

The Diagnostic Value of Pyruvate Kinase Isoenzyme Type M2 for Biliary Tract Carcinoma. A Systematic Review and Meta-Analysis

Wen Qin Wang, Wei Liu, Ning Zhang, Peng Hui Wang, Jian Zhong Cao, Jie Min Chen, Bing Lu Li, Xiao Dong He

Department of General
Surgery,
Peking Union Medical College
Hospital,
Chinese Academy of Medical
Sciences and Peking Union
Medical College,
Beijing,
P.R. China

Address for correspondence:

Xiao Dong He
Department of General
Surgery of Peking Union
Medical College Hospital
(East), Chinese Academy of
Medical Sciences and Peking
Union Medical College, No.1
Shuaifuyuan Wangfujing
Dongcheng District,
Beijing, 100730
P.R. China
hexiaodpumch@163.com

Received: 30.10.2017
Accepted: 05.01.2018

ABSTRACT

Background & Aims: Growing evidence has shown that M2-PK is involved in cancer diagnosis and prognosis. The overall diagnostic accuracy of the pyruvate kinase isoenzyme type M2 (M2-PK) in biliary tract carcinoma (BTC) remains controversial. We performed a meta-analysis to evaluate the diagnostic value of M2-PK for BTC. **Methods:** The online PubMed, Cochrane, Web of Science, and Embase databases were searched for eligible studies published until August 8th, 2017. The Quality Assessment for Diagnostic Accuracy Studies 2 (QUADAS-2) was used to evaluate study quality. All statistical analyses were conducted with Stata 12.0. **Results:** We included 7 studies from 5 articles with 410 patients with BTC and 438 controls. The pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and AUC for M2-PK in the diagnosis of BTC were 0.79 (95%CI 0.70-0.86), 0.81 (95%CI 0.71-0.88), 4.1 (95%CI 2.5-6.8), 0.26 (95%CI 0.16-0.41), 17.159 (95%CI 5.468-54.071), and 0.87 (95%CI 0.83-0.89), respectively. The same indicators assessed for CA19-9 were as follows: 0.70 (95%CI 0.62-0.77), 0.71 (95%CI 0.45-0.87), 2.38 (95%CI 1.2-4.73), 0.43 (95%CI 0.34-0.53), 6.28 (95%CI 2.4-16.44) and 0.73 (95%CI 0.69-0.77), respectively. Additionally, the diagnostic value of M2-PK varied based on characteristics of golden methods and different cut-off values.

Conclusions: This meta-analysis showed that M2-PK had a better diagnostic accuracy for BTC compared with CA19-9, with moderate diagnostic performance. However, prospective studies are required to confirm its diagnostic value.

Key words: Biliary tract cancer – Kinase Isoenzyme Type M2 – diagnosis – accuracy – meta-analysis.

Abbreviations: AC: ampullary carcinoma; BTC: biliary tract cancer; CC: cholangiocarcinoma; CA19-9: carbohydrate antigen 19-9; DOR: diagnostic odds ratio; GBC: gallbladder cancer; M2-PK: pyruvate kinase isoenzyme type 2; NLR: negative likelihood ratio; PLR: positive likelihood ratio; PK: pyruvate kinase; QUADAS-2: The Quality Assessment for Studies of Diagnostic Accuracy 2.

INTRODUCTION

Biliary tract cancers (BTC) include cholangiocarcinoma (CC), gallbladder cancer (GBC) and ampullary carcinoma (AC) [1]. Though the incidence of BTC is low (2-3 persons/100,000 per year in Western countries, 4-6 persons/100,000 per year in Asian countries) [2, 3], its prognosis remains dismal with a poor 5-year survival rate of only 3.2% [4], and a median overall survival of 5-15 months [5] in spite of medical intervention.

Exceptionally, the 5-year survival rates of AC patients were 43.3–61.0% [6-8] for early symptoms and curative surgery stages [9]. Despite a slight downward trend worldwide over the past few decades according to National Cancer Institute (SEER Program; <http://seer.cancer.gov/>) and global cancer statics [10], the prevalence of GBC is constantly rising and is a substantial cause of mortality in some regions, for example, in Shanghai, China [10, 11].

These malignancies have a poor prognosis mostly because they are advanced when diagnosed, there is no effective adjuvant therapy and they have a progressive biological behavior. Curative surgery is almost the single effective treatment until now; however, 35% of the patients miss the timing for surgery at diagnosis and only 50% of those who underwent surgery achieved curative or margin-free resection (R0) [13]. The relief of biliary obstruction and palliative

chemotherapy are the mainstays of therapy. Nevertheless, only 15-40% of patients with BTC respond to chemotherapy [5]. The patients with BTC show a high rate of recurrence and distant metastasis, about 60% to 80% [14].

Thus, early diagnosis is one of the important steps to improve the medical interventions and survival of BTC [15]. Until now, imaging methods, such as ultrasound or computed tomography (CT), were the main diagnostic modality for BTC, but with little promise for early diagnosis in cases with atypical symptoms or without a discernible mass. Meanwhile, no individual and single specific diagnostic tumor biomarkers exist for imaging conflicting results [16, 17]. The National Comprehensive Cancer Network (NCCN) guidelines only take carbohydrate antigen 19-9 (CA19-9) as a baseline of continuous surveillance and do not recommend other tumor markers as diagnostic indicators [18]. CA19-9 has a wide variation in sensitivity (50–90%) and specificity (54–98%) [17, 19-23], is falsely elevated in benign biliary diseases, is relieved in biliary obstruction and sepsis, and undetectable in Lewis antigen negative population [17]. Thus, the early diagnosis of BTC remains challenging [24, 25]. In addition, it is often difficult to differentiate between malignant and benign biliary strictures, such as primary sclerosing cholangitis (PSC) [9, 26].

Consequently, there is a strong need for better reliable tumor biomarkers in addition to the current diagnostic work up for BTC. One such potential biomarker is the pyruvate kinase isoenzyme type M2 (M2-PK) [26]. Pyruvate kinase (PK) in mammals has four isoforms (L, R, M1, and M2). The M2-PK is found primarily in embryonic tissues and specifically in tumor cells, participating in the glycolysis, as the key enzyme to product lactate from pyruvate, known as “aerobic glycolysis” [27, 28]. Therefore, M2-PK specifically assists survival and development of cellular growth of tumor cells [29]. Its role as a diagnostic marker has been explored in many different cancers and diseases such as Barret’s esophageal cancer [30, 31], glioma [32], gastric cancer [29, 33], colon cancer [34, 35], non-small cell lung cancer [36, 37], exocrine pancreatic cancer [38], ovarian cancer [39] and advanced renal cancer [40], with diagnostic sensitivity levels ranging from 43 to 95%.

Several studies have addressed the potential value of M2-PK as a repeatable and non-invasive “liquid biopsy” for BTC. Its

diagnostic role remains equivocal. Moreover, these studies have not been systematically reviewed. Hence, this meta-analysis aimed to quantitatively evaluate the diagnostic efficiency of M2-PK to provide a better understanding of its diagnostic value in BTC.

METHODS

Data sources and search

We performed this systematic review and meta-analysis in accordance with the PRISMA 2009 guidelines. On August 8th, 2017 we systematically searched online PubMed, Cochrane, Web of Science, and Embase databases. Search terms used in our study included biliary tract malignancy, cholangiocarcinoma, gallbladder cancer, ampullary carcinoma, M2 Pyruvate Kinase, M2-PK, sensitivity, specificity and accuracy. We used free terms, MeSH terms as well as abbreviations and synonyms as keywords. We also manually searched the references of eligible articles and relevant reviews to find out other potential articles of interest.

The full electronic search strategy for Pubmed was listed in Table I.

Selection of the studies

We included English written articles; studies in humans; research articles; studies about diagnostic performance of M2-PK in BTC; studies with BTC diagnosed based on pathological examination; only published literature which provided sufficient data to construct the diagnostic 2-by-2 tables.

The exclusion criteria were: duplicate articles; letters, reviews, meta-analyses, editorials, case reports, and studies without sufficient diagnostic data. Two researchers (WWQ and WPH) independently assessed any related articles carefully. Discrepancies were resolved by discussion.

Data extraction and quality assessment

We reviewed the included studies and extracted the details of studies (first author, published date, and country), the clinical characteristics of subjects (number of participants, number of males/females, and pathology of subjects), details about the detection method (specimen type, cut-off value), and

Table I. Full electronic search strategy for Pubmed

Number	Search items	Item count
#1	Pyruvate Kinase[MeSH Terms] ti.ab	6255
#2	(((M-Type Pyruvate Kinase) OR M Type Pyruvate Kinase) OR Pyruvate Kinase, M-Type) OR M2-Type Pyruvate Kinase) OR M2 Type Pyruvate Kinase) OR Pyruvate Kinase, M2-Type ti.ab	11550
#3	Biliary Tract Neoplasms[MeSH Terms] ti.ab	25079
#4	(((((((Biliary Tract Neoplasm) OR Neoplasm, Biliary Tract) OR Neoplasms, Biliary Tract) OR Biliary Tract Cancer) OR Biliary Tract Cancers) OR Cancer, Biliary Tract) OR Cancers, Biliary Tract) OR Cancer of the Biliary Tract) OR Cancer of Biliary Tract ti.ab	28257
#5	#1 OR#2	11550
#6	#3 OR#4	37615
#7	#5 AND #6	5

ti: title; ab: abstraction (only search strategy of Pyruvate Kinase and biliary tract neoplasm are shown in the table; the strategy was same with other terms. MeSH term and free term of each key word was combined by “OR”. Then M2-PK and BTC and diagnosis was combined using “AND”)

diagnostic performance (sensitivity, specificity, and data of 2 by 2 tables). If the study contained the training and validating cohorts, each cohort was considered as an individual study. Study quality was evaluated with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). The QUADAS-2 form is composed of four domains: (1) patient selection, (2) index test, (3) reference standard and (4) flow and timing. For each domain, the risk of bias and concerns about applicability (the latter not applying to the flow and timing domains) were analyzed and rated as low, high and unclear risk. This is a validated tool for the assessment of methodological quality and applicability of diagnostic accuracy studies and to investigate potential sources of heterogeneity [41]. Two reviewers (CJZ and CJM) performed the data extraction and quality assessment independently. Any disagreements were resolved by consensus.

Statistical analysis

Stata 12 was used for all analyses. The sensitivity, specificity, diagnostic odds ratio (DOR), positive likelihood ratio (PLR), and negative diagnostic likelihood ratio (NLR) were graphically displayed after bivariate analysis. The heterogeneity was assessed using the bivariate model via the Spearman correlation analysis method, Cochran-Q, and inconsistency index (I²) tests, respectively. A p value (< 0.05) and I² value (≥50%) indicated significant heterogeneity. When significant heterogeneity existed across studies, a random-effect model was conducted. We conducted covariates in addition to the bivariate model to examine whether sensitivity and specificity were different for the following study characteristics: sample size, specimen type, golden standard, pathology of subject, test method procedure.

We performed a graphical evaluation of reporting bias using a funnel plot specifically designed for reviews of diagnostic test accuracy, representing the log (DOR) of each study against the inverse of the square root of the effective sample size: $ESS=4n_1n_2/(n_1+n_2)^2$, where n1 and n2 are the number of subjects, respectively. We then performed the test on asymmetry proposed by Deeks et al. [42] on the slope of a ESS-weighted regression of log(DOR) against $1/\sqrt{ESS}$, which is known as Deeks publication bias plot [43]. A p value < 0.05 indicates statistical significance.

RESULTS

Literature selection

The number of searched articles relating to M2-PK diagnosis and BTC were 37,615 and 11,550 from Pubmed, 1,364,394 and 43,397 from EMBASE, 237 and 618,679 from Web of Science, 0 and 59,304 from Cochrane library.

Twenty four articles concerning the M2-PK and BTC were primarily identified from four databases (Fig. 1). After 13 duplicated articles were excluded, we screened the titles and abstracts of 11 researches. Six articles were selected for full-text review after excluding 6 articles for the following reasons: letters, reviews, or irrelevant research topic, without valuable data or non-full test. Finally, 5 eligible articles [24-26, 44, 45] containing 7 studies were identified and included in the meta-analysis.

Seven studies from 5 articles including 410 BTC and 438 matched controls from 4 countries were included. The characteristics of each study are presented in Table II. The size of case and control groups ranged from 6 to 115 and 11 to 165,

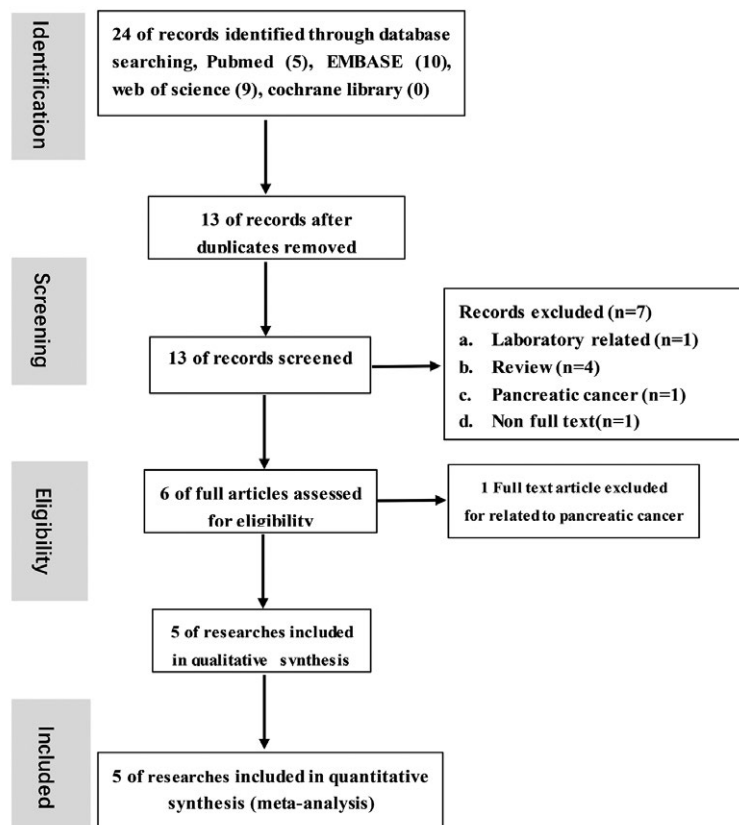


Fig. 1. Flow diagram of the study selection process

respectively. One study was conducted in China, 2 in the USA, 3 in UK, and 1 in Mexico; the patients with BTC were confirmed by pathological and cytological examination; among these studies, M2-PK levels were detected with the ELISA method. All these studies included subjects with benign biliary diseases as controls, and two studies also included healthy individuals as controls.

Methodological quality of the included studies

Overall, the quality of the included study cohorts was good though with slight methodological flaws (QUADAS-2 summary plot is presented in Fig. 2). All studies adopted different prespecified thresholds for differential diagnosis. All studies detailed the procedure adopted to measure M2-PK. Regarding applicability concerns for the index test, all studies met the predefined criteria. For the flow and timing domain, there was an appropriate interval between the index tests and reference standard, and all the measurements of M2-PK were done ahead of the pathology.

Concerning the domain of patient selection bias, three studies did not explicitly report that the patients were consecutive [24-26]. Similarly, one study [26] did not explicitly report whether the measurement of the M2-PK value was interpreted without the knowledge of the results of the surgical-pathological evaluation. In one study, the imaging results were taken as a diagnostic standard [24], hence a different diagnostic method could have introduced bias. Furthermore, four studies did not explicitly state if the pathological evaluation of the specimen was based on fine needle aspiration biopsy / cytology or on histology, as they performed both procedures [25, 26, 45].

According to the criteria, all the 7 studies from 5 articles achieved QUADAS-2 standard (Table II), indicating moderate quality. The details of the quality assessment of each article are presented in Supplementary Table I.

Diagnostic accuracy

To estimate the diagnostic accuracy of M2-PK in BTC, the pooled sensitivity, specificity, PLR, NLR, and DOR were determined. The values were as follows: 0.79 (95%CI 0.70-0.86), 0.81 (95%CI 0.71-0.88), 4.1 (95%CI 2.5-6.8), 0.26 (95%CI 0.16-0.41), and 17.159 (95%CI 5.468-54.071), respectively (Fig. 3). The summary receiver operator characteristic curve (SROC) was also plotted (Fig. 4). M2-PK achieved an area-under-curve (AUC) of 0.87 (95%CI 0.83-0.89), suggesting a moderate accuracy in BTC diagnosis. The same indicators determined for CA19-9 were: 0.70 (95%CI 0.62-0.77), 0.71 (95%CI 0.45-0.87), 2.38 (95%CI 1.2-4.73), 0.43 (95%CI 0.34-0.53), 6.28 (95%CI 2.4-16.44), and AUC 0.73 (95%CI 0.69-0.77).

Threshold analysis

Further, threshold effects were assessed in our study. The Spearman correlation coefficient was 0.98 (p=0.96), suggesting no obvious heterogeneity from threshold effect. Then heterogeneity from non-threshold was evaluated by I². There was a substantial heterogeneity in the pooled sensitivity (I²=80.75, p<0.001) and pooled specificity (I²= 84.72%, p <0.01), thus, a random-effect model was conducted.

Non-threshold analysis

The forest plot in Fig. 5 depicts the sensitivity and specificity of the M2-PK test stratified by the sample size, specimen, same method, index test, subjects, and cut-off value.

Subgroup analysis and meta-regression showed that the studies which used the same gold standard tended to have significantly lower specificity (0.72) than the other ones (0.88) (p<0.01). However, the difference between sensitivities was not significant (p=0.19, 0.79 vs. 0.77). The joint model showed a

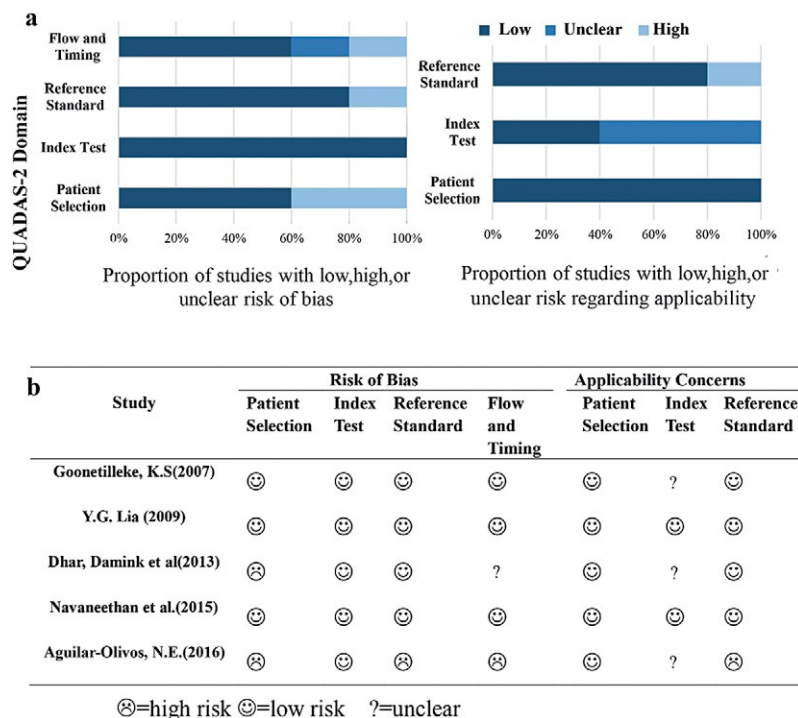


Fig. 2. Methodological evaluation according to QUADAS-2 of the included studies

Table II. The main characteristics of the 7 studies included in the meta-analysis

Author (Year) [Ref]	ID	Country	Study type	Blind method	Study population	Gold standard	Disease	Specimen	Cases number (malignant/benign)	Cut-off (U/mL)	Se %	Sp %
Goonetilleke KS et al. (2007) [44]	1	United Kingdom	prospective	-	-	histology	Peri-ampullary cancer	plasma	115 (76/67)	27.2	66	58
Li YG et al. (2009) [45]	2	China	prospective	-	continuous	histology/FNA†/cytology	Cholangiocarcinoma	plasma	280 (115/165)	18	84	90
Dhar et al. (2013) [26]	3	United Kingdom	-	-	-	histology/cytology	Cholangiocarcinoma/gallbladder cancer	plasma	167 (88/79)	31.7	90.3	84.3
Dhar D et al. (2013) [26]	3	United Kingdom	-	-	-	histology/cytology	Cholangiocarcinoma/gallbladder cancer	Bile	167 (88/79)	24.4	71.0	69.9
Navaneethan U et al. (2015) [25]	4	USA	prospective	-	-	histologically/cytology/FNA	Cholangiocarcinoma/Pancreatic cancer	Bile (discovery cohort)	34 (17/17)	109.1	52.9	94.1
Navaneethan U et al. (2015) [25]	4	USA	prospective	-	-		Cholangiocarcinoma/Pancreatic cancer	Bile (validation cohort)	40 (20/20)	109.1	95	80
Aguilar-Olivos NE et al. (2016) [24]	5	Mexico	prospective	-	-	clinical, radiological/histological	-	Bile	20 (6/11)	0.0216	66.70	90.90

† Fine needle aspiration; - not available; Se: sensitivity; Sp: specificity

p=0.03. Besides, the subgroup categorized into a higher cutoff (≥ 25 U/ml) showed lower sensitivity (0.75) than low M2-PK subgroup (0.81) with statistical significance (p=0.03), while the specificity was similar between the two groups (0.79 vs. 0.82, p=0.19). Sensitivity and specificity did not change significantly, regardless of other covariates.

Publication bias

Deeks funnel plot asymmetry test was performed to evaluate publication bias (Fig. 6). The slope coefficient did not reveal obvious evidence of asymmetry (p=0.94). Potential publication bias among studies cannot be excluded due to their inherent under-power based on the small number of included articles.

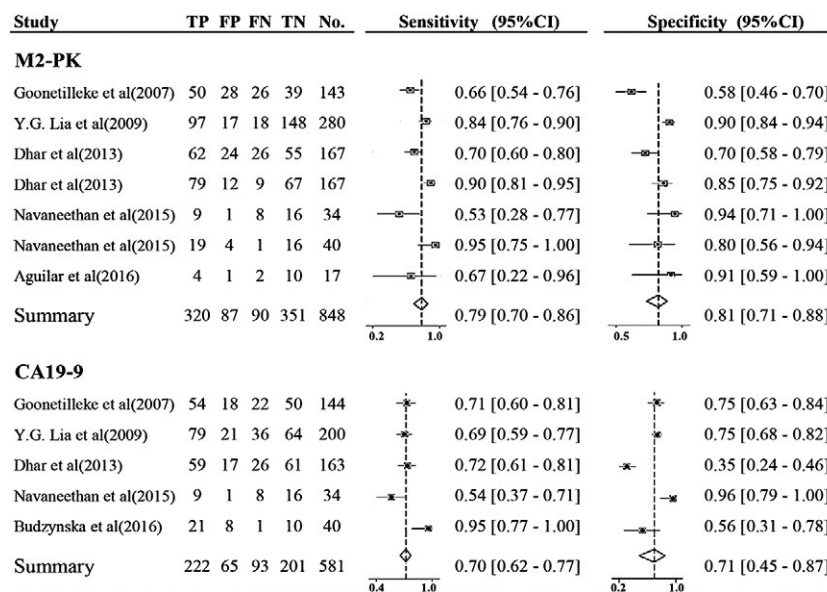


Fig. 3. Pooled results of the meta-analysis. Forest plot shows sensitivity and specificity with corresponding 95% CIs. Studies are ordered chronologically. CI: confidence intervals; TP: true positive; FN: false negative; FP: false positive; TN: true negative; Sn: sensitivity; Sp: specificity.

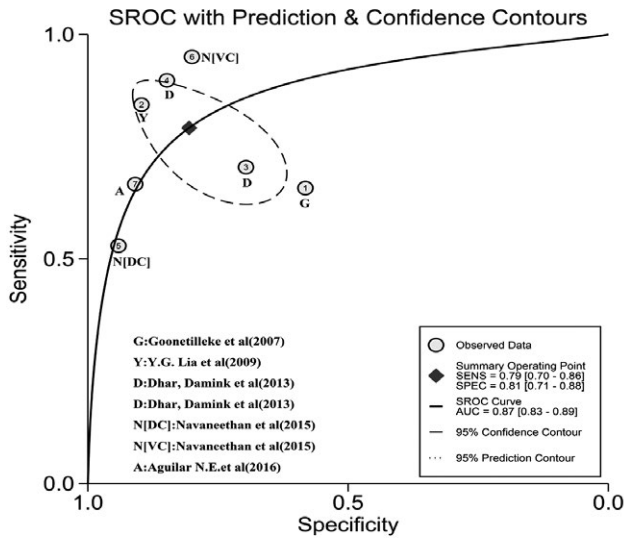


Fig. 4. The summary receiver operator characteristic curve of M2-PK. M2-PK achieved an AUC of 0.87 indicating a moderate accuracy in BTC diagnosis.

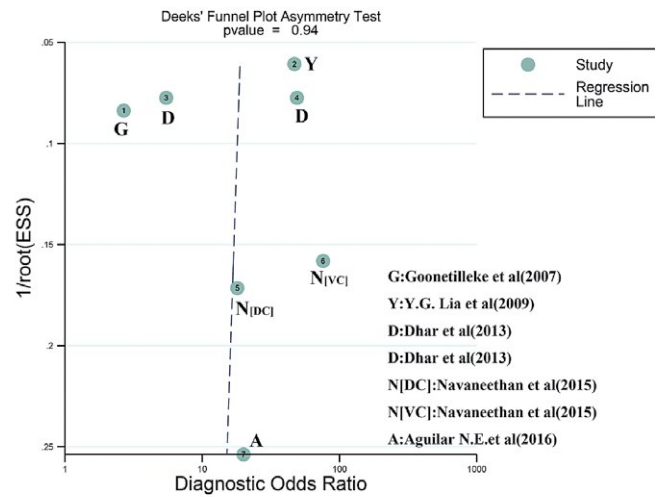


Fig. 6. Deeks funnel plot of publication bias. The dotted line represents the regression line of log (DOR) against root 1/ESS. ESS (effective sample size). The Deeks test for asymmetry is not significant. DC: discovery cohort; VC: validation cohort.

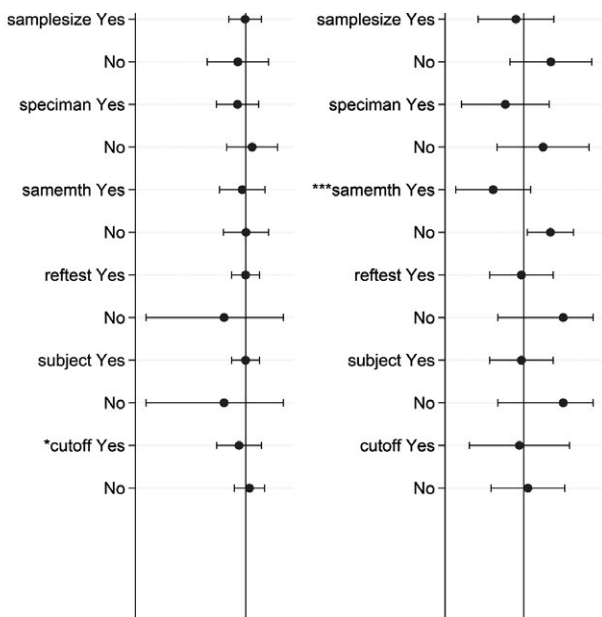


Fig. 5. Subgroup analysis and meta-regression. Sample size (Yes: larger than 100 cases; No: less than 100 cases); Specimen (Yes: plasma; No: bile); Same method (Yes: the same gold standard; No: more than one gold standard); Ref test (Yes: detailed description of the test method; No: not detailed description of the test method); Subject (Yes: detailed description of the subject; No: not detailed description of the subject); Cut-off (Yes: above 25U/ml; No: below 25U/ml).

DISCUSSION

We performed the first meta-analysis on the diagnostic accuracy of M2-PK for detection of BTC based on unique studies. Generally, a correct diagnosis could often be made based on M2-PK assays, with a pooled sensitivity and specificity of 0.79 and 0.81, respectively. The AUC of SROC curves was well above common standards for diagnostic procedures at 0.87 (> 0.8). The compared sensitivity and specificity of CA19-9 in the same study population of 4 studies were 0.70

and 0.71, respectively, and the AUC for CA19-9 was 0.73. In other independent studies, CA19-9 exhibited varied diagnostic performance with a sensitivity of 40–90%, specificity 50–98% [17,19-23], and positive predictive value (PPV) of 16-40% depending on cut-off values [20, 22, 23]. Regarding the other traditional serum markers, the estimated sensitivity and specificity of carcinoembryonic antigen (CEA) were 20%-68% and 82-100%, respectively [15, 46], while for cancer antigen 125 (CA125) sensitivity and specificity were 31.25%, and 96% [47], respectively. The accuracy of the M2-PK assay appears to be modestly stronger than either of these traditional markers.

In grouped analyses stratified by plasma and serum, the sensitivity and specificity of the plasma M2-PK were 0.75 (95% CI,0.63-0.86) and 0.75 (95% CI,0.63-0.88), while those of bile M2-PK were 0.83 (95% CI,0.69-0.97) and 0.86 (95% CI,0.73-0.99). Though bile M2-PK seemed to have better sensitivity and specificity, no statistical significance was reached (p=0.12 and 0.09, respectively) (Supplementary Table II).

The DOR was further calculated to evaluate diagnostic effectiveness [48]. The value of DOR >10 represents a good discriminatory test performance. In this meta-analysis, the DOR for M2-PK assays was 17.195, while the compared DOR for CA19-9 was 6.28 indicating a better ability to correctly discriminate BTC cases from benign lesions by using M2-PK assays. But the fact that M2-PK may rise in patients with benign diseases should not be ignored. This limits the potential use of the M2-PK assay as a diagnostic marker. For example, M2-PK was also elevated in patients with benign pancreatic diseases, inflammatory lung diseases, and even in normal tissues (i.e. normal fresh colonic epithelium) [36, 49, 50].

M2-PK showed a better diagnostic performance than the traditional markers used in other gastrointestinal (GI) cancers, except pancreatic cancer. Its diagnostic specificity was 89% and sensitivity ranged from 48% to 73% [51, 52]. Especially in colon cancer, there are several studies and a meta-analysis about stool

M2-PK providing a high sensitivity (0.78) and specificity (0.77) [34, 35]. However, in pancreatic cancer, M2-PK had an overall specificity of 60% and a sensitivity of 95%, similar to those of CA19-9 [38]. Regarding the nonGI cancers, the sensitivity of M2-PK for detecting ovarian cancer was 70% with a specificity of 65% at a cut-off value of 22 U/ml [39]. In lung cancer, a meta-analysis revealed that M2-PK can be a potential biomarker for diagnosis of non-small cell lung cancer with pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR) of 0.69 (0.65-0.72), 0.92 (0.89-0.94), 7.84 (5.92-10.38), and 0.36 (0.32-0.40) [37]. In addition to the previous non-metastatic cancers, plasma M2-PK concentrations might be a potential biomarker of advanced renal cell cancer with a sensitivity of 66.7% and specificity of 95% [40].

Positive LR and NLR are also important indicators to present overall diagnostic accuracy [53]. Likelihood ratios of a PLR > 10 and an NLR < 0.1 present high accuracy. A PLR value of 4.1 for the M2-PK assays suggests that patients with BTC have more than a 4-fold higher chance of being M2-PK assay-positive compared with benign cases. The NLR was 0.26, implying that one fourth of cases with negative M2-PK test may have BTC. These data suggest that the results of M2-PK for BTC diagnosis should cautiously be interpreted and not used alone. PLR and NLR for CA19-9 were 2.38 (95%CI 1.2-4.73), and 0.43 (95%CI 0.34-0.53) respectively, indicating a worse diagnostic performance than M2-PK.

We found some heterogeneity in the sensitivity and specificity of M2-PK tests among the studies. Considering that positive results have more chance to be published, the Deek's funnel plot asymmetry, which tests the effect of publication bias was conducted, but it did not indicate publication bias. However, we found a higher sensitivity and specificity in the non-same gold standard subgroup rather than in the same method group, suggesting a major source of heterogeneity. Besides, the subgroup categorized into high M2-PK (≥ 25 U/ml) showed lower sensitivity than the low M2-PK subgroup with statistical significance ($p=0.03$), while the specificity was similar, without significant difference ($p=0.19$). Though joint p was 0.70, difference among the cut-off values was another potential source of heterogeneity. However, the other study characteristics including case number, specimen, the measurement of the index test, and the subject of these studies did not contribute to the heterogeneity. The heterogeneity could have been derived from other different methodological characters not included in meta-regression analysis due to incomplete information provided by the primary studies, such as jaundice and TNM staging of patients enrolled. Only one study [44] mentioned that the M2-PK test is unaffected by the presence of jaundice in control patients. In addition, two studies included less than 20 cancer patients [24, 25], which may also have contributed to the poor robustness. Overall, the insignificant publication bias suggested that the heterogeneity of the included studies had been mostly dependent on the objective quality of the research.

Currently, the use of M2-PK in the BTC differential diagnosis is a matter of debate. Hence, we conducted a comprehensive meta-analysis to evaluate the diagnostic accuracy of M2-PK assays, and we concluded that M2-PK has a higher accuracy than the routinely used biomarkers CA19-9 and CA125. The mean sensitivity and specificity (0.76 and

0.80) of M2-PK were even higher than those of ERCP (0.74 and 0.70) [54], the current benchmark of BTC pretreatment diagnosis [4]. Though the LR showed imperfect robustness as a lonely marker [39-41, 51-53], this easy available M2-PK test is highly recommended alongside conventional cytological and histological examinations for BTC by Navaneethan et al. They found that a combination of markers - elevated M2-PK (>109.1 U/l) or CA 19-9 (>33 U/ml) or positive biliary brushing increased the sensitivity (88.2%) and the specificity (88.2%) in diagnosing malignant biliary strictures, with an AUC of 0.89 [25]. In another study, combined M2-PK with CA19-9 did not improve the diagnostic sensitivity and specificity over the M2-PK alone [26].

Our meta-analysis had some limitations. Firstly, it was impossible to determine all sources of heterogeneity. The potential covariates data that may contribute to the heterogeneity were not available from the selected articles including the jaundice, cholangitis, tumor size, metastasis, TNM staging, flow and timing of these studies. One study included healthy subjects as control cases [45] and other studies included different benign diseases as controls. Secondly, a small number of studies were included in the qualitative analysis group though we performed a thorough literature search. Bias may have resulted from incomplete retrieval of this identified research. For example, one study concluded that M2-PK did not have a different expression in benign and malignant diseases, thus no valid data was obtained [45]. The number of the patients was small in the majority of studies due to the low frequency of BTC. This might have weakened the statistical significance. Thirdly, publishing and reporting bias should not be ignored as well. Though the Deeks funnel plot asymmetry test did not reveal obvious evidences of asymmetry, potential publication bias among studies cannot be excluded due to their inherent under-power based on a small number of included articles. Even including only English-language studies might have introduced publishing bias. The papers with positive results tend to be published which may be a reason for the unavoidable bias. Consequently, prospective, large scale and comprehensive studies are required to further confirm the effectiveness and the ease of this diagnostic strategy and potential covariate influences before its implementation in clinical practice.

CONCLUSION

Our meta-analysis showed that M2-PK had a better diagnostic accuracy for BTC compared with CA19-9, with moderate diagnostic performance.

Conflicts of interest: The authors disclose no conflicts of interest.

Authors' contribution: X.D.H. and B.L.: conception and design of the work, and its revision; P.H.W., J.Z.C. and J.M.C.: acquisition of the original data; W.L. and N.Z.: data analysis; W.Q.W.: manuscript drafting and data analysis. All authors read and approved the final version of the manuscript

Acknowledgement: This work was supported by National Natural Science Foundation of China (81372578).

Supplementary material: To access the supplementary material visit the online version of the *J Gastrointest Liver Dis* at <http://www.jgld.ro/wp/archive/n1/a12> and <http://dx.doi.org/10.15403/jgld.2014.1121.271.wqw>

REFERENCES

- Miyazaki M, Yoshitomi H, Miyakawa S, et al. Clinical practice guidelines for the management of biliary tract cancers 2015: the 2nd English edition. *J Hepatobiliary Pancreat Sci* 2015;22:249-273. doi:10.1002/jhbp.233
- Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013;63:11-30. doi:10.3322/caac.21166
- Randi G, Malvezzi M, Levi F, et al. Epidemiology of biliary tract cancers: an update. *Ann Oncol* 2009;20:146-159. doi:10.1093/annonc/mdn533
- Patel T. Worldwide trends in mortality from biliary tract malignancies. *BMC Cancer* 2002;2:10. doi:10.1186/1471-2407-2-10
- Sasaki T, Isayama H, Nakai Y, Koike K. Current status of chemotherapy for the treatment of advanced biliary tract cancer. *Korean J Intern Med* 2013;28:515-524. doi:10.3904/kjim.2013.28.5.515
- Choi SB, Kim WB, Song TJ, Suh SO, Kim YC, Choi SY. Surgical outcomes and prognostic factors for ampulla of Vater cancer. *Scand J Surg* 2011;100:92-98. doi:10.1177/145749691110000205
- Feng JF, Zhou XM, Mao WM. Prognostic analysis of carcinoma of the ampulla of Vater: pancreaticoduodenectomy versus local resection. *Hippokratia* 2012;16:23-28.
- Sessa F, Furlan D, Zampatti C, Carnevali I, Franzi F, Capella C. Prognostic factors for ampullary adenocarcinomas: tumor stage, tumor histology, tumor location, immunohistochemistry and microsatellite instability. *Virchows Arch* 2007;451:649-657. doi:10.1007/s00428-007-0444-1
- Khan SA, Davidson BR, Goldin RD, et al. Guidelines for the diagnosis and treatment of cholangiocarcinoma: an update. *Gut* 2012;61:1657-1669. doi:10.1136/gutjnl-2011-301748
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69-90. doi:10.3322/caac.20107
- Sharma A, Sharma KL, Gupta A, Yadav A, Kumar A. Gallbladder cancer epidemiology, pathogenesis and molecular genetics: Recent update. *W J Gastroenterology* 2017;23:3978-3998. doi:10.3748/wjg.v23.i22.3978
- Hsing AW, Bai Y, Andreotti G, et al. Family history of gallstones and the risk of biliary tract cancer and gallstones: A population-based study in Shanghai, China. *Int J Cancer* 2010;121:832-838. doi:10.1002/ijc.22756
- Nagorney DM, Kendrick ML. Hepatic resection in the treatment of hilar cholangiocarcinoma. *Adv Surg* 2006;40:159-171. doi:10.1016/j.yasu.2006.05.009
- Jarnagin WR, Ruo L, Little SA, et al. Patterns of initial disease recurrence after resection of gallbladder carcinoma and hilar cholangiocarcinoma. *Cancer* 2003;98:1689-1700. doi:10.1002/cncr.11699
- Wen H, Yoo SS, Kang J, et al. A new NMR-based metabolomics approach for the diagnosis of biliary tract cancer. *J Hepatol* 2010;52:228-233. doi:10.1016/j.jhep.2009.11.002
- Kumari S, Tewari S, Husain N, et al. Quantification of Circulating Free DNA as a Diagnostic Marker in Gall Bladder Cancer. *Pathol Oncol Res* 2017;23:91-97. doi:10.1007/s12253-016-0087-0
- Steinberg W. The clinical utility of the CA 19-9 tumor-associated antigen. *Am J Gastroenterol* 1990;85:350-355.
- Benson AL III, D'Angelica MI, Abbott DE, et al. NCCN Guidelines Insights: Hepatobiliary Cancers, Version 1.2017. *J Natl Compr Canc Netw* 2017;15:563-573; doi:10.6004/jnccn.2017.0059
- Chalasan N, Baluyut A, Ismail A, et al. Cholangiocarcinoma in patients with primary sclerosing cholangitis: A multicenter case-control study. *Hepatology* 2000;31:7-11. doi:10.1002/hep.510310103
- Patel AH, Harnois DM, Klee GG, LaRusso NF, Gores GJ. The utility of CA 19-9 in the diagnoses of cholangiocarcinoma in patients without primary sclerosing cholangitis. *Am J Gastroenterol* 2000;95:204-207. doi:10.1111/j.1572-0241.2000.01685.x
- Levy C, Lymp J, Angulo P, Gores GJ, Larusso N, Lindor KD. The value of serum CA 19-9 in predicting cholangiocarcinomas in patients with primary sclerosing cholangitis. *Dig Dis Sci* 2005;50:1734-1740. doi:10.1007/s10620-005-2927-8
- Gores GJ. Early detection and treatment of cholangiocarcinoma. *Liver Transpl* 2000;6:S30-S34. doi:10.1053/jlts.2000.18688
- Hultcrantz R, Olsson R, Danielsson A, et al. A 3-year prospective study on serum tumor markers used for detecting cholangiocarcinoma in patients with primary sclerosing cholangitis. *J Hepatol* 1999;30:669-673. doi:10.1016/S0168-8278(99)80198-6
- Aguilar-Olivos NE, Oria-Hernández JA, Valdovinos-Andraca F, Elizondo-Rivera J, Tellez-Avila FI. Effectiveness of biliary pyruvate kinase M2 for diagnosis of malignant biliary strictures: A prospective and comparative study. *Gastrointest Endosc* 2016;83:AB125. doi:10.1016/j.gie.2016.03.044
- Navaneethan U, Lourdasamy V, Poptic E, Hammel JP, Sanaka MR, Parsi MA. Comparative effectiveness of pyruvate kinase M2 in bile, serum carbohydrate antigen 19-9, and biliary brushings in diagnosing malignant biliary strictures. *Dig Dis Sci* 2015;60:903-909. doi:10.1007/s10620-014-3397-7
- Dhar DK, Olde Damink SW, Brindley JH, et al. Pyruvate kinase M2 is a novel diagnostic marker and predicts tumor progression in human biliary tract cancer. *Cancer* 2013;119:575-585. doi:10.1002/cncr.27611
- Warburg O. On the origin of cancer. *Science* 1956;123:309-314. doi:10.1126/science.123.3191.309
- Christofk HR, Vander Heiden MG, Harris MH, et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* 2008;452:230-233. doi:10.1038/nature06734
- Shiroki T, Yokoyama M, Tanuma N, et al. Enhanced expression of the M2 isoform of pyruvate kinase is involved in gastric cancer development by regulating cancer-specific metabolism. *Cancer Sci* 2017;108:931-940. doi:10.1111/cas.13211
- Koss K, Harrison RF, Gregory J, Darnton SJ, Anderson MR, Jankowski JA. The metabolic marker tumour pyruvate kinase type M2 (tumour M2-PK) shows increased expression along the metaplasia-dysplasia-adenocarcinoma sequence in Barrett's oesophagus. *J Clin Pathol* 2004;57:1156-1159. doi:10.1136/jcp.2004.018150
- Hardt PD, Ewald N. Tumor M2 pyruvate kinase: A tumor marker and its clinical application in gastrointestinal malignancy. *Expert Rev Mol Diagn* 2008;8:579-585. doi:10.1586/14737159.8.5.579
- Mukherjee J, Phillips JJ, Zheng S, Wiencke J, Ronen SM, Pieper RO. Pyruvate kinase M2 expression, but not pyruvate kinase activity, is up-regulated in a grade-specific manner in human glioma. *PLoS One* 2013;8:e57610. doi:10.1371/journal.pone.0057610
- Zhan QS, Wang LS, Jiang HZ, Jiang XQ, Shi RY, Shi-Zhen LI. Diagnosis Value of Detecting Plasma Pyruvate Kinase M2 (M2-PK) in Patients with Gastric Cancer. *Journal of Modern Laboratory Medicine* 2006;(5):11-13.
- Abdullah M, Rani AA, Simadibrata M, Fauzi A, Syam AF. The value of fecal tumor M2 pyruvate kinase as a diagnostic tool for colorectal cancer screening. *Acta Med Indones* 2012;44:94-99.

35. Huang JX, Zhou Y, Wang CH, Yuan WW, Zhang ZD, Zhang XF. Tumor M2-pyruvate kinase in stool as a biomarker for diagnosis of colorectal cancer: A meta-analysis. *J Cancer Res Ther* 2014;10 Suppl:C225-C228. doi:[10.4103/0973-1482.145886](https://doi.org/10.4103/0973-1482.145886)
36. Schneider J, Morr H, Velcovsky HG, Weisse G, Eigenbrodt E. Quantitative detection of tumor M2 -pyruvate kinase in plasma of patients with lung cancer in comparison to other lung diseases. *Cancer Detect Prev* 2000;24:531-535.
37. Liu J, Zhu H, Jiang H, et al. Tumor M2 pyruvate kinase in diagnosis of nonsmall cell lung cancer:A meta-analysis based on Chinese population. *J Cancer Res Ther* 2015;11 Suppl 1:C104-C106. doi:[10.4103/0973-1482.163857](https://doi.org/10.4103/0973-1482.163857)
38. Kumar Y, Gurusamy K, Pamecha V, Davidson BR. Tumor M2-pyruvate kinase as tumor marker in exocrine pancreatic cancer a meta-analysis. *Pancreas* 2007;35:114-119. doi:[10.1097/mpa.0b013e3180537237](https://doi.org/10.1097/mpa.0b013e3180537237)
39. Ahmed AS, Dew T, Lawton FG, et al. M2-PK as a novel marker in ovarian cancer. A prospective cohort study. *Eur J Gynaecol Oncol* 2007;28:83-88.
40. Roigas J, Schulze G, Raytarowski S, Jung K, Schnorr D, Loening SA. Tumor M2 Pyruvate Kinase in Plasma of Patients with Urological Tumors. *Tumor Biol* 2001;22:282-285. doi:[10.1159/000050628](https://doi.org/10.1159/000050628)
41. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2:a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529 -536. doi:[10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)
42. 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. doi:[10.1016/j.jclinepi.2005.01.016](https://doi.org/10.1016/j.jclinepi.2005.01.016)
43. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000;53:1119-1129. doi:[10.1016/S0895-4356\(00\)00242-0](https://doi.org/10.1016/S0895-4356(00)00242-0)
44. Goonetilleke KS, Mason JM, Siriwardana P, King NK, France MW, Siriwardana AK. Diagnostic and prognostic value of plasma tumor M2 pyruvate kinase in periampullary cancer: evidence for a novel biological marker of adverse prognosis. *Pancreas*. 2007;34:318-324. doi:[10.1097/MPA.0b013e31802ee9c7](https://doi.org/10.1097/MPA.0b013e31802ee9c7)
45. Li YG, Zhang N. Clinical significance of serum tumour M2-PK and CA19-9 detection in the diagnosis of cholangiocarcinoma. *Dig Liver Dis* 2009;41:605-608. doi:[10.1016/j.dld.2008.11.010](https://doi.org/10.1016/j.dld.2008.11.010)
46. Siqueira E, Schoen RE, Silverman W, et al. Detecting cholangiocarcinoma in patients with primary sclerosing cholangitis. *Gastrointest Endosc* 2002;56:40-47. doi:[10.1067/mge.2002.125105](https://doi.org/10.1067/mge.2002.125105)
47. Ker CG, Chen JS, Lee KT, Sheen PC, Wu CC. Assessment of serum and bile levels of CA19-9 and CA125 in cholangitis and bile duct carcinoma. *J Gastroenterol Hepatol* 1991;6:505-508. doi:[10.1111/j.1440-1746.1991.tb00896.x](https://doi.org/10.1111/j.1440-1746.1991.tb00896.x)
48. Glas AS, Lijmer JG, Prins MH, Bossel GJ, Bossuyt PM. The diagnostic odds ratio:a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129-1135. doi:[10.1016/S0895-4356\(03\)00177-X](https://doi.org/10.1016/S0895-4356(03)00177-X)
49. Cerwenka H1, Aigner R, Bacher H,et al. TUM2-PK (pyruvate kinase type tumor M2), CA19-9 and CEA in patients with benign, malignant and metastasizing pancreatic lesions. *Anticancer Res* 1999;19:849-851.
50. Aloysius MM, Zaitoun AM, Bates TE, et al. Complete absence of M2-pyruvate kinase expression in benign pancreatic ductal epithelium and pancreaticobiliary and duodenal neoplasia. *BMC Cancer* 2009;9:327. doi:[10.1186/1471-2407-9-327](https://doi.org/10.1186/1471-2407-9-327)
51. Schulze G. The tumor marker tumor M2-PK:an application in the diagnosis of gastrointestinal cancer. *Anticancer Res* 2000;20(6D):4961-4964.
52. Schneider J, Schulze G. Comparison of tumor M2-pyruvate kinase (tumor M2-PK), carcinoembryonic antigen (CEA), carbohydrate antigens CA 19-9 and CA 72-4 in the diagnosis of gastrointestinal cancer. *Anticancer Res* 2003;23:5089-5093.
53. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157-162.
54. Park MS, Kim TK, Kim KW, et al. Differentiation of Extrahepatic Bile Duct Cholangiocarcinoma from Benign Stricture: Findings at MRCP versus ERCP1. *Radiology* 2004;233:234-240. doi:[10.1148/radiol.2331031446](https://doi.org/10.1148/radiol.2331031446)

Supplementary Table I
QUADAS-2 ITEMS

Study	Was a consecutive or random sample of patients enrolled?	Was a case-control design avoided?	Did the study avoid inappropriate exclusions?	Were the index test results interpreted without knowledge of the results of the reference standard?	If a threshold was used, was it prespecified?	Is the reference standard likely to correctly classify the target condition?	Were the reference standard results interpreted without knowledge of the results of the index test?	Was there an appropriate interval between index tests and reference standard?	Did all patients receive a reference standard?	Did all patients receive the same reference standard?	Were all patients included in the analysis?
Goonetilleke, K.S.(2007)	Y	Y	Y	Y	N	Y	U	Y	Y	Y	Y
Y.G Lia(2009)	Y	Y	Y	Y	N	Y	U	U	Y	N	Y
Dhar, Damink et al(2013)	U	U	U	U	N	Y	U	Y	Y	N	N
Dhar, Damink et al(2013)	U	U	U	U	N	Y	U	Y	Y	N	
Navaneethan, Lourdusamy et al.(2015)	Y	Y	Y	Y	N	Y	U	Y	N	N	N
Navaneethan, Lourdusamy et al.(2015)	Y	Y	Y	Y	N	Y	U	Y	N	N	N
Aguilar-Olivos, N.E.(2016)	U	Y	U	Y	N	U	U	Y	Y	N	U

Y, Yes; U, Unclear; N ,No.

Supplementary Table II. The meta-regression and subgroup analysis

Subgroup		Sensitivity (95%CI)	P1	Specificity (95%CI)	P2
Sample size	Yes	0.79 (0.70-0.88)	0.47	0.79 (0.68-0.89)	0.07
	No	0.75 (0.58 - 0.92)		0.88 (0.77 - 1.00)	
Specimen	Yes	0.75 (0.63 - 0.86)	0.12	0.75 (0.63 - 0.88)	0.09
	No	0.83 (0.69 - 0.97)		0.86 (0.73 - 0.99)	
Same method	Yes	0.77 (0.65 - 0.90)	0.19	0.72 (0.61 - 0.82)	0.00
	No	0.79 (0.67 - 0.92)		0.88 (0.82 - 0.95)	
Ref test	Yes	0.79 (0.71 - 0.87)	0.53	0.80 (0.71 - 0.89)	0.59
	No	0.67 (0.24 - 1.00)		0.92 (0.73 - 1.00)	
Subject	Yes	0.79 (0.71 - 0.87)	0.53	0.80 (0.71 - 0.89)	0.59
	No	0.67 (0.24 - 1.00)		0.92 (0.73 - 1.00)	
Cut-off	Yes	0.78 (0.68 - 0.88)	0.19	0.75 (0.66 - 0.84)	0.00
	No	0.82 (0.67 - 0.97)		0.89 (0.81 - 0.97)	

Sample size : Yes : larger than 100 cases ; No : less than 100 cases

Specimen: Yes : plasma ; No : bile

Same method : Yes : the same golden standard ; No : more than one golden standard

Ref test : Yes : detailed description of the test method ; No : not detailed description of the test method

Subject : Yes : detailed description of the subject ; No : not detailed description of the subject

Cut-off : Yes : above 25U/ml ; No : below 25U/ml