence of disease may influence the performance index of diagnostic method. In term of prevalence, LR + and LR − are more stable than sensitivity, specificity, PPV and NPV. According to the thought, combination with B-US (high LR+) and MRCP/hepatobiliary scintigraphy (low LR −) could be the better choice. Because hepatobiliary scintigraphy is radioactive, so we can use a B-US make a definite diagnosis of BA firstly, and then MRCP is performed to exclude BA if cannot confirm. Sung [43] demonstrated that better diagnostic performance of US with MRCP for discrimination between BA and non-BA was achieved (sensitivity, specificity, accuracy, PPV and NPV are 98, 91, 95, 95, 95 and 98, 83, 92, 91, 95%, evaluated by two observer, respectively).

Certainly, we need more clinical studies to assess the combination strategy for diagnosing BA. If it remains a suspense, hepatobiliary scintigraphy is needed. Liver biopsy should be performed in most infants with undiagnosed cholestasis [44].

Limitations

Our study has several limitations. First, although there are not heterogeneities in some subgroups, other subgroups on the same covariate still show the heterogeneities or cannot be analyzed because of too few articles included. So maybe the heterogeneities are caused by other aspects. In fact, we wanted to add one more covariate of mean age of patients (≤ 60 versus > 60 days), whereas only a part of studies show the result. So we gave up and it was regarded as the greatest limitation of our meta-analysis. Certainly, we thought the difference of diagnosis test equipments maybe also cause the heterogeneities. Second, excluding non-English articles and absence of gray articles could cause bias. Third, we excluded all of articles with incorrect or insufficient data to construct diagnostic 2×2 table. We did not contact authors to obtain the raw data, which also lead to bias probably.

Conclusions

The results of this meta-analysis showed that the accuracy rate of percutaneous liver biopsy is better than all of the noninvasive methods. Take into consideration the advantages and disadvantages of the six methods, combination of multidisciplinary noninvasive diagnosis methods is the first choice for differential diagnosis of BA from other causes of neonatal cholestasis.

Acknowledgements We thank Patrick Chung and Professor Vincent for their advices on this study.

Funding This study was funded by the National Science Foundation of China (Grant Number 81570471) and the Tianjin Health Bureau special grant (Grant Number 14KG129).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Research involving human participants and/or animals This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Sokol RJ, Mack C, Narkewicz MR et al (2003) Pathogenesis and outcome of biliary atresia: current concepts. J Pediatr Gastroenterol Nutr 37: 4–21
2. Wildhaber BE, Majno P, Mayr J et al (2008) Biliary atresia: Swiss national study, 1994–2004. J Pediatr Gastroenterol Nutr 46:299–307
3. Whiting PF, Rutjes AW, Westwood ME et al (2011) Quadas-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536
4. Spivak W, Sarkar S, Winter D et al (1987) Diagnostic utility of hepatobiliary scintigraphy with 99mtc-disida in neonatal cholestasis. J Pediatr 110: 855–861
5. Tolia V, Dubois RS, Kagalwalla A, at el (1986) Comparison of radionuclear scintigraphy and liver biopsy in the evaluation of neonatal cholestasis. J Pediatr Gastroenterol Nutr 5:30–34
6. Cox KL, Stadalnik RC, Mcgahan JP et al (1987) Hepatobiliary scintigraphy with technetium-99m disofenin in the evaluation of neonatal cholestasis. J Pediatr Gastroenterol Nutr 6:885–891
7. Park WH, Choi SO, Lee HJ et al (1997) A new diagnostic approach to biliary atresia with emphasis on the ultrasonographic triangular cord sign: comparison of ultrasonography, hepatobiliary scintigraphy, and liver needle biopsy in the evaluation of infantile cholestasis. J Pediatr Surg 32:1555–1559
8. Lee CH, Wang P, Lee TT et al (2000) The significance of functioning gallbladder visualization on hepatobiliary scintigraphy in infants with persistent jaundice. J Nucl Med 41:1209–1213
9. Tan KA, Phua KB, Ooi BC et al (2000) Making the diagnosis of biliary atresia using the triangular cord sign and gallbladder length. Pediatr Radiol 30:69–73
10. Farrant P, Meire HB, Mieli-Vergani G (2001) Improved diagnosis of extraheptic biliary atresia by high frequency ultrasound of the gall bladder. Br J Radiol 74:952–954
11. Sun Y, Zheng S, Qian Q (2011) Ultrasonographic evaluation in the differential diagnosis of biliary atresia and infantile hepatitis syndrome. Pediatr Surg Int 27:675–679
12. Han SJ, Kim MJ, Han A et al (2002) Magnetic resonance cholangiography for the diagnosis of biliary atresia. J Pediatr Surg 37:599–604
13. Azuma T, Nakamura T, Nakahira M et al (2003) Pre-operative ultrasonographic diagnosis of biliary atresia—with reference to the presence or absence of the extrahepatic bile duct. Pediatr Surg Int 19:475–477
14. Lee HJ, Lee SM, Park WH et al (2003) Objective criteria of triangular cord sign in biliary atresia on US scans. Radiology 229:395–400
15. Visrutaratna P, Wongsawasdi L, Lerttumnongtum P et al (2003) Triangular cord sign and ultrasound features of the gall bladder in infants with biliary atresia. Austr Radio 47:252–256
16. Ryeom HK, Choe BH, Kim JY et al (2005) Biliary atresia: feasibility of mangafodipir trisodium-enhanced MR cholangiography for evaluation. Radiology 235:250–258
17. Dehghani SM, Haghighat M, Imanieh MH et al (2006) Comparison of different diagnostic methods in infants with cholestasis. World J Gastroenterol 12:5893–5896
18. Hu Y, Huang Z, Xia L (2006) MR cholangiography and dynamic examination of duodenal fluid in the differential diagnosis between extrahepatic biliary atresia and infantile hepatitis syndrome. J Huazhong Univ Sci Technol Med Sci 26:725–727
19. Humphrey T, Stringer M (2007) Biliary atresia: US diagnosis. Radiology 244:845–851
20. Kim WS, Cheon JE, Youn BJ et al (2007) Hepatic arterial diameter measured with US: adjunct for US diagnosis of biliary atresia. Radiology 245:549–555
21. Takamizawa S, Zaima A, Muraji et al (2007) Can biliary atresia be diagnosed by ultrasonography alone? J Pediatr Surg 42:2093–2096
22. Wongsawasdi L, Ukarapol N, Visrutaratna P et al (2008) Diagnostic evaluation of infantile cholestasis. J Med Assoc Thai 91:345–349
23. Lee MS, Kim MJ, Lee MJ et al (2009) Biliary atresia: color doppler US findings in neonates and infants. Radiology 252:282–289

24. Poddar U, Thapa BR, Das A et al (2009) Neonatal cholestasis: differentiation of biliary atresia from neonatal hepatitis in a developing country. Acta Paediatr 98:1260–1264
25. Rouzrokh M, Sobhiyeh MR, Heibatollahi M (2009) The sensitvity, specificity, positive and negative predictive values of stool color test, triangular cord sign and hepatobiliary scintigraphy in diagnosis of infantile biliary atresia. Iran Red Crescent Med J 11:425–430
26. Yang JG, Ma DQ, Peng Y et al (2009) Comparison of different diagnostic methods for differentiating biliary atresia from idiopathic neonatal hepatitis. Clin Imaging 33:439–446
27. Liu SX, Huang ZH (2010) The value of radionuclide hepatobiliary scintigraphy in combination with determination of bilirubin from duodenal drainage in differential diagnosis of infantile persistent jaundice. Front Med China 4:342–345
28. Aziz S, Wild Y, Rosenthal P et al (2011) Pseudo gallbladder sign in biliary atresia—an imaging pitfall. Pediatr Radiol 41:620–626
29. Jensen MK, Biank VF, Moe DC et al (2012) HIDA, percutaneous transhepatic cholecysto-cholangiography and liver biopsy in infants with persistent jaundice: can a combination of ptcc and liver biopsy reduce unnecessary laparotomy? Pediatr Radiol 42:32–39
30. Elguindi MA, Sira MM, Konsowa HA et al (2013) Value of hepatic subcapsular flow by color doppler ultrasonography in the diagnosis of biliary atresia. J Gastroenterol Hepatol 28:867–872
31. Jiang LP, Chen YC, Ding L et al (2013) The diagnostic value of high-frequency ultrasonography in biliary atresia. Hepatobiliary Pancreat Dis Int 12:415–422
32. Kwatra N, Shalabyrana E, Narayanan S et al (2013) Phenobarbital-enhanced hepatobiliary scintigraphy in the diagnosis of biliary atresia: two decades of experience at a tertiary center. Pediatr Radiol 43:1365–1375
33. Boskovic A, Kitic I, Prokic D et al (2014) Predictive value of hepatic ultrasound, liver biopsy, and duodenal tube test in the diagnosis of extrahepatic biliary atresia in serbian infants. Turk J Gastroenterol 25:170–174
34. El-Guindi MA, Sira MM, Sira AM et al (2014) Design and validation of a diagnostic score for biliary atresia. J Hepatol 61:116–123
35. Liu B, Cai J, Xu Y et al (2014) Three-dimensional magnetic resonance cholangiopancreatography for the diagnosis of biliary atresia in infants and neonates. Plos One 9:e88268
36. Guan YX, Chen Q, Wan SH et al (2015) Effect of different time phases of radionuclide hepatobiliary scintigraphy on the differential diagnosis of congenital biliary atresia. Genet Mol Res 14:3862–3868
37. Jancelewicz T, Barmherzig R, Chung CT et al (2015) A screening algorithm for the efficient exclusion of biliary atresia in infants with cholestatic jaundice. J Pediatr Surg 50:363–370
38. Lee SM, Cheon JE, Choi YH et al (2015) Ultrasonographic diagnosis of biliary atresia based on a decision-making tree model. Korean J Radiol 16:1364–1372
39. Brittain JM, Kvist N, Johansen LS et al (2016) Hepatobiliary scintigraphy for early diagnosis of biliary atresia. Dan Med J 63:A5253
40. Mandana R, Lida S, Shoa HJ et al (2016) Diagnostic value of anti-smooth muscle antibodies and liver enzymes in differentiation of extrahepatic biliary atresia and idiopathic neonatal hepatitis. Afr J Paediatr Surg 13:63–68
41. Shen Z, Zheng S, Dong R et al (2016) Saturation of stool color in HSV color model is a promising objective parameter for screening biliary atresia. J Pediatr Surg 51:2091–2094
42. Ağın M Tümgör G, Alkan M et al (2016) Clues to the diagnosis of biliary atresia in neonatal cholestasis. Turk J Gastroenterol 27:37–41
43. Sung S, Jeon TY, Yoo SY et al (2016) Incremental value of MR cholangiopancreatography in diagnosis of biliary atresia. Plos One 11:e0158132
44. Moyer V, Freese DK, Whitington PF et al (2004) Guideline for the evaluation of cholestatic jaundice in infants: recommendations of the North American Society for pediatric gastroenterology, hepatology and nutrition. J Pediatr Gastroenterol Nutr 39:115–128