



Clinical characteristics of neonatal cholestasis in a tertiary hospital and the development of a novel prediction model for mortality

Ho Jung Choi,^{a,1} Inki Kim,^{b,1} Hye-Jin Lee,^a Hyun Ju Oh,^a Mi Kyoung Ahn,^a Woo Im Baek,^a Yeong Eun Kim,^a Seak Hee Oh,^{a,*} Byong Sop Lee,^a Jung-Man Namgoong,^c Dae Yeon Kim,^d Eun Joo Lee,^d Jung Ok Shim,^e Jae Sung Ko,^d and Kyung Mo Kim^a

^aDepartment of Pediatrics, Asan Medical Center Children's Hospital, University Ulsan College of Medicine, 88, Olympic-ro 43-gil Songpa-Gu, Seoul 05505, South Korea

^bDepartment of Convergence Medicine, Asan Medical Center, Asan Institutes for Life Sciences, University of Ulsan College of Medicine, Seoul, South Korea

^cDivision of Pediatric Surgery, Department of Surgery, Asan Medical Center, University Ulsan College of Medicine, Seoul, South Korea

^dDepartment of Pediatrics, Seoul National University Children's Hospital, Seoul National University College of Medicine, South Korea

^eDepartment of Pediatrics, Korea University College of Medicine, Seoul, South Korea

Summary

Background Few studies have described the aetiologies of neonatal cholestasis, and the overall neonatal cholestasis-related mortality (NCM) rate is unclear. We investigated the aetiology and outcome of neonatal cholestasis in a tertiary hospital and developed an NCM prediction model for these patients.

Methods Patients aged <100 days with serum direct bilirubin (DB) levels of >1.0 mg/dL were retrospectively screened. Diagnostic and laboratory data during the 8-week follow-up period after enrolment between 2005 and 2020 were extracted digitally, and medical charts were reviewed manually by clinicians. Logistic regression was used to derive a prediction model for the 1-year mortality outcome of neonatal cholestasis, and performance evaluation and external validation were conducted for the NCM prediction model.

Findings We enrolled 4028 neonates with DB of >1.0 mg/dL at least once. Prematurity and birth injury (35.4%), complex heart anomalies (18.6%), liver diseases (11.4%), and gastrointestinal anomalies (9.2%) were the most common aetiologies; 398 (9.9%) patients died before one year of age. The peak value of DB was positively correlated to the 1-year mortality rate. In the multivariate analysis, simple laboratory indices, including platelet, prothrombin time, aspartate aminotransferase, albumin, direct bilirubin, creatinine, and C-reactive protein, were independent predictors of 1-year mortality outcome of complete-case subjects. Using these laboratory indices, a logistic regression-based NCM prediction model was constructed. It showed acceptable performances on discrimination (area under the curve, 0.916), calibration (slope, 1.04) and Brier scoring (0.072). The external validation of the sample ($n = 920$) from two other centres also revealed similar performance profiles of the NCM model.

Interpretation Various aetiologies of neonatal cholestasis were identified in a tertiary hospital, resulting in unfavourable outcomes of a large proportion. The NCM prediction model may have the potential to help clinicians to be aware of high-risk neonatal cholestasis.

Copyright © 2022 The Author(s). 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/>)

Keywords: Neonatal cholestasis; Aetiology; Mortality; Prediction model

Introduction

Neonatal cholestasis is defined as reduced bile formation or flow resulting from hepatobiliary dysfunction or obstructive lesions of the bile duct in the neonatal period.¹ Neonatal cholestasis is screened by the

*Corresponding author.

E-mail address: seakhee.oh@amc.seoul.kr (S.H. Oh).

¹ These authors contributed equally to this work.

Research in context

Evidence before this study

Aetiologies of neonatal cholestasis were described in few studies from liver centres with small numbers of subjects. Risks of mortality in neonatal cholestasis have been assessed by a limited dataset of subgroups such as parenteral nutrition.

Added value of this study

This study of the largest cohort of neonatal cholestasis showed various aetiologies and combined medical conditions in detail. Using simple laboratory indices, a novel prediction model for 1-year mortality was developed and validated based on the TRIPOD guidelines.

Implications of all the available evidence

The results show that the prediction model has a potential for clinical application in identifying high-risk neonatal cholestasis.

identification of elevated direct (or conjugated) bilirubin (DB) and bile acids. Neonatal cholestasis occurs in approximately 1 out of every 2500 term neonates.^{2,3} In primary care, the management of neonatal cholestasis requires caution because its causes are known to be biliary atresia (BA) and monogenic liver disease in approximately 25–40% and 25% of cases, respectively.^{1–5} As delayed Kasai operation generally leads to a poor outcome of the native liver, the early detection of BA is crucial.⁴ Infections, galactosemia, tyrosinemia, panhypopituitarism, bile acid synthesis defect, and biliary sludges (or stones) also require prompt management. So, the correct diagnosis of neonatal cholestasis often requires extensive invasive and non-invasive evaluations.¹ Although the aetiologies are not discovered on differential diagnosis or treatments are not available for specific causes, nutritional optimisation and prevention of medical complications from this condition can be beneficial for the appropriate growth or nutritional stabilisation of the patient before liver transplantation. Despite the clinical importance of neonatal cholestasis, currently, a total of fewer than two thousand studies of various study designs are available in the literature on this subject.^{5–23}

There is also little information available on the outcome of neonatal cholestasis in the literature, and the little available information is generally determined and reported by each specific cause and available treatment.^{5–23} Approximately 50% of patients with BA lose their native livers within 2 years after a Kasai operation.⁴ Contrarily, idiopathic cholestasis, including that caused by idiopathic neonatal hepatitis, is often transient and benign; only 1% of patients with idiopathic cholestasis lost their native livers before the age of

2 years in a recent report from the Childhood Liver Disease Research Network study.⁵ Transient cholestasis is also often observed in infants with specific systemic conditions such as sepsis, ischaemic injury from birth asphyxia or congenital heart anomalies, haemolytic disease and defective enteral autonomy with prolonged parenteral nutrition.^{1–6} Premature infants cannot often afford oral nutrition and the risk of cholestasis doubles in cases of parenteral nutrition that lasts for more than 30 days.²⁴ Many single-centre studies have revealed a variety of epidemiologic characteristics and clinical outcomes in neonatal cholestasis, of which cohorts contained heterogeneous patient groups based on the kind of centre (liver-oriented centre vs. other specialised centres; referral centre vs. primary centres).^{5–23}

Predicting the outcome of neonatal cholestasis could be useful. By discriminating unfavourable outcomes such as in-hospital mortality among patients with neonatal cholestasis, clinicians are able to identify high-risk neonatal cholestasis cases and to prepare focused and multidisciplinary strategies for them. There are several mortality prediction models for liver diseases in paediatric hepatology, including the Pediatric End-Stage Liver Disease (PELD) score for chronic liver disease, the King's College Hospital criteria (KCHC) score, and the Liver Injury Units (LIU) for acute liver failure.^{25–27} The training dataset of these models originated from end-stage liver disease and fulminant hepatitis in paediatric groups with wide age ranges, mostly more than 6 months. In addition, the primary goal of these predictions is to estimate the risk-benefit of liver transplantation in affected infants and children. Currently, there is no prediction model for mortality in neonatal cholestasis using a big-enough dataset of neonatal cholestasis cases.

To make the neonatal cholestasis-related mortality (NCM) prediction model clinically useful, clear and sufficient information should be provided in the essential parts of model development-sample aetiologies, variable selection, model construction, full performance evaluation and external validation.^{28,29} Notably, not only accuracy in discrimination (i.e. receiver-operating characteristic curve analysis) but also precision in the calibration of the predicted probability must be included in the evaluation of the new model's performance.³⁰ In addition to reliable performance, a prediction model should provide clinicians insights into algorithms used for predicting mortality in neonatal cholestasis. This is because understanding the 'how and why' by clinicians, which is referred to as the interpretability of models in statistics, applies to medical decision-making in real-life clinical practice.³¹

In this study, we investigated the aetiological characteristics of neonatal cholestasis and its outcome in a tertiary hospital. In addition, using a large dataset of simple laboratory indices in this cohort, we developed an NCM prediction model for 1-year in-hospital mortality in neonates with cholestasis.

Methods

Study population

We conducted a retrospective study on cholestasis in a group of paediatric patients aged less than 100 days in the Seoul Asan Medical Center. Neonatal cholestasis was defined as serum DB >1.0 mg/dL according to a definition of the North American Society for Pediatric Gastroenterology, Hepatology and Nutrition and the European Society for Paediatric Gastroenterology, Hepatology, and Nutrition (NASPGHN/ESPGHN) guideline.¹ The inclusion criteria were as follows:¹ age <100 days (neonatal period),² DB >1.0 mg/dL (cholestasis) at least once and³ occurrence of a first DB >1.0 mg/dL between January 2005 and January 2020 (study period). Based on the study criteria, neonates ($n = 13,344$) who never had DB >1.0 mg/dL were excluded from the analysis. Diagnostic information in the electronic medical records was classified according to the International Classification of Disease (ICD) system.¹⁴ All electric medical records and their ICD codes were manually validated in detail by paediatric hepatologists. The diagnosis of BA was evidenced by intraoperative angiography of the bile duct. Other definitions of idiopathic neonatal hepatitis and idiopathic neonatal cholestasis were used according to the criteria of the Childhood Liver Disease Research Network.⁵ From enrolment to the 8th week of follow-up, simple laboratory indices, such as the white blood cell (WBC, $\times 10^3/\mu\text{L}$) and platelet (PLT, $\times 10^3/\mu\text{L}$) counts; haemoglobin level (Hb, g/dL); prothrombin time (PT INR, international normalized ratio), alanine aminotransferase level (ALT, g/dL), aspartate aminotransferase (AST, IU/L) level and albumin (ALB, g/dL), total bilirubin (TB, mg/dL), DB, creatinine (Cr, mg/dL), and C-reactive protein (CRP, mg/dL) levels, were analysed.

Literature search

To compare the outcomes of the participants of the present study, the Medline and EMBASE databases were searched to extract outcome data from the prospective or retrospective case series or cohort study with >20 subjects with cholestasis during the neonatal period. The total number of subjects, proportion of BA, liver transplantation rate, mortality rate and risk factors for mortality were reviewed. If information on patients' poor outcomes was not available in reported cohorts in any given study, the study was excluded. To evaluate the necessity of prediction model, we searched previous models for cholestatic paediatric patients and checked suggested questionnaires (Supplementary Table 1).³²

Statistical analysis

The outcome (or dependent) variable for this study was survival up to the age of 1 year [i.e.¹ survival group vs.²

mortality group]. Descriptive statistics for the characteristics listed above were provided for 1-year survival and 1-year mortality in subjects included in the model development. Continuous variables were expressed as mean values with standard deviations or median values with interquartile ranges (IQRs) for normally distributed and skewed data, respectively. Differences between variable groups were assessed using the two-sample *t*-test or the Mann-Whitney *U* test for continuous variables as appropriate. For categorical variables, the χ^2 test or Fisher's exact test (in cases where the number of patients involved was ≤ 5) was used. Receiver-operating characteristic (ROC) curve analysis with the area under the curve (AUC) was performed to identify optimal cut-off values of continuous variables. For the multivariate analysis, logistic regression (LR) was performed. As alternative analyses, machine learning was also performed to explore underlying data pattern. All statistical calculations were performed using IBM SPSS version 27.0 (SPSS Inc., Armonk, NY), R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria), and Python version 3.8.2 (Python Software Foundation, Wilmington), as appropriate. A *p*-value of <0.05 was considered statistically significant in statistical models.

Variable selection and missing values

The simple laboratory indices within 8 weeks of enrolment (DB >1.0 mg/dL) were selected as predictor candidates. Based on the fact that the paediatric prediction models for chronic liver disease and acute liver failure were adopted to use peak index values during the specific follow-up durations,^{25–27} not only the initial but also the peak values of each laboratory index were used as candidate variables evaluated in the modelling process. ROC curve analyses were performed to evaluate the discriminatory ability in univariate analysis of each variable and calculate empirical optimal cut-off values based on the Youden index.

Before conducting multiple imputation or regression imputation for missing values,^{33,34} missing values were classified into missing completely at random, missing at random, and missing not at random (MNAR). As MNAR is possibly associated with both independent and dependent variables,³⁵ the plausible mechanism of suspected MNAR was reviewed using descriptive statistics. Clinical judgement on subgroups of neonatal cholestasis may affect the selection of a range of serologic tests, such as PT INR. For example, clinicians may not conduct full laboratory evaluation for transient benign neonatal cholestasis^{20,23} because transient cholestasis without any underlying severe diseases shows favourable survival outcomes.³ As MNAR is still problematic in statistics,³⁵ we conducted both [1] complete-case analysis without imputation and [2] whole-case analysis with imputation.

Model development

The prediction model was developed following the recommended guideline of the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement and Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies (CHARMS) checklist.^{28,36} One of the primary aims of this study was to develop a prediction model for 1-year in-hospital mortality (binary dependent variable). Firstly, multivariate LR was performed to develop an NCM prediction model; binary LR with the forward method³⁷ was derived for the model 1-year in-hospital mortality risk and resulted in regression β coefficients and intercept of the appropriate laboratory values ($\text{logit}[\text{probability}] = \beta_0 + \beta_1 \times 1 + \beta_2 \times 2 + \dots + \beta_n X_n$). Multicollinearity among independent variables was examined before LR, and parameters with correlation coefficients of ≥ 0.85 were removed from the LR analysis. Using logit (probability), the predicted probability of the 1-year mortality event was calculated.

For alternative analyses, decision trees including CART and CHAID modelling,³⁸ Random forests (RFs)³⁹ and extreme gradient boost (XGBoost)⁴⁰ were also performed (Supplementary Method 1 in detail). For machine learning analysis, overall variable importance and Shapley additive explanations (SHAP) value were provided.^{41,42} To promote the interpretability of XGBoost, a SHAP plot was also used to show the complex relationship between independent and dependent variables.

Performance evaluation

To evaluate overall performance, the Brier score (value = $\frac{1}{N} \sum_{k=1}^n [\text{predicted probability}_k - \text{observed event}_k]^2$), which is a metric concept similar to the mean square error, was used.⁴³ Moreover, Nagelkerke's R^2 and the Hosmer-Lemeshow test were used as measures of the overall performance.⁴⁴ A prediction density plot was used to visualise the discriminatory trends of each model. ROC curve analysis was used to evaluate the discrimination of the models and the c-index, which is equivalent to the AUC in the ROC curve analysis in the binary outcome, was a measure of the discriminatory ability. The calibration curve was also used to test the precision of the models, and the calibration slope and intercept were calculated using Hosmer-Lemeshow contingency tables. The sensitivity, specificity, positive predictive value and negative predictive value were used to evaluate the 'clinical usefulness' of models.

Internal and external validation

Both the internal and external validations of the models were performed. Internal validation for the reproducibility of prediction models for the originated dataset was

performed through 10-fold cross-validation and bootstrap re-sampling.^{45,46} Based on the TRIPOD statement,²⁸ the mean difference between the 200 bootstrapping re-samples is defined as the optimism. Estimating the optimism (or bootstrap)-corrected AUC and its confidence intervals (CIs) of the final NCM prediction model was conducted using the internal dataset. The dataset of neonatal cholestasis at two other large tertiary hospitals (Seoul National University Hospital and Korea University Hospital) in South Korea were used for external validation. Using the same modelling frames, performance on the external dataset was quantified using the Brier score, AUC and calibration. In addition, clinical usefulness was determined in the same manner.

Ethical approval

The study was approved by the Institutional Review Board of the Asan Medical Center (#2020-0202), 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.

Role of funding source

The funder had no role in the study design, data collection, data analyses, interpretation, writing of the report, and the decision of paper submission.

Results

Characteristics of neonatal cholestasis

During the study period, 17,372 neonates aged less than 100 days underwent 128,822 serologic DB tests (Supplementary Figure 1). Among them, 4028 neonates had DB of at least ≥ 1 -fold of the DB threshold of >1.0 mg/dL, defined as neonatal cholestasis in the NASPGHN/ESPGHN guideline,¹ with a median of 1.3 mg/dL (IQR: 1.1–1.6 mg/dL). The mean age of these patients at the time of enrolment was 10.2 ± 19.7 days with a median gestational age of 35 weeks (IQR: 32–38 weeks) and median birth weight of 2.36 kg (IQR: 1.5–3.0 kg). There were 2322 (57.6%) male patients. A variety of combined medical conditions were noted in the classification of the ICD code and review of medical charts (2.6 ICD codes per patient). Among them, the primary diagnoses, directly linked to the development of neonatal cholestasis, are listed in Table 1. Birth issues with prematurity and injury (35.4%), congenital heart anomaly with other cardiac issues (18.6%), liver diseases (11.4%), gastrointestinal anomaly and related issues (9.2%), neonatal unconjugated jaundice with DB $<20\%$ of TB (7.4%), respiratory anomaly and other issues (5.5%) and genetic/chromosomal disorders (4.3%) were the most common primary diagnoses in the whole-case dataset.

Primary diagnosis for neonatal cholestasis	Dataset	
	Complete-case <i>n</i> = 2661 (%)	Whole-case <i>n</i> = 4028 (%)
Prematurity and birth-related issues	654 (24.6%)	1426 (35.4%)
Congenital heart anomaly and other problems	663 (24.9%)	751 (18.6%)
Liver origin diseases including biliary atresia and hepatitis	416 (15.6%)	461 (11.4%)
Gastrointestinal anomaly and other problems	345 (13%)	369 (9.2%)
Neonatal jaundice in healthy newborns	58 (2.2%)	299 (7.4%)
Respiratory anomaly and related issues	178 (6.7%)	222 (5.5%)
Genetic/chromosomal disorders	141 (5.3%)	174 (4.3%)
Renal anomalies and urinary tract issues	55 (2.1%)	81 (2.0%)
Malignancy and tumour	57 (2.1%)	59 (1.5%)
Brain anomalies and other neurologic issues	32 (1.2%)	48 (1.2%)
Infection	15 (0.6%)	43 (1.1%)
Haematologic disorders	16 (0.6%)	38 (0.9%)
Hypopituitarism/hypothyroidism	9 (0.3%)	27 (0.7%)
Others	22 (0.8%)	30 (0.7%)

Table 1: Primary diagnosis of neonatal cholestasis.

Among diseases of hepatic origin, BA and neonatal hepatitis were the most common. Neonatal unconjugated jaundice also had the transient condition of DB >1.0 mg/dL, which is a definition of the study. In addition, a variety of genetic disorders and other specific diagnoses combined with neonatal cholestasis, which were not highlighted in the literature,^{5–23} are listed in Supplementary Table 2.

Outcomes of neonatal cholestasis

Among the 4028 neonates with cholestasis, 425 (10.6%) patients died during the whole follow-up period and 398 (9.9%) patients died in the first year of their lives (Figure 1a). The median age at the time of death was 55 days (IQR: 23–128 days) with a sex ratio (male/female) of 1.3 (Figure 1b). Causes of 1-year mortality are listed in Supplementary Table 3; complex heart anomalies, prematurity-related condition, birth injury, foetal hydrops and congenital diaphragmatic hernia were the most common causes of mortality. Among survivors, 54 (1.3%) underwent liver transplantation, with BA being the most common indication (*n* = 46/54, 85%). Notably, transient neonatal cholestasis (or early resolved cholestasis), including idiopathic neonatal hepatitis, idiopathic neonatal cholestasis, and neonatal jaundice with DB >20% of TB, had no mortality-related issues. Among the neonates included in this study, 148 neonates with BA were initially diagnosed or referred to a tertiary hospital without Kasai operation. Among them, 145 (3.6%) patients had Kasai operation, while three patients with no Kasai operation went to liver transplantation later. In BA, one patient died due to sepsis caused by cholangitis and another due to heart failure caused by combined cardiac anomalies. In the literature search,^{5–23} 18 reports (*n* = 2225) were included

(Supplementary Table 4), and the death rate was 10.8% (*n* = 240), which was similar to that found in the current study, although etiologic proportions were different.

Variable selection and plausible mechanism of the loss of missing variables

For comparison of simple laboratory indices between the 1-year mortality group and the survival group, the initial dataset at the time of enrolment and peak value dataset during the 8-week follow-up were selected for analysis (Supplementary Table 5). All the laboratory indices showed significant differences between survival and death groups in the univariate analysis. Among them, the peak value dataset had more variables of higher discriminatory ability in the ROC curve analysis. In addition, based on the fact that current liver-oriented prediction models, such as KCHC, PELD and LIU, use peak values of variables during a certain period in their prediction algorithms,^{25–27} we used the peak values of simple laboratory indices in the development of a new NCM prediction model. Notably, peak DB value was positively correlated to observed mortality rate (Supplementary Figure 2). In multicollinearity analysis, the Spearman correlation coefficient between AST and ALT was 0.9 (Supplementary Figure 3), and ALT was then removed from the LR analysis. The sample size of this study was appropriate to satisfy the minimum criteria of events per selected variables.⁴⁷

As the peak variable dataset was selected, the missing data mechanism was also evaluated; a large proportion of missing values in PT INR and CRP were possibly MNAR (Supplementary Table 6). Among the 4028 patients, the proportion of missing PT INR values was 34% (*n* = 1368). Groups with missing PT INR had only 2.3% (*n* = 32) mortality compared to the 13.7%

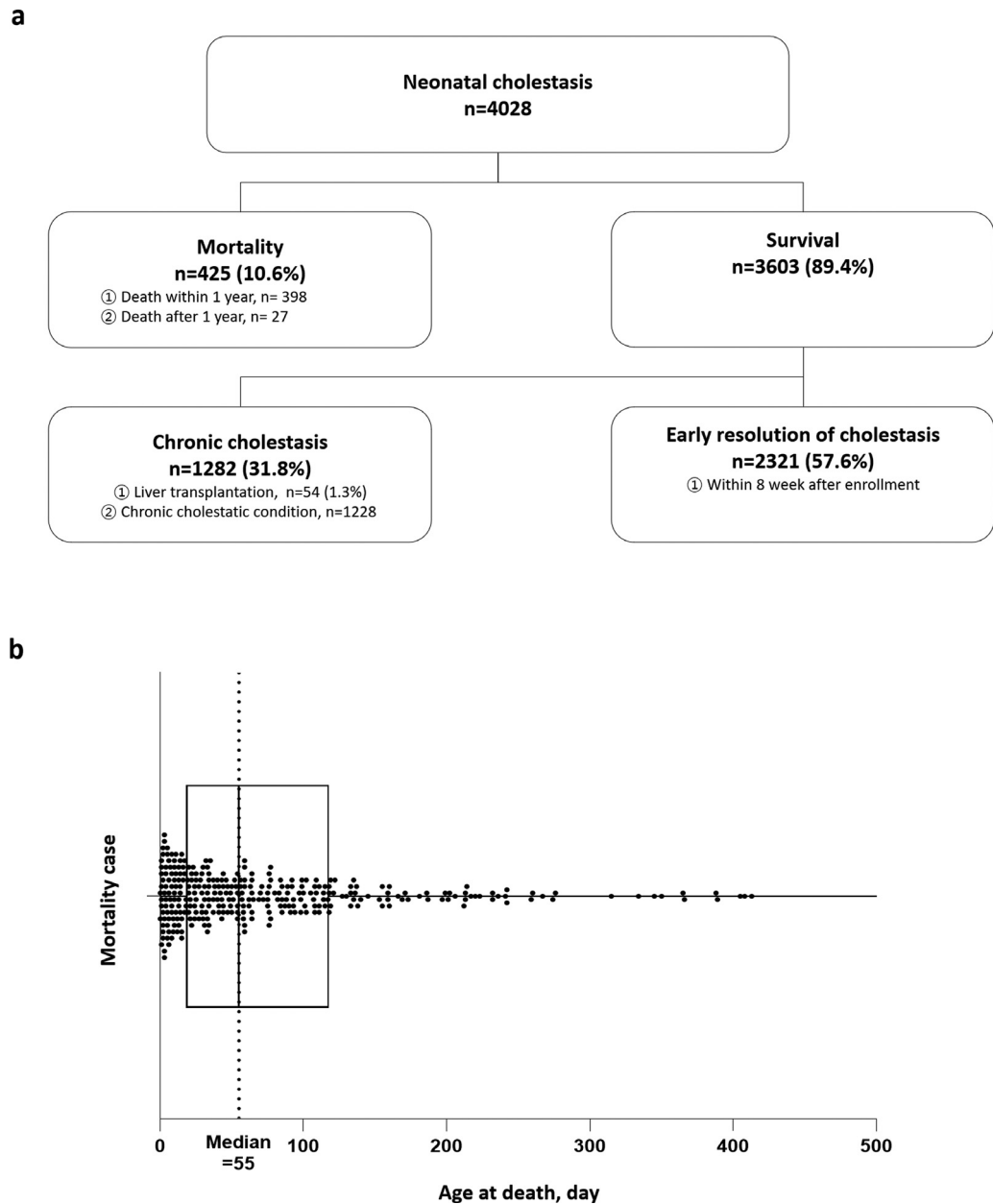


Figure 1. Outcome of neonatal cholestasis. a. The loss of native liver occurred in 12% of patients. b. Time-lapse from birth to demise among 425 mortality cases.

($n = 366$) mortality of the group with PT INR test values. This absence of PT INR values was also associated with higher values of DB, AST, Cr, and CRP (Supplementary Figure 4), suggesting that PT INR may be a cliché of problematic MNAR impacting both independent and dependent variables. As there is currently no established strategy for problematic MNAR,³⁵ we then performed separate analyses for both [1] complete-case analysis ($n = 2661$) without missing PT INR values as a main reference dataset and [2] whole-case analysis

($n = 4028$) with regression imputation as an alternative dataset (Supplementary Figure 5).

Interpretable prediction model development

Multivariate LR was conducted first for complete-case analysis using the forward method to predict 1-year mortality (Supplementary Table 7). The significance of the LR model was assessed by $\chi^2=843.0$ (p -value <0.001) and the Hosmer–Lemeshow test revealed insufficient

goodness-of-fit of the LR model (p -value < 0.05) with a low Nagelkerke's R^2 of 0.54. The logit (p) and its predicted probability were calculated using the following equation:

$$\begin{aligned} \text{Logit}(p) = & -0.016 \times \text{low PLT}(10^3/\mu\text{L}) + 0.095 \times \text{peak PT INR} \\ & + 0.000065 \times \text{peak AST(IU/L)} \\ & - 0.748 \times \text{low ALB(g/dL)} + 0.051 \times \text{peak DB(mg/dL)} \\ & + 0.38 \times \text{peak Cr(mg/dL)} \\ & + 0.027 \times \text{peak CRP(mg/dL)} - 0.382 \end{aligned}$$

$$\text{Probability}(\%) = \frac{I}{I + e^{-\text{Logit}(p)}} \times 100$$

Alternative analysis using machine learning

The odds ratios in this LR analysis cannot reflect the relative importance of continuous variables in the LR model because the scale of each variable is different. Therefore, we performed decision tree analyses for more interpretable models. CART and CHAID were used to visualise classification algorithms to distinguish the death group from the survival group (Figure 2a and Supplementary Figure 6). In both CART and CHAID analyses, initial segregations consisted of subgroups by different values of PLT. These findings suggested that the lowest values of PLT in this dataset are hierarchically higher in the prediction of mortality than variables directly related to hepatic conditions. Based on the fact that multiple illnesses such as sepsis, birth asphyxia, prematurity, liver disease, respiratory distress, and disseminated intravascular coagulation can be combined with thrombocytopenia in neonates,⁴⁸ the severity of thrombocytopenia was probably reflected by the degree of systemic illness, which may affect survival in

patients with neonatal cholestasis. Next, segregations were guided by thresholds of PT INR and AST in the CART model (Figure 2a) and by thresholds of PT INR, AST, ALB and DB in the CHAID model (Supplementary Figure 6). PT INR, AST, ALB and DB are, in general, indicators of liver pathologies.^{25–27}

We also applied flexible machine learning (RF and XGBoost) to explore whether other data patterns between predictors and outcomes were hidden. Complete-case and whole-case datasets were split into the train set (70%) and the test set (30%) in RF and XGBoost (Supplementary Figure 7). Through grid search, hyperparameters were optimised (Supplementary Table 8); otherwise, default settings were applied to both models. The result of variable importance in RF and XGBoost had similar findings of LR and decision tree models (Supplementary Figure 8), with PLT, PT INR, Cr, AST and DB being the most important in RF's complete-case and whole-case analyses. In XGBoost, variable importance, presented as the F score, showed a similar hierarchical order of variables to that shown in RF. In a SHAP summary plot, as a patient's platelet count decreases (approaches blue), the SHAP value impact on 1-year mortality increases (Figure 2b), and as patients' PT INR, CRP, AST and DB increase, the SHAP value impacts also increase. Overall, machine learning modelling agreed with the findings of LR (Figure 2c).

Performance evaluation

Three measures were used to evaluate the apparent performance of the NCM model and other alternative

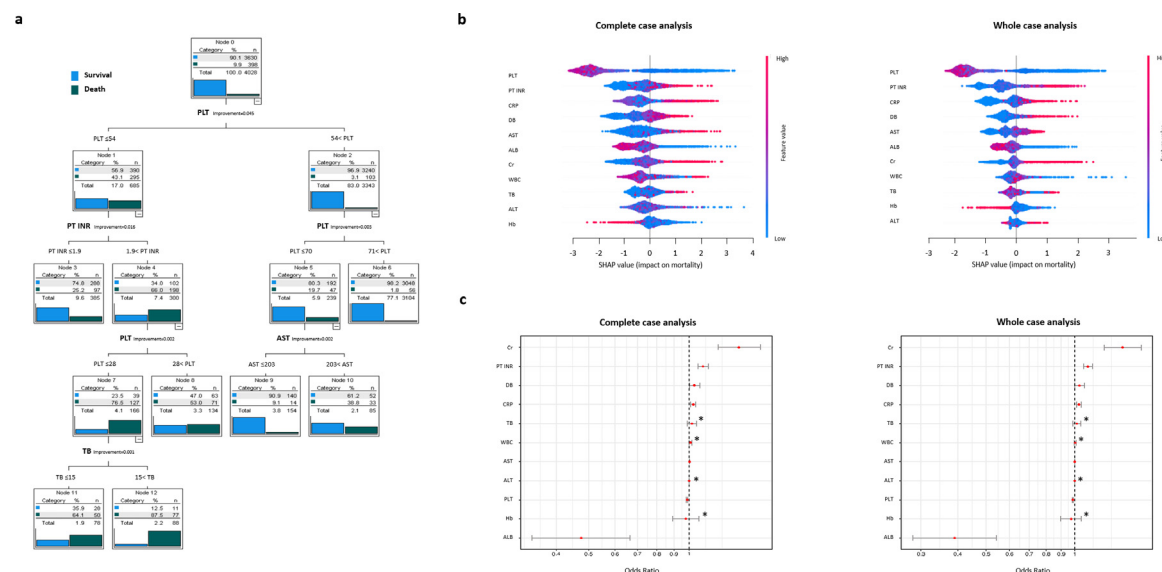


Figure 2. Exploratory analyses on complete-case and whole-case development dataset. a. Example of decision trees to classify 1-year mortality case in whole-case dataset without imputation: CART analysis. b. SHAP plot for XGBoost model; complete-case analysis vs. whole-case analysis with regression imputation. c. Odds ratio plot in LR model. Asterisks indicate no statistical significance in LR.

Prediction model	Dataset	Performance			Clinical usefulness				
		Brier score	AUC	Calibration slope	Threshold	Sensitivity	Specificity	PPV	NPV
Logistic regression	Complete-case	0.072	0.916	1.04	16.1%	86.9%	84.6%	47.3%	97.6%
	Whole-case	0.052	0.937	1.05	13.7%	86.9%	88.4%	45.1%	98.4%
Alternative analyses									
CART*	Complete-case	0.071	0.89	1.01	9%	87.7%	78.6%	39.5%	98.6%
	Whole-case*	0.052	0.89	1.02	9%	82.4%	87.8%	42.6%	97.8%
CHAID*	Complete-case	0.071	0.916	1.06	14%	87.4%	81%	42.4%	97.6%
	Whole-case*	0.052	0.936	1.08	7%	94.2%	77.8%	31.7%	99.2%
Random forest**	Complete-case	0.067	0.921	1.03	18.9%	84%	87.8%	50.9%	97.3%
	Whole-case	0.046	0.954	1.09	14.5%	89%	90.3%	49.8%	98.6%
XGBoost**	Complete-case	0.074	0.919	0.82	5.6%	84.9%	85.6%	47.4%	97.4%
	Whole-case	0.049	0.951	0.93	11%	86.6%	90.9%	51%	98.4%

Table 2: Performance and clinical usefulness of neonatal cholestasis-related mortality (NCM) prediction model. The NCM model was derived from complete-case dataset by logistic regression analysis.

* Missing values were not imputed

** Values of performances were derived from the test dataset (See Supplementary Figure 7). AUC; area under the curve, CART; classification and regression tree, CHAID; chi-square automatic interaction detection, ML; machine learning PPV; positive predictive value, NPV; negative predictive value, XGBoost; extreme gradient boost.

machine learnings; these were the Brier score, discrimination by the AUC value, and calibration. Table 2 suggests that the overall judgement of the performance of LR-based models is acceptable compared to that of other models (Figure 3a and Supplementary Figs. 9 to 13). The AUCs of the LR-based models were 0.916 for the complete-case analysis and 0.937 for the whole-case analysis with regression imputations. The calibration slopes were 1.04 for complete-case analysis and 1.05 for whole-case analysis. The discriminatory abilities were high in RF and XGBoost for the complete-case and whole-case analyses. Furthermore, the differences between complete-cases and whole-case analyses were minimal within analyses. Overall, LR-based NCM prediction model from complete-case dataset showed acceptable performance in predicting the 1-year mortality of neonatal cholestasis.

Subgroup analyses and comparison of other paediatric models

We performed subgroup analysis according to the major aetiologies, such as [1] prematurity with very low birth weight [2], complex heart anomalies, and [3] gastrointestinal anomalies with related issues. The discriminatory abilities of the NCM model were acceptable with AUCs ranging from 0.89 to 0.93 (Supplementary Figure 14), suggesting that the apparent discriminatory performance of the NCM model were applicable to the subgroups.

In addition, we compared our NCM model with existing scoring models (KCHC and PELD) that have been widely used for acute liver failure and chronic liver

disease in the older infant and childhood age groups. The AUCs of the NCM model and PELD score (0.863–0.937) were much higher than those of the KCHC (0.607–0.606) in both complete-case and whole-case analyses (Supplementary Figure 15). Overall, the NCM model showed a significantly higher discriminatory performance compared to the PELD and KCHC models (Supplementary Table 9).

Using another criterion of DB ≥2.0 mg/dL in the Childhood Liver Disease Research Network,⁵ 1608 (40%) patients were identified with a mortality rate of 7.4%. Adding another criterion of cholestasis duration of >2 weeks, persistent cholestasis with DB ≥2.0 mg/dL was observed in 937 (23%) patients with 19.1% mortality. Using the NCM model, we performed subgroup analyses (Supplementary Figure 16). In the performance evaluation of DB ≥2 mg/dL and DB ≥2 mg/dL and duration >2 weeks groups, the AUCs were 0.90 and 0.89, respectively. Other LR analyses using the dataset of the two subgroups were also performed, and all risk variables remained significant, with the exception of albumin (data not shown), indicating that the variables used in the original NCM model and reconstructed models were valid in the subgroups with higher intensity and longer duration of the phenotype. Bootstrap analysis of these LR showed the same variable patterns (data not shown).

Internal and external validations

To validate the NCM model internally in our dataset, the cross-validation and bootstrap re-sampling method were used to estimate the optimism and its 95% CIs. In the

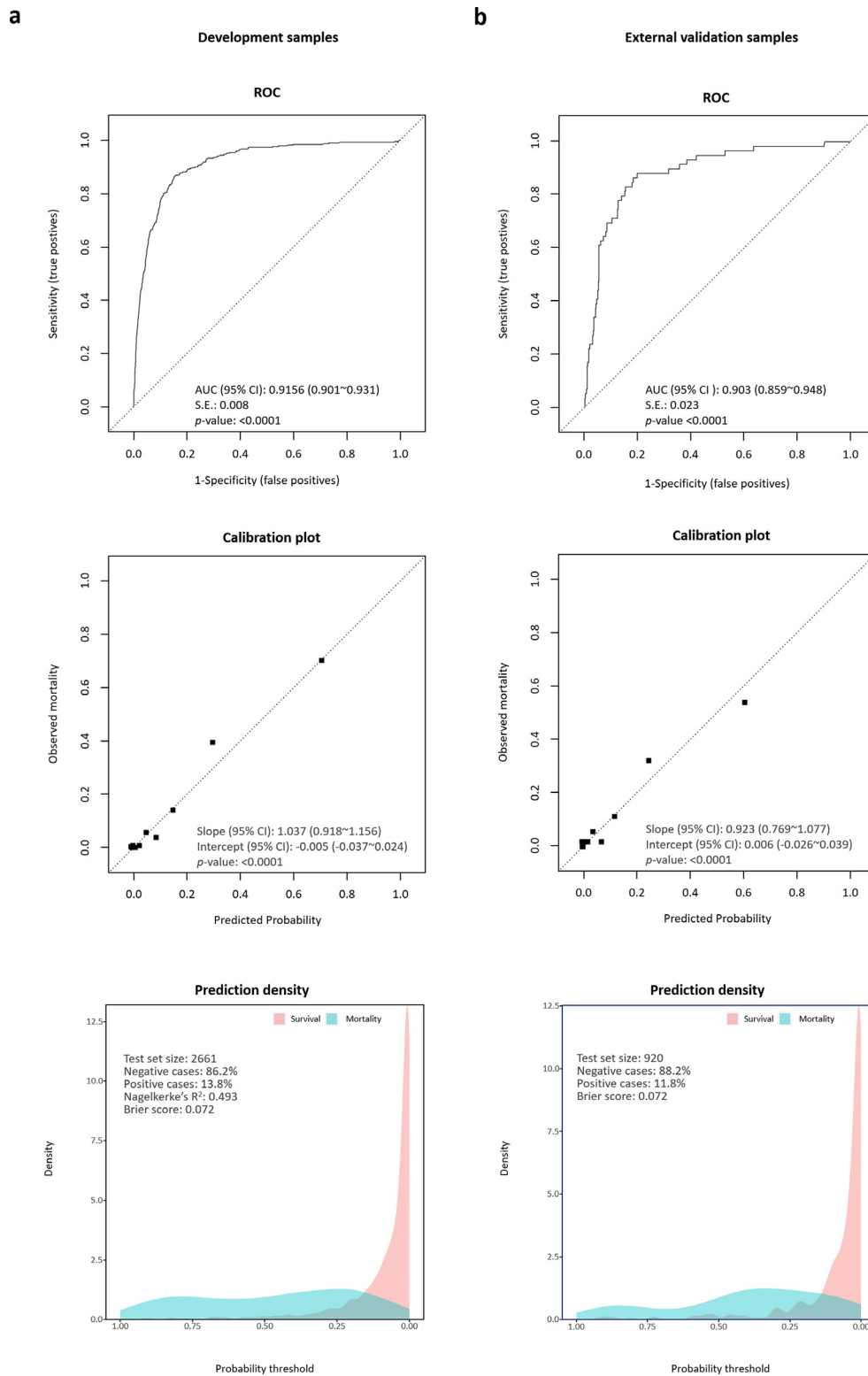


Figure 3. Performance evaluation of LR-based NCM prediction model across development (complete-case) and external validation (complete-case) samples.

10-fold cross-validation, the average AUCs of complete-case and whole-case analyses were 0.913 (95% CI, 0.898–0.927) and 0.936 (95% CI, 0.924–0.948), respectively (Supplementary Table 10). In bootstrap re-sampling, the stability of AUCs was shown on 150 times re-sampling, and the average values of optimism after 500 times re-sampling were 0.003 and 0.0023 for complete-case and whole-case analyses, respectively. Therefore, the optimism-adjusted AUCs were 0.912 (95% CI: 0.897–0.928) and 0.935 (95% CI: 0.922–0.947) for complete-case and whole-case analyses, respectively. Moreover, optimism-adjusted calibration slopes were 1.03 (95% CI, 0.92–1.14) and 1.04 (95% CI, 0.99–1.07), respectively (Supplementary Figure 17). A criterion of internal validation includes obtaining a calibration slope ≥ 0.9 and optimism of AUC ≤ 0.02 .⁴⁹ As a result, no adjustments to the regression coefficients in the NCM prediction model were made.

As external validation is essential in assessing the generalizability of a new model to a plausibly related population,⁵⁰ we collected the retrospective dataset ($n = 2541$ with 174 [6.8%] deaths) of neonatal cholestasis with the same inclusion criteria from two of the largest tertiary hospitals in Korea. Due to insufficient manual reviews of medical charts of external samples, spatial transportability was evaluated in the external validation. Among external samples, the complete-case dataset was available in 36.2% ($n = 920$ with 109 [11.8%] deaths) of all subjects. The NCM prediction model showed acceptable discriminatory performance across the external validation samples (AUC=0.903 and Brier score=0.072) (Figure 3b). Moreover, the level of agreement between the observed and predicted 1-year mortality risk across the external samples was acceptable (calibration slope=0.923, intercept=0.006) compared to the apparent performance. Finally, systemic inspection on the NCM model was performed in accordance with the TRIPOD and CHARM checklists.

Discussion

The present study had two major findings: aetiologic characteristics of neonatal cholestasis in a tertiary hospital and development of a novel neonatal cholestasis-related prediction model for 1-year mortality. First, this is the first report of thousands of subjects with neonatal cholestasis that has never been published in a single centre that describes a variety of aetiologies and their outcomes in detail. A large proportion (54%) of birth issues and heart issues reflected the clinical setting of a tertiary hospital. The variety of liver diseases and genetic disorders also reflected that of a liver transplantation centre, where the largest paediatric transplantation cohort was built in Korea.⁵¹ The proportion (3.6%) of BA was small compared to those reported in other studies (Supplementary Table 4). The authors believe that it is because large number of subjects ($n = 2420$) with

subtle neonatal cholestasis (1 mg/dL <DB <2.0 mg/dL during the 8-week follow-up) enrolled based on the NASPGHN/ESPGHN guideline's cut-off value of DB >1.0 mg/dL.¹ The cut-off DB value of >1 mg/dL used by Harpavat et al. to screen biliary atresia showed an excellent diagnostic accuracy in their well-established population-based study.⁵² In fact, if we adopted another criterion of DB ≥ 2.0 mg/dL in the Childhood Liver Disease Research Network,⁵ 1608 (40%) patients could be enrolled from this dataset. Regardless of the screening value of DB, independent risk factors remained similar in this dataset, and peak DB was positively correlated to the observed mortality rate (Supplementary Figure 2). With regard to mortality in neonatal cholestasis, the literature search showed an overall mortality of 10.8% in these groups (Supplementary Table 4), which was interestingly similar to the value (10.6%) we found in our cohort. However, the interpretation of this coincident finding requires caution because of the different proportions of underlying aetiologies and study inclusions among studies.

Second, a novel prediction model for 1-year mortality named NCM was developed and validated following the recommended steps of the TRIPOD statement.²⁸ This is also the first model to predict the outcome of neonatal cholestasis using large internal and external datasets. Liver-oriented prediction models, such as PELD and KCHC, were also suitable to an extent for further analysis in this cohort (Supplementary Figure 15), with AUCs of 0.678 and 0.841, respectively, as reported in other studies.⁵³ This may indicate that the dataset of our cohort may be appropriate for the analysis of cholestasis. However, the comparison may not be meaningful because the purposes of models, decision rules and datasets of subjects are different.^{25–27} As TRIPOD statement recommends a nomogram for rapid estimation of risk, a nomogram for 1-year overall mortality probability was made based on our NCM prediction analyses (Supplementary Figure 18).

The performance of the NCM prediction model was acceptable in terms of Brier scoring, discrimination and calibration. However, the value of Nagelkerke's R^2 of 0.5 in this LR model indicates that the model explains 50% of data patterns in the LR-based analysis. This low R^2 value may be appropriate because prognostic models generally have R^2 values of 0.2–0.3⁽³²⁾ because substantial future uncertainty within 1 year remains at the individual level in the reality of neonatal cholestasis. Notably, the machine learning RF and XGBoost, known to be efficient classifiers,⁵⁴ agreed with the findings of the LR-based NCM prediction in our study. Especially, the Shapley plot of XGBoost properly explained how the changes and density of variables affect the mortality outcome in this study.

There is limited knowledge about risk factors for mortality in neonatal cholestasis. To the best of our knowledge, no literature identified risk factors for

mortality of neonatal cholestasis, except for a recent study by Santos Silva et al. (Supplementary Table 4). This study identified some predictors of diagnosis and prognosis, which helped build a diagnostic decision algorithm.²³ TB, DB and gamma-glutamyl transferase were related to the poor outcomes in a group diagnosed as having neonatal cholestasis. The findings of Silva's report support the concept of our study using simple laboratory indices in predicting mortality. In the NCM model, the clinical meaning of each predictor is difficult to understand in a clinical perspective, despite rigorous analytic exploration in this study. We speculate that the whole samples have various heterogeneous aetiologies for the explanation. Each patient's condition towards mortality cannot be simply explained. Moreover, the endpoint of the outcome is a future event, death within 1 year. Based on our analyses, lower PLT and ALB levels and higher PT INR, AST, DB, TB, CRP and Cr levels were related to higher mortality risk. Several severe conditions may be shown in combinations of predictors; for example, [low PLT + high PT INR + high CRP + high Cr] in severe septic shock, [high PT INR + high AST + high DB + high TB] in acute liver failure, [high PT INR + high DB] in prematurity with parenteral nutrition-associated liver disease, or [high AST + high DB + high Cr] in ischaemic liver disease from heart failure and congenital heart anomalies. A prospective study in a controlled sample setting is needed to fully understand the specific role of each predictor on future death.

Despite the less interpretable role of predictors, the clinical utility of the NCM prediction model is clear. By providing the probability of mortality, the NCM prediction model can help clinicians be aware of high-risk patients. We believe that patients with a high mortality probability should prioritise the therapeutic approach. The prediction model in medical practice generates decision rules, such as decision support on test ordering, surgical decision-making, using treatment threshold (starting, delaying, and intensifying), cost-effectiveness of treatment, and simple prediction of future events (attention).^{32,55} The decision rule of the NCM prediction model is just a simple prediction for now. However, this NCM prediction model is just about to be listed on the evidence-level I of the prediction model (Supplementary Table 11).⁵⁶ More clinical utility can be evaluated in another cohort.

There are several limitations in this study. First, the study design is retrospective, and a biased cohort is possible. The large sample size in our study cannot equate to the novelty of the research. Conventionally, requirements to derive reliable prediction mode are [1] at least '100 events' and [2] at least 10 events per variable, and preferably 20 in the case of an event rate <20%. The sample size of the complete-case dataset ($n = 2661$) in this study barely meets the preferable sample size requirement. The retrospective study design limits well-controlled data collection. However, the inclusion

criteria are almost similar to a well-designed prospective study for diagnostic screening of BA among neonatal cholestasis.⁵² Despite routine diagnostic and genetic screening, alpha-1-antitrypsin deficiency was absent in our dataset considering the conventional findings.⁵⁶ The prevalence rates of some specific diseases may differ among East Asian subjects.²¹ Moreover, the evaluation for aetiologies was based on the availability of diagnostic modalities in our centre. Second, the principal outcome, which is the 1-year mortality as a hard endpoint, was partially interrupted by various interventions. For example, patients with BA have a Kasai operation at the time of diagnosis and liver transplantation at approximately 1 year of age if the Kasai operation fails.⁵¹ In this context, the nomogram from the NCM prediction model may not be suitable for BA. Surgical techniques and medical care for severe complex anomalies and prematurity have also improved over time. As expected, like a retrospective study, the outcome assessment was not done by the prior protocol, and the study period of outcome assessment was arbitrary. Third, the NCM prediction model provides only the predicted probabilities of mortality risk and the authors believe that the prediction model is inherently away from providing certainty on the individual outcome of the future reality.

Nevertheless, the LR-based NCM prediction model showed internal and external validity, suggesting a possibility of generalisation in a population of children with neonatal cholestasis. Furthermore, this novel NCM model fulfilled all checklists of the TRIPOD and CHARM checklists. Thus, this model strongly suggests the necessity of a prospective study to develop more powerful prediction models. To overcome these limitations, the NCM prediction model needs evidence-level II, assessing the predictive ability when tested prospectively in one setting.⁵⁵ In this light, the prospective cohort of the Childhood Liver Disease Research Network⁵ is vital and needs to expand globally.

In conclusion, the present study provided detailed aetiologic characteristics of the largest paediatric cohort of neonatal cholestasis in a tertiary hospital. In addition, an NCM prediction model for 1-year mortality, which is the first in neonatal cholestasis, showed appropriate performance.

Contributors

S.H.O., J.M.N., and I.K. conceived and designed the study. S.H.O., H.J.C., I.K., and H.J.L. verified and analysed the data and wrote the paper. H.J.O., M.K.A., W.I.B., Y.E.K., B.S.L., D.Y.K., E.J.L. and K.M.K. reviewed the manuscript. J.O.S. and J.S.K. provided the external validation samples. B.S.L., D.Y.K., E.J.L. and K.M.K. participated in the management of the patients and analysed the data. All authors reviewed and approved the manuscript.

Declaration of interests

The authors have no conflicts of interest to declare.

Data sharing

The main data underlying the findings of the current study are provided within the article and its Supplementary Information File. Extra data are available from the corresponding author upon reasonable request after receiving clearance from the institutional review board of our center.

Acknowledgments

This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HR21Co198).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.103890.

References

- Fawaz R, Baumann U, Ekong U, et al. Guideline for the evaluation of cholestatic jaundice in infants: joint recommendations of the North American Society for pediatric gastroenterology, hepatology, and nutrition and the European Society for pediatric gastroenterology, hepatology, and nutrition. *J Pediatr Gastroenterol Nutr.* 2017;64(1):154–168.
- Dick MC, Mowat AP. Hepatitis syndrome in infancy—an epidemiological survey with 10 year follow up. *Arch Dis Child.* 1985;60(6):512–516.
- Hartley JL, Davenport M, Kelly DA. Biliary atresia. *Lancet.* 2009;374(9702):1704–1713. (London, England).
- Serinet MO, Wildhaber BE, Broué P, et al. Impact of age at Kasai operation on its results in late childhood and adolescence: a rational basis for biliary atresia screening. *Pediatrics.* 2009;123(5):1280–1286.
- Hertel PM, Hawthorne K, Kim S, et al. Presentation and outcomes of infants with idiopathic cholestasis: a multicenter prospective study. *J Pediatr Gastroenterol Nutr.* 2021;73(4):478–484.
- Gottesman LE, Del Vecchio MT, Aronoff SC. Etiologies of conjugated hyperbilirubinemia in infancy: a systematic review of 1692 subjects. *BMC Pediatr.* 2015;15:192.
- Mowat AP, Psacharopoulos HT, Williams R. Extrahepatic biliary atresia versus neonatal hepatitis. Review of 137 prospectively investigated infants. *Arch Dis Child.* 1976;51(10):763–770.
- Danks DM, Campbell PE, Jack I, Rogers J, Smith AL. Studies of the aetiology of neonatal hepatitis and biliary atresia. *Arch Dis Child.* 1977;52(5):360–367.
- Hitch DC, Leonard JC, Pysker TJ, Manion CV, Smith EI. Differentiation of cholestatic jaundice in infants. Utility of diethyl-IDA. *Am J Surg.* 1981;142(6):671–677.
- Spivak W, Sarkar S, Winter D, Glassman M, Donlon E, Tucker KJ. Diagnostic utility of hepatobiliary scintigraphy with 99mTc-DISIDA in neonatal cholestasis. *J Pediatr.* 1987;110(6):855–861.
- Motala C, Ireland JD, Hill ID, Bowie MD. Cholestatic disorders of infancy—etiology and outcome. *J Trop Pediatr.* 1990;36(5):218–222.
- Yachha SK, Khanduri A, Kumar M, et al. Neonatal cholestasis syndrome: an appraisal at a tertiary center. *Indian Pediatr.* 1996;33(9):729–734.
- Fischler B, Papadogiannakis N, Nemeth A. Aetiological factors in neonatal cholestasis. *Acta Paediatr.* 2001;90(1):88–92.
- Stormon MO, Dorney SF, Kamath KR, O'Loughlin EV, Gaskin KJ. The changing pattern of diagnosis of infantile cholestasis. *J Paediatr Child Health.* 2001;37(1):47–50.
- Aanpreung P, Laohapansang M, Ruangtrakool R, Kimhan J. Neonatal cholestasis in Thai infants. *J Med Assoc Thai.* 2005;88(Suppl 8):S9–15.
- Tiker F, Tarcan A, Kilicdag H, Gurakan B. Early onset conjugated hyperbilirubinemia in newborn infants. *Indian J Pediatr.* 2006;73(5):409–412.
- Humphrey TM, Stringer MD. Biliary atresia: US diagnosis. *Radiology.* 2007;244(3):845–851.
- Rafeey M, Golzar A, Javadzadeh A. Cholestatic syndromes of infancy. *Pak J Biol Sci.* 2008;11(13):1764–1767.
- Ipek MS, Aydin M, Zenciroglu A, Gokce S, Okumus N, Gulaldi NC. Conjugated hyperbilirubinemia in the neonatal intensive care unit. *Turk J Gastroenterol.* 2013;24(5):406–414.
- Lu FT, Wu JF, Hsu HY, et al. γ -Glutamyl transpeptidase level as a screening marker among diverse etiologies of infantile intrahepatic cholestasis. *J Pediatr Gastroenterol Nutr.* 2014;59(6):695–701.
- Liu P, Guo L, Huang L, et al. Analysis of factors affecting the prognosis of neonatal cholestasis. *Int J Clin Exp Med.* 2015;8(5):8005–8009.
- Abuduxikuer K, Chen R, Wang ZL, Wang JS. Risk factors associated with mortality in neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD) and clinical implications. *BMC Pediatr.* 2019;19(1):18.
- Santos Silva E, Moreira Silva H, Catarino C, Dias CC, Santos-Silva A, Lopes AI. Neonatal cholestasis: development of a diagnostic decision algorithm from multivariate predictive models. *Eur J Pediatr.* 2021;180(5):1477–1486.
- Lauriti G, Zani A, Aufieri R, et al. Incidence, prevention, and treatment of parenteral nutrition-associated cholestasis and intestinal failure-associated liver disease in infants and children: a systematic review. *JPEN J Parenter Enteral Nutr.* 2014;38(1):70–85.
- Freeman RB, Wiesner RH, Harper A, et al. The new liver allocation system: moving toward evidence-based transplantation policy. *Liver Transpl.* 2002;8(9):851–858.
- Jain V, Dhawan A. Prognostic modeling in pediatric acute liver failure. *Liver Transpl.* 2016;22(10):1418–1430.
- Liu E, MacKenzie T, Dobyns EL, et al. Characterization of acute liver failure and development of a continuous risk of death staging system in children. *J Hepatol.* 2006;44(1):134–141.
- Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1–73.
- Damen JA, Hooft L, Schuit E, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ.* 2016;353:i2416. (Clinical research ed).
- Diamond GA. What price perfection? Calibration and discrimination of clinical prediction models. *J Clin Epidemiol.* 1992;45(1):85–89.
- Wang F, Kaushal R, Khullar D. Should health care demand interpretable artificial intelligence or accept “black box” medicine? *Ann Intern Med.* 2020;172(1):59–60.
- Steyerberg EW. *Clinical Prediction Models (Statistics for Biology and Health).* Springer International Publishing.
- Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. (Clinical research ed).
- Enders CK. A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosom Med.* 2006;68(3):427–436.
- Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol.* 2017;9:157–166.
- Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014;11(10):e1001744.
- In Lee K, Koval JJ. Determination of the best significance level in forward stepwise logistic regression. *Commun Stat Simul Comput.* 1997;26(2):559–575.
- Breiman LFJ, Olshen RA, Stone CI. *Classification and Regression Trees.* Belmont, CA: Wadsworth; 1984.
- Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32.
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on*

- Knowledge Discovery and Data Mining*. 2016;785-794 [Online]. Available: <https://arxiv.org/abs/1603.02754>.
- 41 Moore KW, de Waal Malefyt R, Coffman RL, O'Garra A. Interleukin-10 and the interleukin-10 receptor. *Annu Rev Immunol*. 2001;19:683-765.
 - 42 Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.
 - 43 Hilden J, Habbema JD, Bjerregaard B. The measurement of performance in probabilistic diagnosis. III. Methods based on continuous functions of the diagnostic probabilities. *Methods Inf Med*. 1978;17(4):238-246.
 - 44 Nagelkerke NJ. A note on a general definition of the coefficient of determination. *Biometrika*. 1984;78:691-692.
 - 45 Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *J Am Stat Assoc*. 1983;78(382):316-331.
 - 46 Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361-387.
 - 47 Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373-1379.
 - 48 Roberts I, Murray NA. Neonatal thrombocytopenia: causes and management. *Arch Dis Child Fetal Neonatal Ed*. 2003;88(5):F359-FF64.
 - 49 Christodoulou E, van Smeden M, Edlinger M, et al. Adaptive sample size determination for the development of clinical prediction models. *Diagn Progn Res*. 2021;5(1):6.
 - 50 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.
 - 51 Oh SH, Jeong IS, Kim DY, et al. Recent improvement in survival outcomes and reappraisal of prognostic factors in pediatric living donor liver transplantation. *n/a(n/a)* 2022.
 - 52 Harpavat S, Garcia-Prats JA, et al. Diagnostic yield of newborn screening for biliary atresia using direct or conjugated bilirubin measurements. *JAMA*. 2020;323(12):1141-1150.
 - 53 Lee EJ, Kim JW, Moon JS, et al. Development of a prognostic score to predict mortality in patients with pediatric acute liver failure. 2020;70(6):777-82.
 - 54 Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med*. 2016;23(3):269-278. official journal of the Society for Academic Emergency Medicine.
 - 55 Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144(3):201-209.
 - 56 Balistreri WF, Bezerra JA. Whatever happened to "neonatal hepatitis"? *Clin Liver Dis*. 2006;10(1):27-53. v.