



## Research Paper

# Development and Validation of Novel Diagnostic Models for Biliary Atresia in a Large Cohort of Chinese Patients



Rui Dong<sup>a,1</sup>, Jingying Jiang<sup>a,1</sup>, Shouhua Zhang<sup>b,1</sup>, Zhen Shen<sup>a</sup>, Gong Chen<sup>a</sup>, Yanlei Huang<sup>a</sup>, Yijie Zheng<sup>c,\*\*</sup>, Shan Zheng<sup>a,\*</sup>

<sup>a</sup> Department of Pediatric Surgery, Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defect, Shanghai 201102, China

<sup>b</sup> Department of General Surgery, Jiangxi Provincial Children's Hospital, Nanchang, Jiangxi Province, 330006, China

<sup>c</sup> Medical Scientific Liaison Asian Pacific, Abbott Diagnostics Division, Abbott Laboratories, Shanghai 200032, China

## ARTICLE INFO

## Article history:

Received 22 May 2018

Received in revised form 6 July 2018

Accepted 17 July 2018

Available online 1 August 2018

## Keywords:

Biliary atresia

Neonatal cholestasis

Gamma-glutamyl transpeptidase

Nomogram

## ABSTRACT

**Background & aims:** The overlapping features of biliary atresia (BA) and the other forms of neonatal cholestasis (NC) with different causes (non-BA) has posed challenges for the diagnosis of BA. This study aimed at developing new and better diagnostic models for BA.

**Methods:** We retrospectively analyzed data from 1728 newborn infants with neonatal obstructive jaundice (NOJ). New prediction models, including decision tree (DT), random forest (RF), and multivariate logistic regression-based nomogram for BA were created and externally validated in an independent set of 508 infant patients.

**Results:** Five predictors, including gender, weight, direct bilirubin (DB), alkaline phosphatase (ALP), and gamma-glutamyl transpeptidase (GGT) were significantly different between the BA and non-BA groups ( $P < .05$ ), from which DT, RF, and nomogram models were developed. The area under the receiver operating characteristic (ROC) curve (AUC) value for the nomogram was 0.898, which was greater than that of a single biomarker in the prediction of BA. Performance comparison of the three diagnostic models showed that the nomogram displayed better discriminative ability (sensitivity, 85.7%; specificity, 80.3%; PPV, 0.969) at the optimal cut-off value compared with DT and RF, which had relatively similar high sensitivity and PPV (0.941 and 0.947, respectively), but low specificity in the modeling group. In sub-analysis of the discriminative capacity between the nomogram and GGT ( $<300$  or  $\geq 300$ ), we found that the nomogram was superior to the GGT alone in the preoperative diagnosis of BA.

**Conclusions:** The nomogram has demonstrated better performance for the prediction of BA, holding promise for future clinical application.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\* Correspondence to: Dr. Shan Zheng, Department of Pediatric Surgery, Children's Hospital of Fudan University, Shanghai Key Laboratory of Birth Defect, 399 Wan Yuan Rd., Shanghai 201102, China.

\*\* Correspondence to: Dr. Yijie Zheng, Medical Scientific Affairs, Abbott Diagnostics, Shanghai, 200032, China.

E-mail addresses: [yijie.zheng@abbott.com](mailto:yijie.zheng@abbott.com) (Y. Zheng), [szheng@shmu.edu.cn](mailto:szheng@shmu.edu.cn) (S. Zheng).

<sup>1</sup> These authors contributed to the work equally and should be regarded as co-first authors.

## Research in Context

### Evidence Before This Study

Gamma-glutamyl transpeptidase (GGT) has been proposed as a serum marker for differentiating biliary atresia (BA) from neonatal hepatitis in the disease diagnosis. However, the reliability and reproducibility of serum GGT activity alone were limited in an accurate diagnosis of BA. Added Value of This Study

This study of a large cohort of Chinese infant patients has developed and validated a novel nomogram using GGT in combination with other BA-related factors for better diagnosis of BA. Implications of All the Available Evidence

The results demonstrate that this nomogram is superior to the GGT alone in the preoperative diagnosis of BA, and thus holds promise in the clinical application to better predict BA in newborn infants.

## 1. Introduction

Biliary atresia (BA) is an uncommon, but serious disorder in newborn infants, which is characterized by the obstruction of extra- or intra-hepatic bile ducts [1–4]. If left undiagnosed and untreated, BA can rapidly progress into biliary cirrhosis and hepatic failure, which will require liver transplantation, and can even lead to death within 2–3 years after birth, in a proportion of BA patients [5–8]. Although this disease rarely occurs among infants worldwide, the incidence of BA is high in the Asia-Pacific region.

Currently in Eastern Asia, BA has an overall incidence of approximately 1.51 in 10,000 live births, which is markedly greater than that in the United States [1–3]. In fact, in our hospital, which is one of the largest pediatric hospitals in China, as many as 400 infant patients per year are diagnosed with BA. A majority of these patients received the Kasai operation and postoperative conventional treatment with medications (e.g. antibiotics, hormones, ursodeoxycholic acid). In our previous study, a two-year survival rate was 53.7% in BA patients surviving with their native livers, while the remaining BA patients required subsequent liver transplantations, but the two-year survival rates of these patients were unavailable because of difficulty in patient tracking [9]. The key to restoring the flow of the bile ducts and obtaining good clinical outcomes is diagnosing and treating the disease early. However, the misdiagnosis of BA can result in inappropriate treatment and unnecessary surgery [10–15]. In our previous study, we retrospectively analyzed data obtained from 602 BA surgery cases, of which only 86% were postoperatively confirmed with BA by pathological studies [16]. Therefore, it is critical to establish reliable models for the early detection and diagnosis of BA. Unfortunately, the definitive diagnosis and confirmation of BA in suspected infants generally requires a liver biopsy and intraoperative cholangiography (IOC) during the surgical procedure, and these diagnostic methods have turned out to be invasive, time-consuming, and costly [6, 17–20]. Obviously, there is an urgent need for a reliable and better diagnostic approach to distinguish BA from other form of neonatal cholestasis (NC) with different causes.

Serum activity of gamma-glutamyl transpeptidase (GGT), as a non-invasive marker, has been extensively studied and proposed for the diagnosis of BA [21–28]. In fact, GGT >300 U/L, or a daily increase in its serum activity of 6 U/L for differentiating BA from neonatal hepatitis, had an accuracy of 85% and 88%, respectively [22]. El-Guindi and colleagues reported that the serum activity of GGT at a cutoff value (>286 U/L) had a sensitivity of 76.7% and specificity of 80% for the diagnosis of BA [29]. In our previous study, GGT activity in serum also showed good performance ability in discriminating BA from other

causes in the Chinese population [21]. However, the reliability, accuracy, and reproducibility of GGT activity alone was questionable. For example, it has been demonstrated that healthy infants at birth have higher levels GGT [23], and the normal range for levels of GGT may vary dependent of age. Indeed, GGT corrected with age has shown improvement in the accuracy of predicting BA. Until now, diagnostic models using GGT in combination with other BA-related factors, which are anticipated to offer a better approach for the diagnosis of BA, have not been developed and evaluated for the diagnosis of BA.

In the present study, the demographic, clinical, and laboratory data from a large-scale of infant patients with neonatal obstructive jaundice (NOJ) were analyzed to examine the association between a number of risk factors and BA. New prediction models, including decision tree (DT), random forest (RF), and multivariate logistic regression-based nomogram were developed and validated for the diagnosis of BA. The results obtained through this study may offer a novel and better algorithm for the diagnosis of BA and hold potential for clinical application.

## 2. Patients and Methods

### 2.1. Human Subjects and Study Design

In this study, demographic, clinical, and laboratory test data of 1728 infant patients with NOJ between January 2012 and December 2017 at the Children's Hospital of Fudan University were collected, reviewed, and analyzed. Of these, 1512 patients with BA were assigned to the BA group, while 216 patients had other causes of NC, including 196 patients with neonatal hepatitis, 10 with alagille syndrome, and 8 with biliary hypoplasia, who were allocated to the non-BA group. Intraoperative cholangiography and subsequent histological examination of liver biopsies were used for diagnostic confirmation of BA and non-BA. The following inclusion criteria for BA patients were used with intent in this study: (1) Pediatric patients were diagnosed as BA by intraoperative cholangiography in combination with histological features of liver biopsies, showing ductular proliferation, canalicular and cellular bile stasis, portal or periportal inflammation, swelling and vacuolization of biliary epithelial cells, edema and monocytic inflammatory cell infiltration of portal tracts, fibrosis with the presence of bile plugs in the portal tract bile ducts, hepatocyte ballooning, and end-stage cirrhosis; (2) No other severe systematic deformity was present, such as BA splenic malformation syndrome. The inclusion criteria for pediatric patients with cholestasis were cholestasis without BA, as confirmed by intraoperative cholangiography, and no other severe malformation in other systems. Infants who had bile duct dysplasia and/or malformation of other systems were excluded from the current study.

This study was reviewed and approved by the Institutional Review Board (IRB) at the Children's Hospital of Fudan University, with a waiver of requirement for informed consent due to the nature of this retrospective study. The study was performed in compliance with the Declaration of Helsinki, and other relevant regulations.

### 2.2. Development and Validation of Decision Tree Model, Random Forest Model, and Logistic Regression-based Nomogram for the Diagnosis of BA

Decision tree (DT) was conducted via R package *rpart*, and a DT plot was drawn via *rattle* package. In brief, the root node asked, or the first question: Was  $\ln(\text{GGT}) < 4.8$  in the patient? In generation of classification trees, “no” indicated a branch to the right, while “yes” represented a branch to the left. Terminal nodes were eventually for the prediction of BA.

Random forest (RF), a tree-based ensemble consisting of tree-structured classifiers, was built for the prediction via *RF* package with 500 regression trees. The importance of variables was shown in a figure, using mean decrease accuracy and Gini.

A diagnosis nomogram was constructed, based on multivariate logistic regression analysis, using the *rms* package. The independent variables included gender, weight, DB, ln(ALP), and ln(GGT). Decision curve analysis (DCA) was performed to finalize the ranges of threshold probabilities within which the nomograms were clinically valuable by *rmda* package.

### 2.3. Statistical Analysis

Statistical analysis was performed using SAS 9.4 and R software. Gender was described by n (%). Ln-transformation was conducted for right-skewed distributed variables including ALP and GGT. Continuous variables, which were not normally distributed, including

weight, DB, ln(ALP) and ln(GGT), were expressed as median and quartiles (Q1, Q3). In univariate analysis, the Chi-squared test was conducted for gender, while Wilcoxon tests were used and performed for all continuous variables. ROC curves were constructed to calculate the best cutoff point and area under curve (AUC) for DB, ln(ALP) and ln(GTT) as single predictor separately using the training data. The sensitivity, specificity, PPV, and NPV were used to show the predictive properties using validation data. The performance of the nomogram was measured by the C-index and calibration curve with 1000 Bootstrap resample. In the present study, both internal and external validations were conducted with training data and validation data for all prediction models. Statistically significant difference was defined as a *P*-value <0.05.

**Table 1**  
Characteristics of the study subjects and univariate analysis.

Item	Non-BA	BA	Total	Method	Statistic	<i>P</i> value
Gender				Chi-square test	$\chi^2=66.09$	<0.0001
* Male (%)	174 (80.56)	773 (51.12)	947 (54.80)			
* Female (%)	42 (19.44)	739 (48.88)	781 (45.20)			
Age (days)				Rank-sum test	$Z = 0.35$	0.7254
* N (Missing)	216 (0)	1512 (0)	1728 (0)			
* Mean $\pm$ SD	74.44 $\pm$ 24.28	73.73 $\pm$ 24.86	73.82 $\pm$ 24.78			
* Median	71.00	71.00	71.00			
* Q1, Q3	58.00, 84.00	57.00, 86.00	57.00, 85.00			
* Min, Max	33.00, 175.00	3.00, 200.00	3.00, 200.00			
Weight (kg)				Rank-sum test	$Z = -5.38$	<0.0001
* N (Missing)	193 (23)	1297 (215)	1490 (238)			
* Mean $\pm$ SD	4.78 $\pm$ 1.17	5.22 $\pm$ 0.99	5.16 $\pm$ 1.03			
* Median	5.00	5.00	5.00			
* Q1, Q3	4.00, 5.50	4.50, 6.00	4.50, 6.00			
* Min, Max	2.07, 10.00	2.00, 8.50	2.00, 10.00			
TB (mmol/l)				Rank-sum test	$Z = -0.65$	0.5142
* N (Missing)	216 (0)	1510 (2)	1726 (2)			
* Mean $\pm$ SD	168.47 $\pm$ 54.97	170.23 $\pm$ 74.38	171.57 $\pm$ 58.00			
* Median	160	151.3	160.20			
* Q1, Q3	130.80, 194.80	125.70, 198.30	134.20, 195.50			
* Min, Max	76.80, 387.90	74.90, 489.10	55.40, 533.20			
DB (mmol/l)				Rank-sum test	$Z = -4.17$	<0.0001
* N (Missing)	216 (0)	1510 (2)	1726 (2)			
* Mean $\pm$ SD	121.96 $\pm$ 54.23	128.91 $\pm$ 40.99	128.04 $\pm$ 42.91			
* Median	109.65	122.30	121.60			
* Q1, Q3	86.75, 140.30	101.90, 147.60	100.10, 147.00			
* Min, Max	11.70, 342.20	27.30, 337.20	11.70, 342.20			
ALT (U/L)				Rank-sum test	$Z = 0.66$	0.5116
* N (Missing)	216 (0)	1504 (8)	1720 (8)			
* Mean $\pm$ SD	129.03 $\pm$ 114.50	110.87 $\pm$ 99.68	113.15 $\pm$ 101.80			
* Median	99.00	91.00	91.50			
* Q1, Q3	56.00, 157.00	60.00, 137.50	59.00, 140.00			
* Min, Max	8.00, 670.00	4.00, 2641.00	4.00, 2641.00			
AST (U/L)				Rank-sum test	$Z = -0.29$	0.7726
* N (Missing)	213 (3)	1484 (28)	1697 (31)			
* Mean $\pm$ SD	198.83 $\pm$ 173.15	168.06 $\pm$ 101.17	171.92 $\pm$ 113.14			
* Median	138.00	144.00	144.00			
* Q1, Q3	89.00, 267.00	105.00, 200.50	103.00, 203.00			
* Min, Max	18.00, 1146.00	20.00, 1027.00	18.00, 1146.00			
ln(ALP)				Rank-sum test	$Z = 3.65$	0.0003
* N (Missing)	184 (32)	1395 (117)	1579 (149)			
* Mean $\pm$ SD	6.44 $\pm$ 0.43	6.31 $\pm$ 0.41	6.32 $\pm$ 0.41			
* Median	6.41	6.32	6.34			
* Q1, Q3	6.17, 6.72	6.06, 6.57	6.07, 6.59			
* Min, Max	5.57, 7.45	4.42, 7.60	4.42, 7.60			
ln(GGT)				Rank-sum test	$Z = -15.30$	<0.0001
* N (Missing)	186 (30)	1354 (158)	1540 (188)			
* Mean $\pm$ SD	5.21 $\pm$ 0.89	6.47 $\pm$ 0.83	6.32 $\pm$ 0.94			
* Median	5.16	6.59	6.42			
* Q1, Q3	4.61, 5.71	5.92, 7.10	5.69, 7.04			
* Min, Max	2.30, 7.78	3.56, 8.59	2.30, 8.59			

### 3. Results

#### 3.1. Demographic, Laboratory, and Clinical Characteristics of the Study Subjects

A total of 1728 infant patients, spanning the period between January 2012 and December 2017, who met the eligibility criteria were retrospectively enrolled, of which 1512 (87.5%) patients were diagnosed as BA, while 216 (12.5%) patients were confirmed to have cholestasis with other cause or non-BA. The demographic, laboratory, and clinical characteristics of the study subjects in the BA and non-BA groups were summarized in Table 1. The mean age of the infant patients was 73.8 (SD, 24.8) days, with 73.7 (SD, 24.9) in the BA group and 74.4 (SD, 24.3) in the non-BA group. A majority of non-BA patients were male (80.6%), while the gender distribution was nearly equal in the BA group (51% male, 49% female). The detailed description of other characteristics including weight, TB, DB, ALT, AST, ln(ALP) and ln(GGT) were also listed in Table 1. As a result, gender, weight, DB, ln(ALP), and ln(GGT) were identified to have significant differences between the BA and non-BA groups ( $P < .05$ ), whereas the two groups did not show any differences in age, TB, ALT, and AST ( $P > .05$ ).

Preoperative levels of total bilirubin, direct bilirubin, and GGT were significantly higher in the BA group ( $P < .05$ ), whereas the non-BA group had higher alkaline phosphatase levels (ALP) ( $P < .05$ ).

#### 3.2. Univariate Logistic Regression Analysis of Variables Significantly Associated with BA

To determine the independent variables associated with BA, univariate logistic regression analysis was performed. Statistically significant differences in the variables, including gender, weight, DB, ln(ALP), and ln(GGT), were identified between the BA and non-BA groups (Table 1) ( $P < .05$ ). ln(GGT) showed a good independent prediction property with an AUC  $> 0.8$ . However, the AUC of DB and ln(ALP) were  $< 0.6$  (Table 2).

#### 3.3. Establishment and Validation of the Decision Tree Model in Predicting BA

The decision tree (DT) for prediction of BA included 5 study variables: gender, weight, DB, ln(ALP), and ln(GGT), and were constructed via R package *rpart*. In an establishment of the DT model, the first question, also known as the root node, queried (1) was ln(GGT)  $< 4.8$  in the patient? In classification trees, “no” represented a branch to the right. If the answer was “no”, the second question asked: (2) was ln(GGT)  $< 5.7$  in the patient? The infant patients who

did not meet the criteria were classified as BA (terminal node 7). For those patients who met the criteria, the tree further queried: (3) was the gender of the patient male? If the answer was “no”, the patients were classified as BA (terminal node 13). For those male patients, the next question asked: (4) was the weight of the patient  $< 3.8$ ? If “yes”, the patients were classified as BA (terminal node 24). For those who did not meet the criteria, the fifth question was: (5) was DB  $< 74$  in the patient? The patients who did not fulfill the criteria were classified as BA (terminal node 50). If “no”, the next question queried: (6) was the weight of the patient  $< 5.2$ ? If “no”, the patients were classified as BA (terminal node 103). For those who did not meet the criteria, the subsequent question was: (7) was DB  $\geq 134$  in the patient? The patients were classified as BA with different probabilities (terminal nodes 204 and 205).

For the patients with ln(GGT)  $< 4.8$  in the question 1 (root node) in the left branch, the first question queried: (1) was the gender of the patient male? If the answer was “yes”, the second question asked: (2) was the weight of the patient  $< 6.2$ ? the patients were classified as BA with different probabilities (terminal nodes 8 and 9). If the answer was “no” in question 1, the next question asked: (3) was ln(GGT)  $< 4.2$  in the patient? The patients were classified as BA with different probabilities (terminal nodes 10 and 11). As shown in Table 2 and Fig. 1, the DT revealed that the probability of BA was 0.96 when ln(GGT) was  $> 5.7$ , whereas the probability of non-BA was 0.93 when ln(GGT) was  $< 5.7$  with male gender and weight  $< 3.8$ .

#### 3.4. Establishment and Validation of the Random Forest Model in Predicting BA

A random forest (RF) classification algorithm was created using RF package with a 500 regression tree for the prediction of BA. As with other models, all variables were included and time lags with more than five steps were trained. The importance of each variable was subsequently measured by calculating how much reduction each variable offers when they were added to the RF model. As shown in Fig. 2, ln(GTT) was the most important variable and was most closely related to BA, which was followed by gender, weight, DB, and ln(ALP) by mean decrease accuracy, and by DB, ln(ALP), weight, and gender by mean decrease Gini.

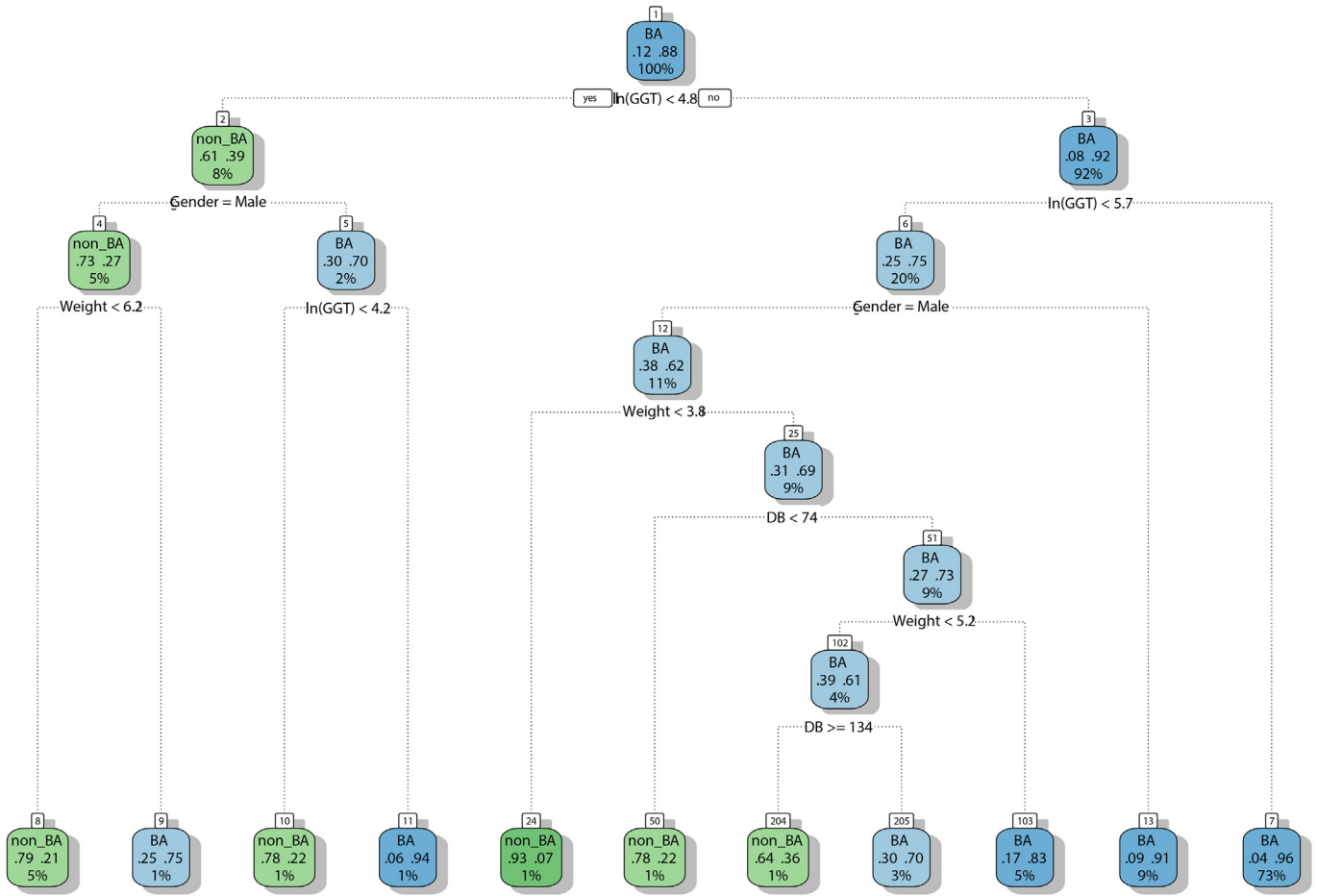
#### 3.5. Establishment and Validation of the Logistic Regression-based Nomogram in Predicting BA

A nomogram to predict BA was developed on the basis of multivariate logistic regression analysis using the five factors which were identified to be statistically different between the BA and non-BA groups,

**Table 2**  
Prediction properties of internal and external validation.

Method	Internal validation						External validation			
	AUC	Cutoff	SEN	SPE	PPV	NPV	SEN	SPE	PPV	NPV
Decision Tree*	/	/	0.980	0.554	0.941	0.791	0.910	0.405	0.784	0.653
Random Forest*	/	/	0.974	0.605	0.947	0.760	0.917	0.446	0.798	0.692
DB	0.567	93.7	0.837	0.318	0.900	0.212	0.646	0.537	0.769	0.389
ln(ALP)	0.572	76.9	0.950	0.146	0.890	0.286	0.878	0.372	0.769	0.563
		6.3	0.656	0.468	0.900	0.157	0.514	0.545	0.729	0.320
ln(GGT)	0.845	7.0	0.950	0.121	0.887	0.249	0.997	0	0.703	0
		5.8	0.786	0.796	0.966	0.338	0.885	0.124	0.706	0.313
GGT $< 300$ (U/l)		4.9	0.950	0.408	0.921	0.528	0.997	0	0.703	0
		0.000	1.000	/	0.344	0.000	1.000	/	0.508	
GGT $\geq 300$ (U/l)			0.960	0.200	0.965	0.178	0.945	0.033	0.867	0.083
Logistic Regression <sup>§</sup>	0.898	0.85	0.857	0.803	0.969	0.434	0.712	0.760	0.876	0.526
		0.65	0.950	0.599	0.945	0.622	0.837	0.570	0.823	0.595
GGT $< 300$ (U/l)			0.448	0.957	0.952	0.477	0.284	0.890	0.714	0.563
GGT $\geq 300$ (U/l)			0.951	0.350	0.971	0.237	0.900	0.367	0.905	0.355

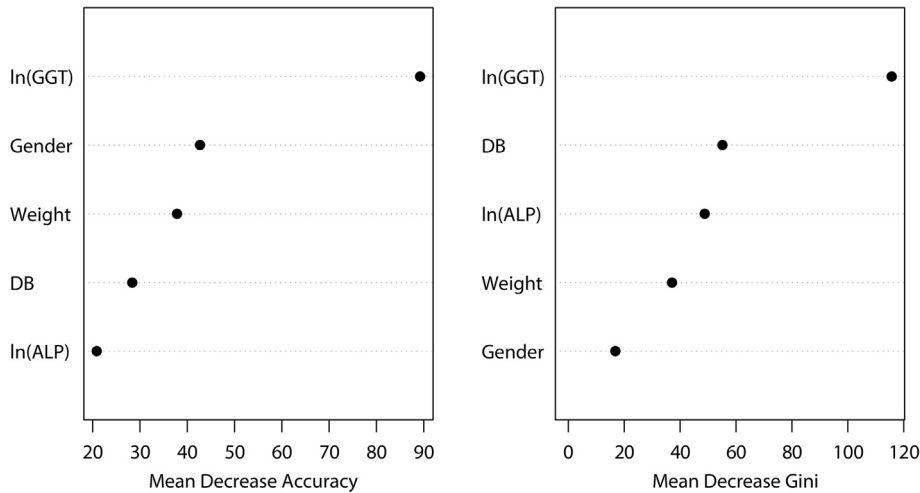
Abbreviations: AUC, area under receiver operating characteristic (ROC) curve; SEN, sensitivity; SPE, specificity; PPV, positive prediction value; NPV, negative prediction value. Note: \*Based on the combination of gender, weight, DB, ln(ALP) and ln(GGT).<sup>§</sup>The external validation was based on the cutoff value.



**Fig. 1.** Decision tree for the prediction of BA using gender, weight, DB, ln(ALP), and ln(GGT). The DT included five variables: gender, weight, DB, ln(ALP), and ln(GGT), and were built via R package *rpart*. In the formation of DT model, the root node queried: was ln (GGT) <4.8 in the patient? In classification trees, “no” represented a branch to the right, while “yes” indicated a branch to the left. A total of 11 terminal nodes were generated for the DT model.

including gender, weight, DB, ln(ALP), and ln(GGT). The relationship between these factors and BA was assessed using the multivariate logistic regression analysis and the resulting data were presented in Table 3. The odds ratios for BA were calculated for these factors. We identified that gender, weight, DB, ln (ALP), and ln(GGT) were significantly associated with BA, and thus were used as predictors to build

the nomogram prediction model for BA (Fig. 3A). As shown in Fig. 3A, there were 8 rows in the nomogram, with the rows ranging from 2 to 6 representing the included variables. The points of the five variables were added up to the total points, which were displayed in the row 7 and corresponded to the risk probability in the prediction of BA in the row 8, and the nomogram showed the risk of BA as a percentage. The



**Fig. 2.** Random forest for evaluation of the importance of the study variables in the prediction of BA. RF classification algorithm using RF package with 500 regression tree was constructed for the prediction of BA using the five variables (gender, weight, DB, ln(ALP), and ln(GGT)), in time lags with more than five steps were trained. The importance of each variable was subsequently evaluated in RF.

**Table 3**  
The logistic regression analysis to construct the nomogram for the prediction of BA.

Parameter	$\beta$	Wald $\chi^2$	OR	95% CI	P value
Intercept	-12.0165	35.0449	/	/	<0.0001
Gender	1.8558	52.6817	6.397	3.876 10.559	<0.0001
Weight	0.6704	38.4313	1.955	1.582 2.417	<0.0001
DB	0.0081	12.2552	1.008	1.004 1.013	0.0005
ln(ALP)	-0.5294	4.0630	0.589	0.352 0.985	0.0438
ln(GGT)	1.8263	147.9369	6.211	4.627 8.336	<0.0001

Abbreviation: OR, odds ratio; CI, confidential interval.

area under ROC curve (AUC) value of 0.898 for the nomogram was obtained, which was greater than the AUC values of 0.848 for ln(GGT), 0.572 for ln(ALP), and 0.567 for DB in the prediction of BA (Fig. 3B, Table 3).

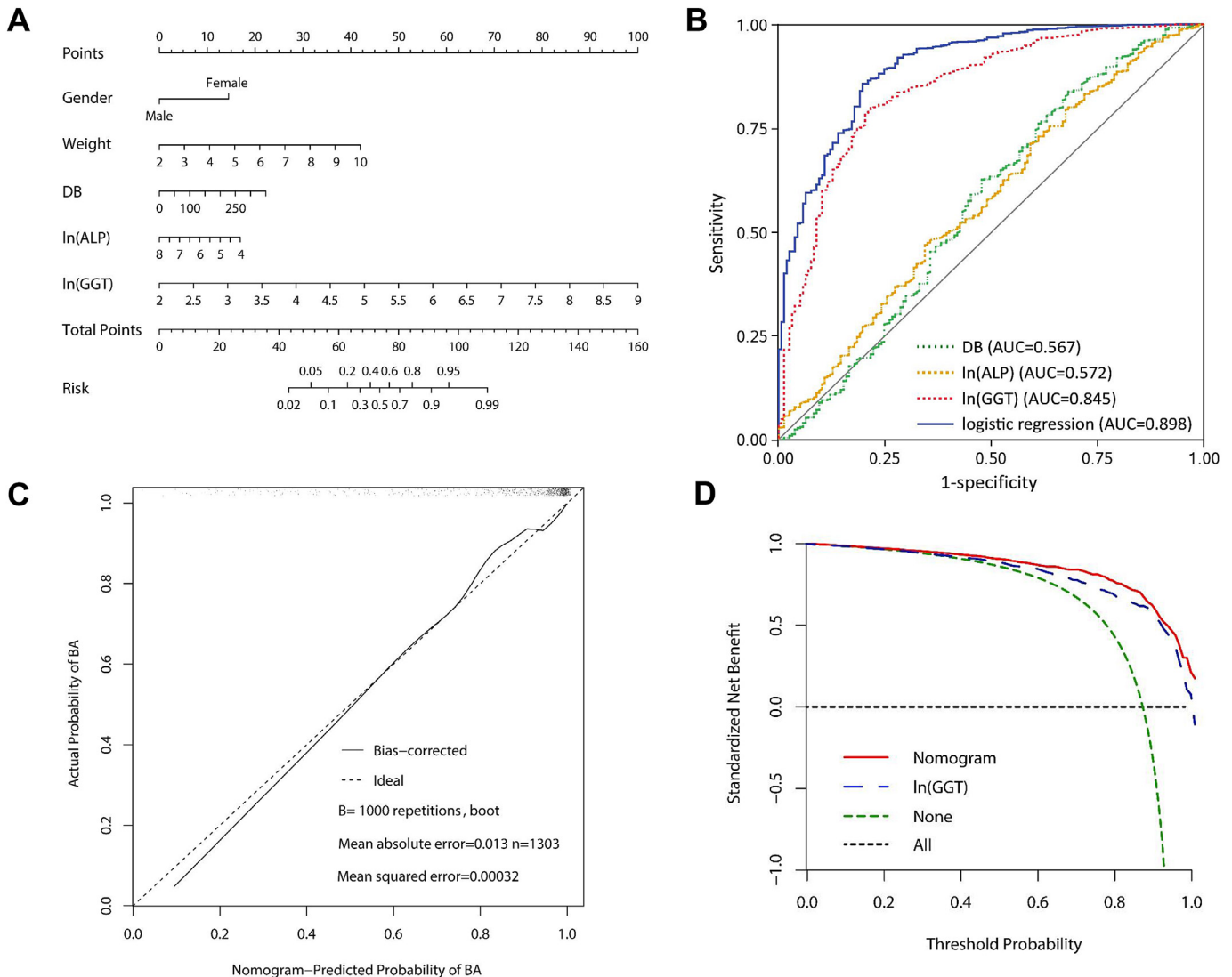
The calibration blots with 1000 Bootstrap resample were illustrated in Fig. 3C, showing that the nomogram-predicted probabilities of BA were similar to the actual probabilities of BA, indicating that the prediction was in good agreement with the actual observation, in terms of the probability of BA (Fig. 3C). These findings also suggested that the

discrimination ability of the nomogram for prediction of BA could be generalizable to the other populations and may be clinically applicable.

Furthermore, decision curve analysis (DCA) was applied to render clinical validity to the nomogram and ln(GGT) for diagnosis of BA. The results corroborated good clinical applicability of the nomogram and ln(GGT) in predicting BA, because the ranges of threshold probabilities were wide and practical (Fig. 3D). The specific standardized net benefit of the nomogram and ln(GGT) under different threshold probabilities were presented in Suppl. Table 1. DCA displayed a net benefit of 9.4% at 80% of the threshold probability, which was superior to ln(GGT) and 30.2% superior to the baseline model (Fig. 3D, Suppl. Table 1).

### 3.6. Performance Comparison of the Three Models for the Diagnosis of BA

After having successfully constructed the DT, RF, and nomogram for the diagnosis of BA, we made performance comparison of the three models. The nomogram demonstrated greater discriminative ability with the sensitivity of 85.7%, specificity of 80.3%, and PPV of 0.969 at the optimal cut-off value in contrast to the other two algorithms DT and RF, which displayed relatively similar high values of sensitivity



**Fig. 3.** Multivariate logistic regression-based Nomogram to predict the probability of BA. Nomogram for prediction of BA was created using the following five predictors: gender, weight, DB, ln(ALP), and ln(GGT). (A) The construction of the nomogram using gender, weight, DB, ln(ALP) and ln(GGT) as predictors. (B) Receive operating characteristic (ROC) plots. The area under ROC curve (AUC) value was 0.898 for the formulated nomogram for the diagnosis of BA. (C) The calibration curve for the prediction model. The nomogram-predicted probabilities of BA were similar to the actual probabilities of BA. (D) Decision curve analysis (DCA) of the prediction nomogram for BA, showing wide and practical ranges of threshold probabilities with a net benefit of 9.4% at 80% of threshold probability.

and PPV (0.941 for DT and 0.947 for RF), but low specificity in the modeling group. In this regard, the nomogram has more potential than the DT and RF in the clinical application. Moreover, the validation of the multivariate logistic regression-based nomogram showed high stability and reproducibility. It was of note that the three models displayed similarly high values of PPV (0.941 for DT, 0.947 for RF, and 0.95 for nomogram at the cut-off value set at 0.65), and therefore DT, RF and nomogram were able to identify patients with BA. Meanwhile, we noticed that the NPV values of the three models were not high, suggesting that the three models were not very helpful for the excluding diagnosis of BA.

We also compared the discriminative capacity between the nomogram and each risk predictor alone, particularly GGT. When the study subjects were stratified by GGT into two subgroups (GGT < 300 U/L and  $\geq$  300 U/L), the diagnosis sensitivity and specificity were 0 and 1 for GGT < 300 U/L, 0.960 and 0.200 for GGT  $\geq$  300 U/L, whereas those for the nomogram were 0.448 and 0.957 in the subgroup (GGT < 300 U/L), and 0.951 and 0.350 in the subgroup (GGT  $\geq$  300 U/L). In addition, the nomogram displayed the consistency of the performance between the modeling and validation sets in the nomogram. However, the sensitivity, specificity, and PPV values (0.786, 0.795, and 0.966) for GGT alone in the modeling group were not well reproduced in the validation group (0.885, 0.124, and 0.706, respectively). Therefore, the nomogram was superior to the GGT alone in the diagnosis of BA.

In summary, the discriminative capacities of the nomogram outperformed the DT and RF, as well as GGT alone, in the diagnosis of BA. Moreover, the specificity of the nomogram to identify the patients with BA was among the highest of all the three models developed. Due to the above reasons, we believe that the nomogram was the most appropriate to predict BA among infant patients with NOJ.

#### 4. Discussion

The accurate diagnosis of BA using the existing diagnostic approaches is challenging primarily due to the overlapping features between BA and the other forms of NC with different causes, also referred to as non-BA. Aside from that reason, the current diagnostic methods are costly, time-consuming, and highly invasive. As a serum biomarker, GGT has been used for the diagnosis of BA in newborn infants who have been suspected of suffering from neonatal cholestasis [21–24, 26, 27]. However, the reliability and reproducibility of GGT alone needs to be improved. The development of new diagnostic models using GGT, in combination with other BA-associated risk factors, has a potentially better capacity for distinguishing BA from non-BA, and therefore could be clinically significant. The present study, based on a large sample size of 1728 cases, has the following main novel findings: (1) levels of DB, ALP, and GGT were significantly higher in BA patients; (2) the AUC value for the multivariate logistic regression-based nomogram was greater than that for ln(GGT), ln(ALP), or DB alone in the prediction of BA; (3) the discriminatory ability was significantly improved when GGT was combined with additional risk predictors, including weight, gender, DB, and ALP; and (4) our results support that the nomogram established in this study had better performance, and therefore holds promise for clinical application for BA diagnosis.

Early detection and accurate diagnosis of BA has been critical for timely intervention with implementation of the Kasai operation to restore bile flow and slowdown the progression of this disease in newborn infants [13, 18, 19, 30]. The current preoperative approaches for the diagnosis of BA primarily include several medical imaging techniques, such as ultrasound imaging of the liver, cross-sectional magnetic resonance imaging (MRI), and cholangiopancreatography (MRCP). The duodenal tube test (DTT) and liver biopsy [6, 19, 29, 30] are two other preoperative screening techniques. The existing diagnostic methods appear to have a number of limitations, either being costly, time consuming, technically difficult, or highly invasive.

Recent progresses have been made in the development of noninvasive serum biomarkers for diagnosing BA, of which GGT has been extensively investigated and verified. Multiple studies, including ours, have demonstrated that the GGT levels are higher in patients with BA than in non-BA controls, and the reliability of GGT was age-dependent [21, 23, 24]. We recently demonstrated that GGT levels were significantly greater in younger infants with BA (age < 30 days) than the older patients [21]. In the same study, the diagnostic value of GGT levels was highest among infant patients aged 61–90 days with a sensitivity of 82.8% and specificity of 81.6% in the discrimination of BA from non-BA cases [21]. Until now, limited studies have been performed on GGT levels coupled with additional risk factors. To date, few studies have been carried out to establish a diagnostic model using both non-invasive markers and other risk factors. In the present study, we identified a number of risk factors which were significantly different between BA and non-BA patients, including weight, age, DB, ALP, and GGT. Higher expression of GGT was detected in the BA patients as compared to non-BA patients. When all five risk factors were considered in the development of the multivariate logistic regression-based nomogram, the discrimination ability and diagnostic value were improved. Furthermore, the nomogram established in this study has turned out to be feasible and accurate for the diagnosis of BA, and thereby has potential for clinical application.

It has been reported that GGT > 300 U/L or a daily increase of 6 U/L can be used to differentiate BA from neonatal hepatitis with an accuracy of 85% and 88%, respectively [21]. However, a portion of infant patients eventually diagnosed as BA by intraoperative cholangiography had GGT < 300 U/L, which has posed a special challenge for the preoperative differential diagnosis of BA. In this regard, the present established nomogram, exhibiting better value in the preoperative diagnosis of BA among infants with GGT < 300 U/L, could improve the clinical diagnosis rate of BA and reduce the false positive rate, which apparently merits attention.

While the present study has offered useful information about the value of the nomogram from the diagnosis of BA, it has a number of limitations that must be acknowledged. First, the nomogram was established based on a single-center cohort study. However, our center is the largest treatment center for BA nationwide. Secondly, the study was conducted retrospectively, and selection bias might exist. Thirdly, this nomogram is only based on regular clinical characteristics and liver function, while other biomarkers were not assessed. Fourthly, this nomogram may not confirm the precise diagnosis for early detection in the atypical course. Lastly, the sensitivity and specificity of the nomogram may be further improved in the future.

Most recently, we investigated circulating microRNAs (miRNAs) using serum miRNA microarray analysis and identified miR-4429 and miR-4689 as potential biomarkers for the diagnosis of BA [31]. In our previous study, we found that AUC of miR-4429 was 0.789 with sensitivity of 83.3% and specificity of 80.0%, while the AUC of miR-4689 was 0.722 with sensitivity of 66.7% and specificity of 80.0% for the prediction of BA, suggesting that circulating miR-4429 and miR-4689 may play a role in the diagnosis of BA [31]. Our studies have shown that GGT, combined with other factors in the nomogram, significantly improved the discriminatory ability in the diagnosis of BA. It is worthwhile carrying out further studies to find better and more effective serum marker combinations, such as miRNA-4429 and miR-4689 [31], for the early diagnosis of BA among the patients with NOJ. Zahm et al. revealed that the AUC value of miR-200b/429 was > 0.80, suggesting promising diagnostic performance for BA [32]. Wang and colleagues reviewed 38 independent studies investigating the early differential diagnostic methods of BA in patients with infantile cholestasis, and reported that the sensitivities and specificities varied, ranging from 77% to 93% and 84% to 97%, respectively [33]. However, there are no commercially available standard kits for examination of the above small molecular markers.

To date, there are still few effective markers that have been evaluated in clinical practice, except for GGT. Most recently, Kim and colleagues established a MRI-based DT model for the diagnosis of BA among infants with jaundice and reported high performance with sensitivity of 97.3%, specificity of 94.8%, and accuracy of 96.2% [34]. The present nomogram, which was generated based upon serum markers, particularly GGT, had relatively lower sensitivity (85.7%) and specificity (80.3%). However, infants are usually uncooperative and therefore need procedural pediatric sedation or a general anesthetic (GA) to undergo MRI examination. In addition, laboratory tests are less costly than MRI. Thus, the combination of the nomogram with MRI-based DT for the diagnosis of BA in infants may need further investigation. Additionally, the integration of features from US and MRI findings may be used to establish a more effective model for diagnosing BA, and such a study is underway in our center.

Taken together, our study has successfully established novel models, including DT, RF, and multivariate logistic regression-based nomogram, to prioritize patients and to diagnose BA among patients suffering from NOJ. Of these, the multivariate logistic regression-based nomogram has demonstrated better performance for the prediction of BA and holds promise for clinical application in the diagnosis of BA.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ebiom.2018.07.025>.

### Funding Sources

This study received financial support from Shanghai Hospital Development Center (SHDC12014106), Shanghai Key Disciplines (no.2017ZZ02022), the National Natural Science Foundation of China (no. 81370472, no. 81770519, no. 81771633, no. 81401243 and no. 81500394), Shanghai Rising-Star Program (A type) (no. 15QA1400800), the Science Foundation of Shanghai Excellent Youth Scholars (no. 2017YQ042), and the Science Foundation of Shanghai (no. 16411952200, no. 16140902300 and no. 17411960600).

### Author Contributions

Shan Zheng and Yijie Zheng conceived and designed the study. Rui Dong, Jingying Jiang and Shouhua Zhang analysis the data and wrote the paper. Zhen Shen, Gong Chen and Yanlei Huang reviewed and edited the manuscript. All authors read and approved the manuscript.

### Conflict of Interest

Yijie Zheng is an employee of Abbott Diagnostics.

### References

- [1] Mack, C.L., 2015]. What causes biliary atresia? Unique aspects of the neonatal immune system provide clues to disease pathogenesis. *Cell Mol Gastroenterol Hepatol* 1 (3), 267–274.
- [2] Mack, C.L., Feldman, A.G., Sokol, R.J., 2012]. Clues to the etiology of bile duct injury in biliary atresia. *Semin Liver Dis* 32 (4), 307–316.
- [3] Chiu, C.Y., Chen, P.H., Chan, C.F., Chang, M.H., Wu, T.C., Taiwan Infant Stool Color Card Study G, 2013]. Biliary atresia in preterm infants in Taiwan: a nationwide survey. *J Pediatr* 163 (1), 100–103 [e101].
- [4] Alagille, D., 1984]. Extrahepatic biliary atresia. *Hepatology* 4 (1 Suppl), 7S–10S.
- [5] Bassett, M.D., Murray, K.F., 2008]. Biliary atresia: recent progress. *J Clin Gastroenterol* 42 (6), 720–729.

- [6] Chardot, C., 2006]. Biliary atresia. *Orphanet J Rare Dis* 1, 28.
- [7] Sokol, R.J., Shepherd, R.W., Superina, R., Bezerra, J.A., Robuck, P., Hoofnagle, J.H., 2007]. Screening and outcomes in biliary atresia: summary of a National Institutes of Health workshop. *Hepatology* 46 (2), 566–581.
- [8] Feldman, A.G., Mack, C.L., 2015]. Biliary atresia: clinical lessons learned. *J Pediatr Gastroenterol Nutr* 61 (2), 167–175.
- [9] Dong, R., Song, Z., Chen, G., Zheng, S., Xiao, X.M., 2013]. Improved outcome of biliary atresia with postoperative high-dose steroid. *Gastroenterol Res Pract* 2013, 902431.
- [10] Lai, H.S., Chen, W.J., Chen, C.C., Hung, W.T., Chang, M.H., 2006]. Long-term prognosis and factors affecting biliary atresia from experience over a 25 year period. *Chang Gung Med J* 29 (3), 234–239.
- [11] de Vries, W., Homan-Van der Veen, J., Hulscher, J.B., et al., 2011]. Twenty-year transplant-free survival rate among patients with biliary atresia. *Clin Gastroenterol Hepatol* 9 (12), 1086–1091.
- [12] Shinkai, M., Ohhama, Y., Take, H., et al., 2009]. Long-term outcome of children with biliary atresia who were not transplanted after the Kasai operation: >20-year experience at a children's hospital. *J Pediatr Gastroenterol Nutr* 48 (4), 443–450.
- [13] Chen, G., Zheng, S., Sun, S., et al., 2012]. Early surgical outcomes and pathological scoring values of older infants ( $\geq 90$  d old) with biliary atresia. *J Pediatr Surg* 47 (12), 2184–2188.
- [14] Sira, M.M., Taha, M., Sira, A.M., 2014]. Common misdiagnoses of biliary atresia. *Eur J Gastroenterol Hepatol* 26 (11), 1300–1305.
- [15] Serinet, M.O., Broue, P., Jacquemin, E., et al., 2006]. Management of patients with biliary atresia in France: results of a decentralized policy 1986–2002. *Hepatology* 44 (1), 75–84.
- [16] Sun, S., Chen, G., Zheng, S., et al., 2013]. Analysis of clinical parameters that contribute to the misdiagnosis of biliary atresia. *J Pediatr Surg* 48 (7), 1490–1494.
- [17] Boskovic, A., Kitic, I., Prokic, D., Stankovic, I., Grujic, B., 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 (2), 170–174.
- [18] Chardot, C., Serinet, M.O., 2006]. Prognosis of biliary atresia: what can be further improved? *J Pediatr* 148 (4), 432–435.
- [19] Chardot, C., Carton, M., Spire-Bendelac, N., Le Pommelet, C., Golmard, J.L., Auvert, B., 1999]. Prognosis of biliary atresia in the era of liver transplantation: French national study from 1986 to 1996. *Hepatology* 30 (3), 606–611.
- [20] Jiang, L.P., Chen, Y.C., Ding, L., et al., 2013]. The diagnostic value of high-frequency ultrasonography in biliary atresia. *Hepatobiliary Pancreat Dis Int* 12 (4), 415–422.
- [21] Chen, X., Dong, R., Shen, Z., Yan, W., Zheng, S., 2016]. Value of gamma-glutamyl transpeptidase for diagnosis of biliary atresia by correlation with age. *J Pediatr Gastroenterol Nutr* 63 (3), 370–373.
- [22] Liu, C.S., Chin, T.W., Wei, C.F., 1998]. Value of gamma-glutamyl transpeptidase for early diagnosis of biliary atresia. *Zhonghua Yi Xue Za Zhi (Taipei)* 61 (12), 716–720.
- [23] Cabrera-Abreu, J.C., Green, A., 2002]. Gamma-glutamyltransferase: value of its measurement in paediatrics. *Ann Clin Biochem* 39 (Pt 1), 22–25.
- [24] Rendon-Macias, M.E., Villasis-Keever, M.A., Castaneda-Mucino, G., Sandoval-Mex, A.M., 2008]. Improvement in accuracy of gamma-glutamyl transferase for differential diagnosis of biliary atresia by correlation with age. *Turk J Pediatr* 50 (3), 253–259.
- [25] Tang, K.S., Huang, L.T., Huang, Y.H., et al., 2007]. Gamma-glutamyl transferase in the diagnosis of biliary atresia. *Acta Paediatr Taiwan* 48 (4), 196–200.
- [26] Maggiore, G., Bernard, O., Hadchouel, M., Lemonnier, A., Alagille, D., 1991]. Diagnostic value of serum gamma-glutamyl transpeptidase activity in liver diseases in children. *J Pediatr Gastroenterol Nutr* 12 (1), 21–26.
- [27] Wang, H., Malone, J.P., Gilmore, P.E., et al., 2010]. Serum markers may distinguish biliary atresia from other forms of neonatal cholestasis. *J Pediatr Gastroenterol Nutr* 50 (4), 411–416.
- [28] Ceriotti, F., 2017]. Quality specifications for the extra-analytical phase of laboratory testing: reference intervals and decision limits. *Clin Biochem* 50 (10–11), 595–598.
- [29] El-Guindi, M.A., Sira, M.M., Sira, A.M., et al., 2014]. Design and validation of a diagnostic score for biliary atresia. *J Hepatol* 61 (1), 116–123.
- [30] Zagory, J.A., Nguyen, M.V., Wang, K.S., 2015]. Recent advances in the pathogenesis and management of biliary atresia. *Curr Opin Pediatr* 27 (3), 389–394.
- [31] Dong, R., Shen, Z., Zheng, C., Chen, G., Zheng, S., 2016]. Serum microRNA microarray analysis identifies miR-4429 and miR-4689 are potential diagnostic biomarkers for biliary atresia. *Sci Rep* 6, 21084.
- [32] Zahm, A.M., Hand, N.J., Boateng, L.A., Friedman, J.R., 2012]. Circulating microRNA is a biomarker of biliary atresia. *J Pediatr Gastroenterol Nutr* 55 (4), 366–369.
- [33] Wang, L., Yang, Y., Chen, Y., Zhan, J., 2018]. Early differential diagnosis methods of biliary atresia: a meta-analysis. *Pediatr Surg Int* 34 (4), 1–18.
- [34] Kim, Y.H., Kim, M.J., Shin, H.J., et al., 2018]. MRI-based decision tree model for diagnosis of biliary atresia. *Eur Radiol* 28 (8), 1–10.