Summer 8-14-2020

Urinary Bile Acid Indices as Diagnostic and Prognosic Biomarkers for Liver Diseases

Jawaher Alamoudi
University of Nebraska Medical Center

URINARY BILE ACID INDICES AS DIAGNOSTIC AND PROGNOSTIC BIOMARKERS FOR LIVER DISEASES

by

Jawaher Abdullah Alamoudi

A DISSERTATION

Presented to the Faculty of

the University of Nebraska Graduate College

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

Department of Pharmaceutical Sciences
Graduate Program

Under the Supervision of Professor Yazen Alnouti

University of Nebraska Medical Center
Omaha, Nebraska

May, 2020

Supervisory Committee:

Yazen Alnouti, Ph.D.

Jane Meza, Ph.D.

Justin Mott, Ph.D.

Marco Olivera, M.D.

ACKNOWLEDGMENTS

First and foremost, I thank God (Allah) for giving me all the help, strength, and determination to complete my thesis.

I would like to express my deep sense of gratitude to a number of people who have been constantly supporting and encouraging me throughout my graduate studies. I am deeply indebted to my mentor Dr. Yazen Alnouti for providing me the opportunity to work in his lab. Dr. Alnouti has been very instrumental in my research career. He has educated me on numerous qualities that are required for becoming a good academic researcher.

I would like to thank my committee members Dr. Jane Meza, Dr. Marco Olivera, and Dr. Justin Mott for their valuable inputs, suggestions and guidance during my studying here at UNMC. I am indebted to Dr. Nagsen Gautam for nurturing my research skills and for all the suggestions and inputs during my studies. Also, I would like to thank Princess Nourah Bint Abdulrahman University for providing me a scholarship and financial support to complete my graduate studies.

I owe a special thanks to my family including my mom and dad, my sister, and my brothers for their love, prayers, support, and encouragement throughout the years. Mom, dad I do not know how to thank you enough for providing me with the opportunity to be where I am today. I love you so much. Also, I would like to thank my lovely daughters (Juwan, Sulaf, and Kienda) who are the pride and joy of my life. I love you more than anything and I appreciate all your patience and support during mommy's Ph.D. study.

Finally, I do not know how to thank my steadfast husband, my best friend, and love of my life, Hasan Albar, for his enduring love, for believing in me long, for always showing how proud he is of me, and for keeping things going. He gave me support, help, discussed ideas and prevented several wrong turns. Hasan thank you for being so understanding, and for putting up with me through the toughest moments of my life. I thank God for

enlightening my life with your presence. You are the best thing that has ever happened to me. I dedicate this Ph.D. thesis to you all.

URINARY BILE ACID INDICES AS DIAGNOSTIC AND PROGNOSTIC BIOMARKERS FOR LIVER DISEASES

Jawaher Abdullah Alamoudi, Ph.D.
University of Nebraska Medical Center, 2020

Supervisor: Yazen Alnouti, Ph.D.

Hepatobiliary diseases result in the accumulation of toxic bile acids (BA) in the liver, blood, and other tissues which may contribute to an unfavorable prognosis. We compared the urinary BA profile between 300 patients with hepatobiliary diseases vs. 103 healthy controls. Also, we investigated the use of the urinary BA profile to develop survival models to predict the prognosis of hepatobiliary diseases. The urinary BA profile, a set of non-BA parameters, and the adverse events of liver transplant and/or death were monitored in patients with cholestatic liver diseases for up to 7 years. The BA profile was characterized by calculating BA indices, which quantify the composition, metabolism, hydrophilicity, and toxicity of the BA profile. Total and individual BA concentrations were higher in all patients. The percentage of secondary BA (DCA and LCA) was markedly lower, while the percentage of primary BA (CDCA, CA, and HCA) was markedly higher in patients compared to controls. In addition, the percentage of taurine-amidation was higher in patients. The increase in non-12α-OH BA was more profound than 12α-OH BA (CA and DCA) causing a decrease in the 12α-OH/non-12α-OH ratio in patients. This trend was stronger in patients with more advanced liver diseases as reflected by the model for end-stage liver disease (MELD) score and the presence of hepatic decompensation. The percentage of sulfation was also higher in patients with more severe forms of liver diseases. In general, BA indices had much lower inter- and intra-individual variability compared to absolute concentrations of the individual and total BA. In addition, BA indices demonstrated high area under the receiver operating characteristic (ROC) curves, and

changes of BA indices were associated with the risk of having a liver disease as determined by the logistic regression analysis, which demonstrated their use as diagnostic biomarkers for cholestatic liver diseases.

We have developed and validated the bile-acid score (BAS) model (a survival model based on BA indices) to predict the prognosis of cholestatic liver diseases. The BAS model was more accurate and results in higher true-positive and true-negative prediction of death compared to both non-BAS and MELD models. Both 3- and 5-year survival probabilities markedly decreased as a function of BAS. Moreover, patients with high BAS had a 4-fold higher rate of death and lived for an average of 11 months shorter than subjects with low BAS. The increased risk of death with high vs. low BAS was also 2-4-fold greater and the shortening of lifespan was 6-7-month lower compared to MELD or non-BAS. One application for BAS could be to define the most seriously ill liver patients, who may need earlier intervention such as liver transplantation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

UPLC: ultra-performance liquid chromatography

LC-MS/MS: liquid chromatography-tandem mass spectrometry

BA: bile acids

G-BA: glycine-amidated bile acids

T-BA: taurine-amidated bile acids

MCA: muricholic acid

CA: cholic acid

HCA: hyocholic acid

MDCA: murideoxycholic acid

UDCA: ursodeoxycholic acid

HDCA: hyodeoxycholic acid

DCA: deoxycholic acid

CDCA: chenodeoxycholic acid

LCA: lithocholic acid

HI: hydrophobicity index

CYP8B1: Cytochrome P450 Family 8 Subfamily B Member 1

IS: internal standard

MeOH: methanol

U: unamidated

G: glycine

T: taurine

SULT2A1: sulfotransferase 2A1

BAT: bile acid-coenzyme A: amino acid N-acyltransferase

BSEP: bile salt export pump

MDR3: multidrug resistance protein 3

MRP2: multidrug resistance–associated protein 2

TGR5: G-protein-coupled membrane receptor

FXR: farsenoid-X-receptor

ROC: Receiver operating characteristics

AUC: Area under the ROC curve

MELD: Model for end-stage liver disease

CTP: Child-Turcotte-Pugh

AST: Aspartate transaminase

ALT: Alanine transaminase

INR: International Normalized Ratio

PSC: primary biliary cholangitis

PSC: primary sclerosing cholangitis

NASH: nonalcoholic steatohepatitis

LFT: liver function test

OR: Odds ratio

HR: hazard ratio

LT: liver transplant

PH: proportional hazard

KM: Kaplan-Meier

BAS: bile-acids score

BAS_0: average bile-acids score

CHAPTER 1

INTRODUCTION

1.1. Bile acids (BA) structure, function, and toxicity

Bile acids (BA) are amphipathic steroid molecules synthesized in the liver from approximately 500 mg of cholesterol each day [1]. **Figure 1.1** shows the chemical structure of the major BA including cholic acid (CA), muricholic acid (MCA), chenodeoxycholic acid (CDCA), Hyocholic acid (HCA), deoxycholic acid (DCA), ursodeoxycholic acid (UDCA), murideoxycholic acid (MDCA), hyodeoxycholic acid (HDCA), lithocholic acid (LCA), their glycine (G) and taurine (T) amidates, and sulfate conjugates. According to their chemical structure, BA can be categorized into tri-OH (HCA, CA, and MCA), di-OH (UDCA, MDCA, CDCA, HDCA, and DCA), and mono-OH BA (LCA).

BA have many physiological functions such as the fat absorption, cholesterol elimination, and protection against bacterial overgrowth [2, 3]. In addition, BA have been identified as signalizing molecules/hormones that exert their functions at the molecular level, via binding to numerous receptors, primarily, the surface G-protein-coupled membrane receptor (TGR5), and the nuclear farsenoid-X-receptor (FXR) [4]. BA are involved in the regulation of their own homeostasis, glucose and lipid metabolism, thyroid hormone signaling, energy expenditure, and cellular immunity [3, 5, 6]. In addition, BA such as UDCA and norursodeoxycholic acid (norUDCA), as well as BA analogs such as 6α-ethyl-CDCA (obeticholic acid, OCA) and 12-monoketocholic acid (12-MKC) are proven to have therapeutic applications for the treatment of cholestatic liver diseases and the metabolic syndrome [6-8].

In contrast to their physiological functions, BA are also cytotoxic and exhibit pathological effects at high concentrations due to their direct detergent effects on biological membranes, cancer promoting effects, and necrotic and apoptotic effects via mitochondria and endoplasmic reticulum-mediated toxicities [6, 9]. When present at high concentration BA bind to the lipid bilayer and solubilize plasma membrane components.

Elevated intracellular levels of BA can lead to loss of the mitochondrion integrity. This can lead to the permeabilization of mitochondrial membranes, which provokes depolarization of the organelle, mitochondrial swelling, and uncoupling of oxidative phosphorylation. Ultimately this leads to mitochondrial collapse, release of cytochrome c, and activation of apoptosis. Also, apoptosis can be initiated in the endoplasmic reticulum (ER). Accumulation BA induce ER stress, which in turn can activate programmed cell death pathways. BA toxicity is thought to be highly correlated with hydrophobicity. The efficiency of BA to solubilize membrane lipids, such as fatty acids, cholesterol or phospholipids, is enhanced with increasing BA hydrophobicity; more hydrophobic bile acids are more cytotoxic [6].

Individual BA differ markedly in their physicochemical properties, physiological, and pathological roles. Lipophilicity of BA is influenced by both the side chain structures and BA nucleus [10]. There is inverse relationship between the lipophilicity and number of hydroxyl (OH) groups present on steroid nucleus. Furthermore, amidation of BA side chain with G or T decreases pKa, reduces lipophilicity and increases solubility. Mono-OH BA (LCA) is more lipophilic than di-OH BA (DCA and CDCA), which in turn more lipophilic than tri-OH BA (MCA and CA). Also, lipophilicity is determined by the position and stereochemistry of (OH) groups, where the CA (tri-OH BA) is more lipophilic than UDCA (di-OH BA). The di-OH BA lipophilicity is in the order of DCA, CDCA, followed by UDCA. Amidation increases the acidity of unconjugated BA, where pKa is decreased from 5.5 for the unamidated BA to 4.5 and 1.5 for those with G- and T-amidation, respectively [11]. This results in complete ionization of BA at physiological pH, which markedly decreased lipophilicity and increases their solubility and therefore, reduces membrane permeability. BA with two “faces” are planar molecules. The OH groups are only present at one face of the molecule, making it hydrophilic. The other face does not have OH groups, thus hydrophobic.

BA with OH groups presented on both faces (α and β orientation) of the cholesterol ring are more hydrophilic than their counterparts with the same number of OH groups in the α orientation. In β-orientation, OH groups are located above the steroid nucleus and are equatorial to the plane of the steroid nucleus, while in α-orientation OH groups are located below the steroid nucleus and are axial to the plane of the steroid nucleus. MCA contains a 6α-orientated OH group that is equatorially positioned to the steroid nucleus plane, thus it escapes from this rule. The equatorial location of OH groups confers polarity to the hydrophobic concave side of the steroid nucleus. Therefore MCA, containing both 6α and 7β-orientated OH groups, and UDCA, with its 7β-orientated OH group, are more hydrophilic than other BA with the same number of OH groups axially positioned to the steroid nucleus [5].

BA are amphipathic molecules and their anions in water self-associate to form micelles. The critical micelle concentration (CMC) values of mono-OH BA (LCA:0.5 mM) is more than Di-OH BA (DCA:3 mM, CDCA:4 mM, UDCA:7 mM) which in turn more than tri-OH BA (CA: 11 mM). DCA and CDCA have critical micellar concentrations lower than that of CA; thus, they are more cytotoxic at any given concentration. The hydrophobicity index (HI) of BA is calculated based on the capacity factor and retention time on a C18 column. HI of BA ranged from +1.46 for (LCA) the hydrophobic BA to - 0.94 for (T-UDCA) the hydrophilic BA. Membrane damage correlates with the hydrophobicity and detergent effect of individual BA. The rank order of BA cytotoxicity, from highest to lowest is: LCA > CDCA, DCA > CA > UDCA [12].

The structural differences in individual BA have the significant effect on the specificity of activation of BA receptors. For example, secondary BA (LCA and DCA) are more potent TGR5 activators compared to primary BA (CDCA and CA). In addition, the affinity of T-amidated BA is higher than G-amidated and unamidated BA [13]. FXR

activation is limited to primary BA (CDCA) and to a lesser extent to secondary BA (DCA and LCA) [14], while MCA act as FXR antagonist [15].

Individual BA also differ in their pathological effects, for example, the most hydrophobic BA (LCA) causes cholestasis in rats, mice, and hamsters, whereas the more hydrophilic BA (CA) causes hypercholeresis. In addition, LCA is 100 times more potent mutagenic than CDCA and 50 times more potent than CA in causing red blood cell (RBC) hemolysis. The G-amidates are more cytotoxic than T-amidates and induce more cell membrane lysis than the corresponding T-amidates [16, 17]. Therefore, the amount and the composition of the BA pool must be tightly regulated to maintain normal physiological levels and to prevent the toxicity caused by the accumulation of the more toxic BA.

1.2. BA synthesis, metabolism, transport, and enterohepatic recirculation

BA synthesis occurs in liver cells via cytochrome P450-mediated oxidation of cholesterol in a many-step process [18]. The major pathway for the synthesis of the BA is initiated by hydroxylation of cholesterol at the 7α position by the action of CYP7A1 enzyme [1]. This pathway is called as the "classic" or "neutral" pathway of BA synthesis. Next step of BA synthesis includes the 3β-OH oxidation and the C5-C6 double bond isomerization via the microsomal C27-3β-hydroxysteroid dehydrogenase (C27-3β-HSD). The resulting intermediate is either, passed on directly to the next step or, hydroxylated by the microsomal CYP8B1 at the 12α position. The 12α hydroxylated intermediates and those which escaped 12α hydroxylation have the C4-C5 double bond and their C3 oxo decreased to yield 3α-OH intermediates by the 3α-HSD and oxosteroid 5β reductase [19]. 12α hydroxylation will eventually produce CA, whereas intermediates which were not hydroxylated will eventually produce CDCA. CDCA and CA are the primary BA in humans. The following step in BA synthesis is the hydroxylation and oxidation to a carboxylic acid by the mitochondrial CYP27A1 enzyme at the C27 position followed by ligation to

coenzyme A by the bile acid coenzyme-A synthetase (BAS) [20]. The side chain of these C27 intermediates are subsequently shortened to C24 BAs via β-oxidation in the peroxisomes. The last step in BA synthesis is the amidation of the BA-CoA with an amino acid, usually glycine (G) or taurine (T), by the BA-CoA: amino acid N acyltransferase (BAT).

There are alternative pathways for BA synthesis, that do not require the initiation by CYP7A1 [21]. These pathways are initiated by the hydroxylation of the cholesterol side chain at the C27, C25, or C24 positions by various enzymes such as, CYP46A1 and CYP27A1. The resulting oxysterols are then hydroxylated at the 7α position by CYP39A1 and CYP7B1. Alternative BA pathways appear more important in conditions associated with deficiency in CYP7A1 activity. The alternative pathways produce predominantly CDCA.

BA synthesized in the liver and secreted into bile, which flows through the bile duct to the intestine. BA are absorbed from the intestine, got back to the liver, and re-secreted into bile. This cycle is called the enterohepatic recirculation of BA [18]. BA are excreted from the liver into bile via efflux transporters, which are the bile salt export pump (BSEP), multidrug resistance–associated protein 2 (MRP2), and multidrug resistance protein 3 (MDR3), located in the canalicular membrane of hepatocytes [22]. Secreted bile is stored in gallbladder, which, under the influence of cholecystokinin secretion after meal ingestion, contracts to empty its contents into the duodenum. Most amidated BA in the small intestine, are actively absorbed in the ileum, while unconjugated BA are passively absorbed throughout the intestinal tract [23, 24]. Partial deamidation takes place by the bacteria in the small intestine, and the liberated unconjugated BA are passively absorbed [23, 25]. Unabsorbed BA are passed along from the small to large intestine. BA undergo bacterial transformation in the large intestine, including deamidation, de-hydroxylation, epimerization, dehydrogenation, and desulfation to produce secondary BA [18, 26]. LCA

and DCA are secondary BA produced from the de-hydroxylation of CDCA and CA. BA that are not absorbed from the large and small intestine are excreted in feces. Ninety-five percent of BA excreted in bile are re-absorbed throughout the intestinal tract and less than five percent are excreted in feces. Absorbed BA are carried in the portal vein and extracted via the liver by active or passive diffusion. Most BA in hepatocytes are amidated, but other metabolic pathways take place such as, sulfation, hydroxylation or glucuronide conjugation. The newly synthesized and reabsorbed BA are then excreted in bile to complete the enterohepatic cycle.

1.3. BA and hepatobiliary diseases

BA have deleterious effects on the liver including cholestasis and alterations in liver structure and hepatocyte ultrastructure in several animal species [18, 27]. Cholestatic liver diseases is a diverse group of hepatobiliary diseases associated with a reduction in bile flow due to impairment of bile production or impairment of bile flow into bile duct [28]. The impediment in bile flow cause the accumulation of BA in the liver, which spills out into the systemic circulation and extrahepatic tissues and eventually into urine. Numerous clinical and preclinical studies have shown up to a 100-fold increase in BA concentrations in the blood and urine during various liver diseases [18, 29-32].

There is ample evidence from animal and human studies to indicate that accumulation of BA in the liver, systemic blood, and extrahepatic tissues can worsen the liver condition that lead to their accumulation. The accumulation of toxic BA in cholestasis induces hepatotoxicity, extrahepatic toxicity such as encephalopathy, which may contribute to the unfavorable liver disease prognosis [33]. For example, BA concentrations were shown to correlate with the progression of damage to the liver and bile duct in cholestatic rats, rabbits, and in humans [34-37]. In addition, several animal studies indicate that the accumulation of toxic BA is linked to the occurrence and severity of

complications after liver transplantation. Other studies shown that patients with low concentrations of BA are less likely to develop hepatobiliary complications after transplantation [36]. The imbalance of bile salt transporters is one of the main causes of cholestasis. The intracellular accumulation of toxic bile components initiates a cascade of different feedback loops mediated by BA receptor, which leads to the upregulation of proteins involved in hepatic bile secretion, while hepatocellular uptake of bile acids and bile acid synthesis is inhibited [38]. Therefore, the damage to hepatocytes and cholangiocytes by accumulated BA in liver diseases may be better linked to more toxic individual BA rather than total BA levels [28]. The evidence from animal and human studies indicates that the accumulation of certain toxic BA lead to unfavorable prognosis of hepatobiliary diseases.

1.4. BA as biomarkers of liver diseases

The etiology and pathogenesis of most cholestatic diseases are poorly understood and a pharmacological cure for these diseases is not yet available. Most cholestatic diseases progress toward end stage liver failure, which likely requires liver transplantation. After liver transplantation, a large proportion of the overall complications occur after liver transplant [39]. Chronic liver diseases are led to over 41,000 deaths in the United States in 2017, making it the 11th leading cause of mortality [40].

The most commonly used biomarkers for the diagnosis of liver diseases include aspartate transaminase (AST), alanine transaminase (ALT), alkaline phosphatase (ALP), λ-glutamyltransferase (GGT), albumin, and bilirubin [41, 42]. However, these serum markers are not specific to the liver or bile duct injuries and may increase in non-hepatobiliary diseases. Also, they require severe cell injury at advanced disease stages before their levels increase in the blood [41, 42]. Multifactorial models with multiple parameters based on these biomarkers are also frequently used and offer advantages

compared to the use of their individual biomarker components such as the Child-Turcotte-Pugh (CTP) score. The CTP score, originally developed to predict portosystemic shunt surgery outcomes in cirrhotic patients, formed the basis on which liver disease severity was assessed. However, the usefulness of CTP was limited by a number of inherent problems [43].

More recently, the mayo model for end-stage liver disease (MELD) was developed to predict three-month mortality of patients with end-stage liver disease [44, 45]. MELD is calculated based on serum bilirubin, creatinine, and international normalized ratio (INR), which are related to both liver and renal functions. The MELD score is superior to other prognostic models in patients with end-stage liver disease, such as CTP score. MELD is also currently used in many countries to classify liver diseases' patients awaiting transplantation to identify patients with the highest priority for liver transplant (LT) [46]. Since its implementation, MELD led to a sharp decrease in the number of individuals waiting for liver transplant and decreased mortality on the waiting list without affecting post-transplant survival [46, 47]. Although mainly adopted for use in patients waiting for liver transplant, the MELD score has additionally proved to be an effective predictor of outcome in other circumstances, for example, patients with fulminant hepatic failure or alcoholic hepatitis and patients with cirrhosis going for surgery [47]. However, despite its widespread application, MELD has some limitations. MELD is based on three objective laboratory variables, that are not necessarily liver specific. For example, serum bilirubin can be elevated in cases of hemolysis or sepsis. Serum creatinine can also be elevated from an underlying kidney disease that unrelated to hepatorenal syndrome and is a poor surrogate of renal function in cirrhotic patients. In addition, patients may have an elevated INR which can be secondary to warfarin use. Any of these conditions can increase the MELD score and overestimate the liver disease severity [47, 48]. Furthermore, several

studies have shown that patients with cholestatic liver diseases may still have high mortality rates despite having low MELD scores [48, 49].

Further diagnosis and assessment of liver disease prognosis is empirical and rely on invasive and risky endoscopic procedures and histological evaluation of liver biopsies [50]. Therefore, there is a critical need to find noninvasive biomarkers of liver injury that can be used to help diagnose, evaluate, and predict the prognosis and the therapeutic outcomes of hepatobiliary diseases.

Because of their known hepatotoxicity and established accumulation in liver diseases, BA were extensively investigated for decades as biomarkers for numerous hepatobiliary diseases [32, 51-53] including intrahepatic cholestasis of pregnancy [54, 55], biliary atresia [52, 56], primary biliary cirrhosis [57], primary sclerosing cholangitis [57], alcoholic liver diseases [58, 59], nonalcoholic fatty liver disease [60], and viral hepatitis [61, 62]. Currently, BA are clinically used as biomarkers for the diagnosis of intrahepatic cholestasis of pregnancy and biliary atresia in infants [51, 53, 63].

1.5. BA indices

Despite these extensive efforts the potential use of BA as a marker for liver diseases have never translated into the clinic due to major limitation, which prevented them from being used as a reliable biomarker of hepatobiliary diseases in the clinic. First, due to the marked differences in the physiological and pathological properties of the various BA, detailed profiling of the more toxic and relevant individual BA rather than total BA concentration may better correlate with the liver condition during hepatobiliary diseases [29, 31, 64]. More importantly, the extreme inter-and intra-individual variability of total and individual BA concentrations makes it very hard to determine the normal baseline ranges in the absence of liver diseases. BA have marked inter- individual variability due to many factors such as gender, alcohol consumption, and weight [52]. In

addition, high intra-individual variability of serum and urinary BA levels was reported due to numerous factors such as food ingestion, diurnal variation, and medication intake [30]. In an effort to address these limitations, we have developed the concept of “BA Indices”. BA indices are ratios calculated from the absolute concentration of individual BA and their metabolites. These ratios quantify in detail, the composition, metabolism, hydrophilicity, formation of secondary BA, and toxicity of the BA profile [28, 65, 66].

Because BA indices are ratios, we proposed that BA indices should have lower variability than the absolute BA concentrations used to calculate them. Indeed, we have demonstrated that BA indices offered numerous advantages over absolute total and individual BA concentrations including low inter- and intra-individual variability and were resistant to covariate influences such as age, BMI, food consumption, gender, and moderate alcohol consumption [28, 65]. For example, the relative standard deviation (RSD) from 66-256% for total and individual absolute BA concentrations in serum, while it was less than 100% and as low as 13% for BA indices in the same healthy controls’ population. After food ingestion, serum BA increase due to the release of cholecystokinin, which stimulates gallbladder contraction resulting in increasing bile flow into the intestine [65]. Therefore, feeding status should be controlled before serum BA can be utilized as a reliable biomarker. We have reported that one hour after a standardized meal, the absolute total and most of the individual BA concentrations more than doubled and increased more than 10-fold in some individuals, while BA indices changed by only 10-40% in the same individuals [65]. The same trend was also observed in urine, where BA indices had much smaller variability with as low as 10% RSD and were resistant to feeding status compared to absolute BA concentrations in the same population [65]. Therefore, noninvasive urinary BA indices have a better potential in the clinic than the absolute serum or urine BA concentrations as biomarkers for hepatobiliary diseases. In addition, data from animal models have shown that urinary BA concentrations can replace serum liver

enzymes as a diagnostic indicator of liver disease in humans [58, 61]. Furthermore, we have showed that BA indices may outperform currently used liver enzymes as biomarkers such as AST and ALT for the diagnosis of cholestatic hepatobiliary diseases in humans [28].

1.6. Research hypothesis and objectives

We hypothesized that BA indices can serve as a biomarker for hepatobiliary diseases diagnosis and prognosis.

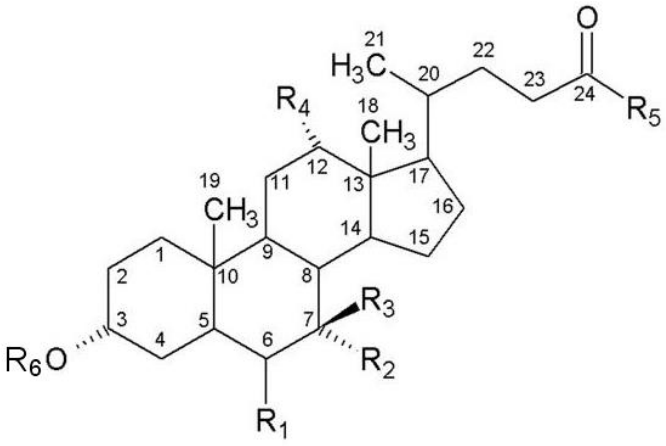
To test this hypothesis, we proposed the following specific aims:

1. Investigate the use of the BA profile in urine as a biomarker for the diagnosis of liver diseases.
2. Develop survival models based on BA indices to predict the prognosis of hepatobiliary diseases.

In chapter 2 of this thesis, we compared the urinary BA profile between healthy controls and patients with hepatobiliary diseases. The BA profile was also compared between patients with different severity of liver disease as determined by model for end-stage liver disease (MELD). In addition, the BA profile was compared between compensated and decompensated patients. The usefulness of BA indices as biomarkers was determined using various statistical tests and their utility was compared with the currently used liver enzyme tests in clinic for diagnosing a liver disease.

In chapter 3, we have developed survival models based on BA indices and non-BA parameters to predict the prognosis of hepatobiliary diseases. The performance of the different models in predicting the occurrence of adverse events of death only and death and/or liver transplant were compared between the different models using various statistic outcomes.

Figure 1.1 The chemical structure of major BA and their glycine, taurine, and sulfate conjugates.



Bile acid	R_1	R_2	R_3	R_4
Tri-OH BA				
Cholic acid (CA)	H	OH	H	OH
α-Muricholic acid (α-MCA)	β-OH	OH	H	H
β-Muricholic acid (β-MCA)	β-OH	H	OH	H
ω-Muricholic acid (ω-MCA)	α-OH	H	OH	H
Hyocholic acid (HCA)	α-OH	OH	H	H
Di-OH BA				
Chenodeoxycholic acid (CDCA)	H	OH	H	H
Deoxycholic acid (DCA)	H	H	H	OH
Ursodeoxycholic acid (UDCA)	H	H	OH	H
Mono-OH BA				
Lithocholic acid (LCA)	H	H	H	H
R_5				
Unamidated BA	OH			
Glycine-amidated BA (G-BA)	NH_2CH_2COOH			
Taurine-amidated BA (T-BA)	$NH_2CH_2CH_2SO_3H$			
R_6				
Unsulfated BA	H			
Sulfated BA	SO_3H			

1.7. References

1. Russell, D.W., *The enzymes, regulation, and genetics of bile acid synthesis.* Annu Rev Biochem, 2003. **72**: p. 137-74.
2. Hofmann, A.F., *The continuing importance of bile acids in liver and intestinal disease.* Arch Intern Med, 1999. **159**(22): p. 2647-58.
3. Monte, M.J., et al., *Bile acids: chemistry, physiology, and pathophysiology.* World J Gastroenterol, 2009. **15**(7): p. 804-16.
4. Khurana, S., J.P. Raufman, and T.L. Pallone, *Bile acids regulate cardiovascular function.* Clin Transl Sci, 2011. **4**(3): p. 210-8.
5. Thomas, C., et al., *Targeting bile-acid signalling for metabolic diseases.* Nat Rev Drug Discov, 2008. **7**(8): p. 678-93.
6. Maillette de Buy Wenniger, L. and U. Beuers, *Bile salts and cholestasis.* Dig Liver Dis, 2010. **42**(6): p. 409-18.
7. Sharma, R., A. Long, and J.F. Gilmer, *Advances in bile acid medicinal chemistry.* Curr Med Chem, 2011. **18**(26): p. 4029-52.
8. Ethanic, M., et al., *Pharmacological Applications of Bile Acids and Their Derivatives in the Treatment of Metabolic Syndrome.* Front Pharmacol, 2018. **9**: p. 1382.
9. Pauli-Magnus, C. and P.J. Meier, *Hepatocellular transporters and cholestasis.* J Clin Gastroenterol, 2005. **39**(4 Suppl 2): p. S103-10.
10. Roda, A., et al., *Bile acid structure-activity relationship: evaluation of bile acid lipophilicity using 1-octanol/water partition coefficient and reverse phase HPLC.* J Lipid Res, 1990. **31**(8): p. 1433-43.
11. Hofmann, A.F. and K.J. Mysels, *Bile acid solubility and precipitation in vitro and in vivo: the role of conjugation, pH, and Ca2+ ions.* J Lipid Res, 1992. **33**(5): p. 617-26.
12. Perez, M.J. and O. Briz, *Bile-acid-induced cell injury and protection.* World J Gastroenterol, 2009. **15**(14): p. 1677-89.
13. Kawamata, Y., et al., *A G protein-coupled receptor responsive to bile acids.* J Biol Chem, 2003. **278**(11): p. 9435-40.
14. Makishima, M., et al., *Identification of a nuclear receptor for bile acids.* Science, 1999. **284**(5418): p. 1362-5.
15. Jiang, C., et al., *Intestine-selective farnesoid X receptor inhibition improves obesity-related metabolic dysfunction.* Nat Commun, 2015. **6**: p. 10166.
16. Powell, A.A., et al., *Bile acid hydrophobicity is correlated with induction of apoptosis and/or growth arrest in HCT116 cells.* Biochem J, 2001. **356**(Pt 2): p. 481-6.
17. Rust, C., et al., *Phosphatidylinositol 3-kinase-dependent signaling modulates taurochenodeoxycholic acid-induced liver injury and cholestasis in perfused rat livers.* Am J Physiol Gastrointest Liver Physiol, 2005. **289**(1): p. G88-94.
18. Alnouti, Y., *Bile Acid sulfation: a pathway of bile acid elimination and detoxification.* Toxicol Sci, 2009. **108**(2): p. 225-46.
19. Penning, T.M., et al., *Structure-function of human 3 alpha-hydroxysteroid dehydrogenases: genes and proteins.* Mol Cell Endocrinol, 2004. **215**(1-2): p. 63-72.
20. Mihalik, S.J., et al., *Participation of two members of the very long-chain acyl-CoA synthetase family in bile acid synthesis and recycling.* J Biol Chem, 2002. **277**(27): p. 24771-9.

21. Axelson, M. and J. Sjovall, *Potential bile acid precursors in plasma--possible indicators of biosynthetic pathways to cholic and chenodeoxycholic acids in man.* J Steroid Biochem, 1990. **36**(6): p. 631-40.
22. Byrne, J.A., et al., *The human bile salt export pump: characterization of substrate specificity and identification of inhibitors.* Gastroenterology, 2002. **123**(5): p. 1649-58.
23. Aldini, R., et al., *Intestinal absorption of bile acids in the rabbit: different transport rates in jejunum and ileum.* Gastroenterology, 1996. **110**(2): p. 459-68.
24. Kuipers, F., et al., *Separate transport systems for biliary secretion of sulfated and unsulfated bile acids in the rat.* J Clin Invest, 1988. **81**(5): p. 1593-9.
25. Takikawa, H., et al., *Absorption of unconjugated bile acids and tauroursodeoxycholate in the rat intestine.* J Gastroenterol Hepatol, 1997. **12**(12): p. 815-21.
26. Philipp, B., *Bacterial degradation of bile salts.* Appl Microbiol Biotechnol, 2011. **89**(4): p. 903-15.
27. Palmer, R.H., *Bile acids, liver injury, and liver disease.* Arch Intern Med, 1972. **130**(4): p. 606-17.
28. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases II. Signature profiles in patients.* Toxicol Sci, 2015. **143**(2): p. 308-18.
29. Makino, I., et al., *Sulfated and nonsulfated bile acids in urine, serum, and bile of patients with hepatobiliary diseases.* Gastroenterology, 1975. **68**(3): p. 545-53.
30. Summerfield, J.A., et al., *Evidence for renal control of urinary excretion of bile acids and bile acid sulphates in the cholestatic syndrome.* Clin Sci Mol Med, 1977. **52**(1): p. 51-65.
31. Takikawa, H., T. Beppu, and Y. Seyama, *Urinary concentrations of bile acid glucuronides and sulfates in hepatobiliary diseases.* Gastroenterol Jpn, 1984. **19**(2): p. 104-9.
32. van Berge Henegouwen, G.P., et al., *Sulphated and unsulphated bile acids in serum, bile, and urine of patients with cholestasis.* Gut, 1976. **17**(11): p. 861-9.
33. LaRusso, N.F., et al., *Primary sclerosing cholangitis: summary of a workshop.* Hepatology, 2006. **44**(3): p. 746-64.
34. Dueland, S., et al., *Regulation of cholesterol and bile acid homoeostasis in bile-obstructed rats.* Biochem J, 1991. **280 (Pt 2)**: p. 373-7.
35. Kawai, H., et al., *Efficacy of urine bile acid as a non-invasive indicator of liver damage in rats.* J Toxicol Sci, 2009. **34**(1): p. 27-38.
36. Buis, C.I., et al., *Altered bile composition after liver transplantation is associated with the development of nonanastomotic biliary strictures.* J Hepatol, 2009. **50**(1): p. 69-79.
37. Geuken, E., et al., *Rapid increase of bile salt secretion is associated with bile duct injury after human liver transplantation.* J Hepatol, 2004. **41**(6): p. 1017-25.
38. Wagner, M., et al., *Role of farnesoid X receptor in determining hepatic ABC transporter expression and liver injury in bile duct-ligated mice.* Gastroenterology, 2003. **125**(3): p. 825-38.
39. Bolondi, G., et al., *Predictive factors of short term outcome after liver transplantation: A review.* World J Gastroenterol, 2016. **22**(26): p. 5936-49.
40. Kochanek, K.D., et al., *Deaths: final data for 2017.* 2019.
41. Ramaiah, S.K., *A toxicologist guide to the diagnostic interpretation of hepatic biochemical parameters.* Food Chem Toxicol, 2007. **45**(9): p. 1551-7.
42. Ozer, J., et al., *The current state of serum biomarkers of hepatotoxicity.* Toxicology, 2008. **245**(3): p. 194-205.

43. Boin, I.F., et al., *Liver transplant recipients mortality on the waiting list: long-term comparison to Child-Pugh classification and MELD.* Transplant Proc, 2004. **36**(4): p. 920-2.
44. Kamath, P.S., et al., *A model to predict survival in patients with end-stage liver disease.* Hepatology, 2001. **33**(2): p. 464-70.
45. Wiesner, R., et al., *Model for end-stage liver disease (MELD) and allocation of donor livers.* Gastroenterology, 2003. **124**(1): p. 91-6.
46. Kamath, P.S., W.R. Kim, and G. Advanced Liver Disease Study, *The model for end-stage liver disease (MELD).* Hepatology, 2007. **45**(3): p. 797-805.
47. Lau, T. and J. Ahmad, *Clinical applications of the Model for End-Stage Liver Disease (MELD) in hepatic medicine.* Hepat Med, 2013. **5**: p. 1-10.
48. Cholongitas, E., et al., *MELD is not enough--enough of MELD?* J Hepatol, 2005. **42**(4): p. 475-7; author reply 478-9.
49. Freeman, R.B., *MELD: the holy grail of organ allocation?* J Hepatol, 2005. **42**(1): p. 16-20.
50. Nusrat, S., et al., *Cirrhosis and its complications: evidence based treatment.* World J Gastroenterol, 2014. **20**(18): p. 5442-60.
51. Huang, W.M., et al., *Intrahepatic cholestasis of pregnancy: detection with urinary bile acid assays.* J Perinat Med, 2007. **35**(6): p. 486-91.
52. Muraji, T., et al., *Urinary sulfated bile acid concentrations in infants with biliary atresia and breast-feeding jaundice.* Pediatr Int, 2003. **45**(3): p. 281-3.
53. Sinakos, E. and K.D. Lindor, *Bile acid profiles in intrahepatic cholestasis of pregnancy: is this the solution to the enigma of intrahepatic cholestasis of pregnancy?* Am J Gastroenterol, 2010. **105**(3): p. 596-8.
54. Ambros-Rudolph, C.M., et al., *The importance of serum bile acid level analysis and treatment with ursodeoxycholic acid in intrahepatic cholestasis of pregnancy: a case series from central Europe.* Arch Dermatol, 2007. **143**(6): p. 757-62.
55. Heikkinen, J., *Serum bile acids in the early diagnosis of intrahepatic cholestasis of pregnancy.* Obstet Gynecol, 1983. **61**(5): p. 581-87.
56. Shinohara, T., et al., *Efficacy of urinary sulfated bile acids for diagnosis of bacterial cholangitis in biliary atresia.* Pediatr Surg Int, 2005. **21**(9): p. 701-4.
57. Trottier, J., et al., *Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing cholangitis: a pilot study.* Dig Liver Dis, 2012. **44**(4): p. 303-10.
58. Simko, V., S. Michael, and R.E. Kelley, *Predictive value of random sample urine bile acids corrected by creatinine in liver disease.* Hepatology, 1987. **7**(1): p. 115-21.
59. Stiehl, A., et al., *Biliary and urinary excretion of sulfated, glucuronidated and tetrahydroxylated bile acids in cirrhotic patients.* Hepatology, 1985. **5**(3): p. 492-5.
60. Dasarathy, S., et al., *Elevated hepatic fatty acid oxidation, high plasma fibroblast growth factor 21, and fasting bile acids in nonalcoholic steatohepatitis.* Eur J Gastroenterol Hepatol, 2011. **23**(5): p. 382-8.
61. Simko, V. and S. Michael, *Urinary bile acids in population screening for inapparent liver disease.* Hepatogastroenterology, 1998. **45**(23): p. 1706-14.
62. Takikawa, H., et al., *Glucuronidated and sulfated bile acids in serum of patients with acute hepatitis.* Dig Dis Sci, 1986. **31**(5): p. 487-91.
63. Bezerra, J.A., et al., *Biliary Atresia: Clinical and Research Challenges for the Twenty-First Century.* Hepatology, 2018. **68**(3): p. 1163-1173.
64. Meng, L.J., et al., *Profiles of bile acids and progesterone metabolites in the urine and serum of women with intrahepatic cholestasis of pregnancy.* J Hepatol, 1997. **27**(2): p. 346-57.

65. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases I. Stability of the baseline profile in healthy subjects.* Toxicol Sci, 2015. **143**(2): p. 296-307.
66. Thakare, R., et al., *Species differences in bile acids I. Plasma and urine bile acid composition.* J Appl Toxicol, 2018. **38**(10): p. 1323-1335.

CHAPTER 2

URINARY BILE ACIDS AS DIAGNOSTIC BIOMARKERS FOR LIVER DISEASES

2.1. Introduction

Bile acids (BA), the end products of cholesterol metabolism, are synthesized in liver and excreted into bile, which flows via the bile duct to the small intestine. BA have many physiological functions such as cholesterol absorption and elimination, fat absorption, and maintenance of healthy microbiome [1, 2]. In addition, BA have been identified as signalizing molecules/hormones that exert their functions at the molecular level, via binding to numerous receptors, primarily, the surface G-protein-coupled membrane receptor (TGR5), and the nuclear farsenoid-X-receptor (FXR) [3]. In that capacity, BA are involved in the regulation of their own homeostasis, glucose and lipid metabolism, energy expenditure, cellular immunity, and thyroid hormone signaling [2, 4, 5].

In contrast to their vital physiological functions, BA are also cytotoxic and exhibit pathological effects at high concentrations due to their direct detergent effects on biological membranes, cancer promoting effects, and necrotic and apoptotic effects via mitochondria and endoplasmic reticulum-mediated toxicities [5, 6]. BA have deleterious effects on the liver including cholestasis and alterations in liver structure and hepatocyte ultrastructure in several animal species [7, 8].

Cholestatic liver diseases is a diverse group of hepatobiliary diseases associated with a reducing in bile flow due to impairment in bile production or failure of bile flow through the canaliculi into bile duct [9]. The impediment in bile flow cause the accumulation of BA in the liver, which spills out into the systemic circulation and extrahepatic tissues and eventually into urine. Numerous clinical and preclinical studies have shown up to a 100-fold increase in BA concentrations in the blood and urine during various liver diseases [8, 10-13].

There is ample evidence from animal and human studies to indicate that accumulation of BA in the liver, systemic blood, and extrahepatic tissues can worsen the liver condition that lead to their accumulation. The accumulation of toxic BA in cholestasis induces hepatotoxicity, extrahepatic toxicity such as encephalopathy, which may contribute to the unfavorable liver disease prognosis [14]. For example, BA concentrations were shown to correlate with the progression of damages to the liver and bile duct in cholestatic rats, rabbits, and in humans [15-18]. In addition, several animal studies indicate that the accumulation of toxic BA is linked to the occurrence and severity of complications after liver transplantation. Other studies shown that patients with low concentrations of BA are less likely to develop hepatobiliary complications after transplantation [17].

The most commonly used biomarkers for the diagnosis of liver diseases include alanine transaminase (ALT), aspartate transaminase (AST), alkaline phosphatase (ALP), λ-glutamyltransferase (GGT), serum creatinine, protime, international normalized ratio (INR), albumin and bilirubin [19, 20]. However, these serum markers are not specific to the liver or bile duct injuries and may increase in non-hepatobiliary diseases. Also, they require severe cell injury at advanced disease stages before their levels increase in the blood [19, 20]. Multifactorial models with multiple parameters based on these biomarkers are also frequently used and offer advantages compared to the use of their individual biomarker components such as the Child-Turcotte-Pugh (CTP) score, and model for end-stage liver disease (MELD) score. Further diagnosis and assessment of liver disease prognosis is empirical and rely on invasive and risky endoscopic procedures and histological evaluation of liver biopsies [21].

Because of their known hepatotoxicity and established accumulation in liver diseases, BA were extensively investigated for decades as biomarkers for numerous

hepatobiliary diseases [13, 22-24] including intrahepatic cholestasis of pregnancy [25, 26], biliary atresia [23, 27], primary biliary cirrhosis [28], primary sclerosing cholangitis [28], alcoholic liver diseases [29, 30], nonalcoholic fatty liver disease [31], and viral hepatitis [32, 33]. Currently, BA are clinically used as biomarkers for the diagnosis of intrahepatic cholestasis of pregnancy and biliary atresia in infants [22, 24, 34].

Despite these extensive efforts the potential use of BA as a marker for liver diseases have never translated into the clinic due to major limitation, which prevented them from being used as a reliable biomarker of hepatobiliary diseases in the clinic. First, due to the marked differences in the physiological and pathological properties of the various BA, detailed profiling of the more toxic and relevant individual BA rather than total BA concentration may better correlate with the liver condition during hepatobiliary diseases [10, 12, 35]. More importantly, the extreme inter-and intra-individual variability of total and individual BA concentrations makes it very hard to determine the normal baseline ranges in the absence of liver diseases. BA have marked inter- individual variability due to many factors such as gender, alcohol consumption, and weight [52]. In addition, high intra-individual variability of serum and urinary BA levels was reported due to numerous factors such as medication intake, food ingestion, and diurnal variation [30].

In an effort to address these limitations, we have developed the concept of “BA Indices”. BA indices are ratios calculated from the absolute concentration of individual BA and their metabolites (**Table 2.1).** These ratios quantify in detail, the composition, metabolism, hydrophilicity, formation of secondary BA, and toxicity of the BA profile [9, 36, 37]. Because BA indices are ratios, we proposed that BA indices should have lower variability than the absolute BA concentrations used to calculate them. Indeed, we have demonstrated that BA indices offered numerous advantages over absolute total and individual BA concentrations including low inter- and intra-individual variability and were

resistant to covariate influences such as age, gender, BMI, food consumption, and moderate alcohol consumption [9, 36]. For example, the relative standard deviation (RSD) from 66-256% for total and individual absolute BA concentrations in serum, while it was less than 100% and as low as 13% for BA indices in the same healthy controls' population. After food ingestion, serum BA rise due to the release of cholecystokinin, that stimulates gallbladder contraction resulting in increasing bile flow into the intestine [36]. Therefore, feeding status should be controlled before serum BA can be utilized as a reliable biomarker. We have reported that one hour after a standardized meal, the absolute total and most of the individual BA concentrations more than doubled and increased more than 10-fold in some individuals, while BA indices changed by only 10-40% in the same individuals [36]. The same trend was also observed in urine, where BA indices had much smaller variability with as low as 10% RSD and were resistant to feeding status compared to absolute BA concentrations in the same population [36]. Therefore, noninvasive urinary BA indices have a better potential in the clinic than the absolute serum or urine or serum BA concentrations as biomarkers for hepatobiliary diseases. Furthermore, we have showed that BA indices may outperform currently used liver enzymes as biomarkers such as AST and ALT for the diagnosis of cholestatic hepatobiliary diseases in humans [9].

In this study, we have expanded on our previous pilot study, where we have recruited 300 patients with liver diseases and 103 control subjects over a period of 7 years. We have compared the urinary BA profile between healthy controls and patients with hepatobiliary diseases. In addition, the BA profile was compared between patients with different severity levels of liver disease as determined by mayo model for end-stage liver disease (MELD) score and disease compensation status. The utility of BA indices as biomarkers was determined using various statistical tests and was compared with the currently used liver enzyme tests in clinic for diagnosing a liver disease.

2.2. Materials and methods

2.2.1. Study participants

For controls, 103 healthy subjects without liver diseases (71 female and 32 male) between the ages of 19 and 65 years were recruited by the Clinical Research Center at the University of Nebraska Medical Center (UNMC) (Omaha, NE, USA). The study was approved by the institutional review board (IRB) at UNMC and written informed consents were provided for all participating subjects. Inclusion criteria included the absence of diabetes, normal liver functions, as verified by serum liver enzymes levels of aspartate aminotransferase (AST) <56 U/L, alanine aminotransferase (ALT) <50 U/L, and gamma-glutamyl transferase (GGT) <78 U/L, and no- or moderate alcohol drinking. According to the "2010 Dietary Guidelines for Americans" by the Department of Agriculture and the Department of Health and Human Services (http://www.cnpp.usda.gov/Publications/DietaryGuidelines/2010/PolicyDoc/PolicyDoc.pdf), moderate drinking is defined as the consumption of alcoholic beverages up to 2 drinks/ day for men and up to 1 drink/ day for women. Standard drinks contain the same amount of alcohol and are defined as 1.5 ounces of liquor, 5 ounces of wine, or 12 ounces of beer.

Urine samples were collected from controls at fasting conditions over multiple visits. Urine samples were initially collected in the first visit at fasting conditions. Samples were then collected also at fasting conditions, one (second visit), two (third visit), and four (fourth visit) weeks after the first visit. Thirty milliliters of urine were collected from these subjects in each visit.

For patients, new and existing patients of the UNMC hepatology clinic, who were diagnosed with one or multi-hepatobiliary conditions due to chronic hepatitis C (n=71) , hepatitis B (n=15), alcoholic liver disease (n=116), primary biliary cholangitis (PBC) (n=12), primary sclerosing cholangitis (PSC) (n=17), autoimmune hepatitis (n=26), alpha-

1-antitrypsin deficiency (n=6), nonalcoholic fatty liver disease/nonalcoholic steatohepatitis (NASH) (n=56), carcinoma (n=25), cryptogenic cirrhosis (n=11), polycystic liver disease (n=5), elevated liver function test (LFT) (n=21), and unknown etiology (n=5), were enrolled in this study. A total of 300 patients (143 female and 157 male) between the ages of 19 and 83 years were recruited into the study after signing an informed consent. Thirty milliliters of urine samples were collected from patients on their first visit and also their follow-up visits to the hepatology clinic. All urine samples were stored in -80 °C until analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). To study the association between the BA profile and the severity of the liver disease, patients were divided into three groups based on their model for end-stage liver disease (MELD) score: low- MELD (6–15 score), medium- MELD (16–25), and high MELD (26–40). As there were only four subjects who had a high MELD score in the current study, this group was not included while performing the statistical analysis. In addition, patients were also categorized according to the compensation status (compensated patients and patients with decompensated liver diseases) as diagnosed by the hepatologists. Decompensated patients have severe complications including ascites, encephalopathy, bleeding varices, or jaundice, while compensated patients did not have any of these complications.

2.2.2. Non-BA parameters

The performance of potential biomarkers from the urinary BA profile has also been compared with the performance of existing markers of liver function including alanine transaminase (ALT), aspartate transaminase (AST), serum creatinine, albumin, bilirubin, international normalized ratio (INR), protime, AST/ ALT ratio, and AST/ platelet ratio index (APRI).

For healthy controls, total bilirubin in serum was analyzed using QuantiChrom^{TM} Bilirubin assay kit (BioAssay Systems, Hayward, California). ALT, AST, serum creatinine,

and albumin were measured using the Beckman Coulter reagents (Beckman Coulter, Inc, Brea, California). INR and protime were measured using STANeoplastine “CI PLUS 10” reagent kit (Diagnostica Stago Inc, Parsippany, New Jersey). For patients, the same information was extracted from their health records after providing a written consent to participate in the study.

2.2.3. Bile acid quantification by liquid chromatography–tandem mass spectrometry

BA concentrations were quantified by LC–MS/MS, as we described previously with some modifications [37-39]. A Waters ACQUITY ultra performance liquid chromatography (UPLC) system (Waters, Milford, MA, USA) coupled to an Applied Biosystem 4000 Q TRAP® quadrupole linear ion trap hybrid mass spectrometer with an electrospray ionization (ESI) source (Applied Biosystems, MDS Sciex, Foster City, CA, USA) was used to perform the LC-MS/MS analysis. All chromatographic separations were performed with an ACQUITY UPLC® BEH C_{18} column (2.1x 150 mm, 1.7 µm) equipped with an ACQUITY UPLC C_{18} guard column (Waters, Milford, MA, USA). The following MS source settings were used: temperature, 500°C; ion spray voltage, −4000 V; collision gas pressure, high; curtain gas, 20; gas-1, 35; gas-2 35 (arbitrary units); Q1/Q3 resolution, unit; and interface heater, on. Mobile phase consisted of 7.5 mM ammonium bicarbonate, have been adjusted to pH 9.0 by using ammonium hydroxide (mobile phase A) and 30% acetonitrile in methanol (mobile phase B) at a total flow rate of 0.2 ml/min. The gradient profile was held at 52.5% mobile phase B for 12.75 minutes, increased linearly to 68% in 0.25 minutes, held at 68% for 8.75 minutes, increased linearly to 90% in 0.25 minutes, held at 90% for one minute and finally brought back to 52.5% in 0.25 minutes and then followed by 4.75 minutes re-equilibration (total run time of 28 minutes per sample).

2.2.4. Measurement and Calculation of BA indices

We measured the following BA: Cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA), lithocholic acid (LCA), ursodeoxycholic acid (UDCA), tauro-cholic acid (T-CA), tauro-chenodeoxycholic acid (T-CDCA), tauro-deoxycholic acid (T-DCA), tauro-lithocholic acid (T-LCA), tauro-ursodeoxycholic acid (T-UDCA), glyco-cholic acid (G-CA), glyco-chenodeoxycholic acid (G-CDCA), glyco-deoxycholic acid (G-DCA), glyco-lithocholic acid (G-LCA), glyco-ursodeoxycholic acid (G-UDCA), β-muricholic acid (MCA), tauro-β-muricholic acid (T-MCA), hyocholic acid (HCA), glyco-hyocholic acid (G-HCA), tauro-hyocholic acid (T-HCA), hyodeoxycholic acid (HDCA), glyco-hyodeoxycholic acid (G-HDCA), tauro-hyodeoxycholic acid (T-HDCA), murideoxycholic acid (MDCA), isolithocholic acid (isoLCA), isodeoxycholic acid (isoDCA), 7-oxoLCA, 12-oxoLCA, 12-oxoCDCA, 3-dehydroCA, norDCA, LCA-sulfate (LCA-S), UDCA-sulfate (UDCA-S), CDCA-sulfate (CDCA-S), DCA-sulfate (DCA-S), CA-sulfate (CA-S), glyco-lithocholic acid-sulfate (G-LCA-S), glyco-ursodeoxycholic acid-sulfate (G-UDCA-S), glyco-chenodeoxycholic acid-sulfate (G-CDCA-S), glyco-deoxycholic acid-sulfate (G-DCA-S), glyco-cholic acid-sulfate (G-CA-S), tauro-lithocholic acid-sulfate (T-LCA-S), tauro-ursodeoxycholic acid-sulfate (T-UDCA-S), tauro-chenodeoxycholic acid-sulfate (T-CDCA-S), tauro-deoxycholic acid-sulfate (T-DCA-S), and tauro-cholic acid-sulfate (T-CA-S).

In addition to the absolute concentration of individual and total BA, the BA profile in urine was characterized using BA “indices”, as we have described previously [36, 38]. **Table 2.1** lists the BA indices that were evaluated in the current study. BA indices describe the composition, hydrophilicity, formation of 12α-OH BA by CYP8B1, metabolism, and formation of secondary BA by intestinal bacteria. The composition indices were calculated as the ratio of the concentration of individual BA in all of their forms (unamidated, amidated, unsulfated and sulfated) to the total concentration of BA. Hydrophilicity indices

include the percentages of the BA pool exist as mono-, di-, or tri-OH BA as well as the hydrophobicity index (HI) of the BA pool. The percentages of mono-OH BA (LCA), di-OH BA (UDCA, MDCA, HDCA, DCA and CDCA) and tri-OH BA (CA, MCA and HCA) were calculated as the ratio of the concentration of the sum of the respective BA in all their forms to the total concentration of BA. HI was calculated according to the Heuman index, which based on the relative contributions of the individual BA to the total BA pool and their HIs [40].

12α-OH BA are formed by CYP8B1 in the liver and include DCA, CA, Nor-DCA, and 3-dehydroCA.Therefore, CYP8B1 activity can be measured by the ratio of 12α-OH BA to the remaining of all other BA (non-12α-OH BA). Another marker for CYP8B1 is the ratio of CA to CDCA because CA is formed by the 12α hydroxylation of CDCA. In the same way, the ratio of 12α-OH (DCA, CA, Nor-DCA, and 3-dehydroCA in all of their forms) to non-12α-OH (HDCA, CDCA, UDCA, LCA, MDCA, MCA, HCA, 12-oxo-CDCA , 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, isoDCA in all of their forms) was calculated.

BA are primarily metabolized by sulfation, and glycine (G), and taurine (T) amidation in the liver. The percentage of sulfation of individual BA was calculated as the ratio of the concentration of sulfated BA, in both the unamidated and amidated forms, to the total concentration of individual BA in all of their forms (unamidated, amidated, unsulfated, and sulfated). The percentage of amidation of individual BA was calculated as the ratio of the concentration of amidated BA, in both the unsulfated and sulfated forms, to the total concentration of individual BA in all of their forms (unamidated, amidated, unsulfated, and sulfated). In addition, percentages of amidation were divided into the percentages of BA existing as taurine (T) or as glycine (G) amidates.

Primary BA are synthesized in the liver and secreted into the intestine via bile, where they are metabolized by intestinal bacteria into secondary BA. The ratio of primary (CA, CDCA, MCA and HCA in all of their forms) to secondary BA (DCA, LCA, UDCA, HDCA, MDCA, Nor-DCA, 12-oxo-CDCA, 3-dehydroCA, 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, and isoDCA in all of their forms) was calculated.

2.2.5. Preparation of standard solutions and calibration curves

For the preparation of standard solutions and calibration curves, blank matrices were obtained by charcoal stripping as described previously [9, 36-39]. Briefly, Stock solutions of individual unsulfated BA and the IS ($^2H_4$-G-CDCA) were prepared in methanol (MeOH) at a concentration of 10 mg/mL and stock solutions of individual sulfated BA were prepared in deionized water at a concentration of 1 mg/mL. Human urine was incubated with 100 mg/mL activated charcoal for two hours to remove endogenous BA from the matrix. The mixture was then centrifuged at 13000 x g for 10 min, and the supernatant was aspirated and filtered using a 0.22-µm nylon filter. The filtrate from the stripped urine matrix was used for preparing the calibration curve. Eleven-point calibration curve was prepared by spiking 10 µL of the appropriate standard solutions and 10 µL of the IS stock ($^2H_4$-G-CDCA) into 100 µL of the stripped urine matrices. The final concentration of IS was 500 ng/ml and the dynamic range of the standard curves for the various unsulfated and sulfated BA analytes was 1-1000 ng/ml.

2.2.6. Sample preparation

Urine samples were extracted using solid phase extraction as described previously [9, 36-39]. Briefly, 100 µl of urine samples were spiked with 10 µl of internal standard (IS), vortexed and loaded on to Supelclean™ LC-18 SPE cartridges pre-conditioned with 4 ml MeOH, followed by 4 ml H_2O. Loaded cartridges were then washed with 3 mL H_2O and eluted with 4 mL MeOH. The eluates were evaporated under vacuum at room temperature

and reconstituted in a 100 µL of 50 % MeOH solution. Ten microliters of reconstituted samples were injected for LC-MS/MS analysis.

2.2.7. Statistical analysis

Independent sample-t-test and Mann-Whitney test were used to study the demographic differences between controls and patients because the sample size was >30 [41]. Independent sample-t-test was used for continuous variables and Mann-Whitney test was used for categorical variables. The demographic variables were (age, BMI, gender, and race). Subjects were divided into four age groups (19-29, 30-41, 42-53, 54-83 years), and the variable age was studied as both a categorical and a continuous variable. Subjects were also divided into three BMI groups (normal: BMI < 25, overweight: BMI 25–29.9, and obese: BMI ≥ 30) and the effect of BMI was studied as both a categorical and a continuous variable. Also, subjects were divided into five race groups (White, Black, Asian, Hispanic, Others), and the variable race was studied as a categorical variable.

Urine samples were collected from controls and patients on their first visit and follow-up visits. Mixed effects models were used to compare patients vs. controls and the demographic variables were included as covariates. Statistically significant covariates were returned to the mixed effects models as interaction terms with the primary group, i.e. patients vs. control.

BA indices were compared between controls, low- MELD (patients), and medium-MELD (patients) groups using mixed effects models followed by pairwise comparisons using Bonferroni's adjustment if the p-value was <0.05. BA indices were compared between compensated and decompensated patients using mixed effects models. Mixed effects models were also used to determine the association between non-BA parameters including (creatinine, ALT, AST, MELD score, APRI, AST/ALT, bilirubin, albumin, INR, and

protime) and BA indices. Receiver operating characteristic curve (ROC) analyses were used to determine cut-off values of BA as markers for the diagnosis of liver diseases with optimum sensitivity and specificity. The areas under the ROC curve (AUC) values were compared between urinary BA profiles and non-BA parameters. The mixed effects models were used to compare BA indices with AUC>0.7 between controls and the patients with specific disease subtypes including chronic hepatitis C, hepatitis B, Laennec/alcoholic cirrhosis, primary biliary cholangitis, primary sclerosing cholangitis, autoimmune hepatitis, alpha-1-antitrypsin deficiency, nonalcoholic fatty liver disease/ nonalcoholic steatohepatitis, carcinoma, cryptogenic cirrhosis, polycystic liver disease, elevated liver function test, and unknown etiology.

Univariate logistic regression analysis was used to determine the association between BA profiles and the likelihood of developing a liver disease. From the logistic regression analysis, odds ratios (ORs) were calculated for a 10% and 20% change from the mean value of BA indices in the healthy controls.

P-value of 0.05 was considered significant for all the statistical tests described above. All statistical analysis was performed using the Statistical Product and Service Solutions (SPSS) software, version 25 (IBM corporation, Armonk, NY, USA).

2.3 Results

2.3.1.1. Demographics

Table 2.2 shows a summary of the demographics of both patients and controls participants. We enrolled 103 controls (32 males and 71 females) and 300 patients (157 males and 143 females), who were treated for cholestatic liver diseases in the University of Nebraska Medical Center (UNMC), over the period from November of 2011 - December of 2018. To compare the demographics between the two groups, age and BMI covariates were compared as both continuous and categorical variables using t-test, and mann-whitney test, respectively. While gender and race were compared as categorical variables using mann-whitney test. Age, Gender, and BMI were significantly different between control and patients (p-value < 0.05), while race was not different. Therefore, the statistically significant demographic variables (age, BMI, and gender) were included as covariates in the mixed effects models to compare BA indices between patients and controls.

2.3.1.2. Differences in BA between patients vs. controls are not due to differences in demographics

Table 2.3 shows the Association between demographics covariates and BA indices in controls and patients. Because some of the covariates (age, BMI, and gender) were significantly different between the two groups (**Table 2.2**), we reran the univariate mixed effect analysis with these covariates (multivariate analysis). First, association between these covariates and BA indices was identified, and then the covariates with significant association with BA indices were incorporated in the multivariate mixed effect analyses as interaction terms with the group (patients and controls). This way, we detect any association between these covariates and BA indices, and if this relationship exists, we verify if this relationship is similar between the two groups. The overall goal is to ensure

that differences in the primary variables (BA indices) between the groups are not due to differences in the demographic covariates. We did not find any difference in the association between covariates and BA indices between the two groups except for the % primary and % secondary BA with gender.

For example, gender was significantly different between patients vs. controls (**Table 2.2**: 71 out of 103 vs. 143 out of 300) and was associated with some BA indices such as % LCA ($p = 0.047$ in **Table 2.3**). However, this association between % LCA and gender was not different in the patient vs. control groups ($p = 0.682$ in **Table 2.3**). Therefore, despite the differences in gender between the groups and the association between this demographic covariate and some BA indices, the difference in these BA indices between the groups is not due to this covariate.

2.3.2. BA profiles in controls vs. patients

Table 2.4 shows the absolute concentrations of major urinary BA in controls and patients. **Table 2.5** compares representative absolute BA concentrations and indices between controls and patients. **Appendix Table A** shows the full list of BA concentrations and indices. BA indices were compared between patients vs. controls using univariate mixed effects models. Total BA was 5.9-fold higher in patients compared with controls. All individual BA concentrations were higher in patients, except MDCA, but to different extents. The highest increase was in UDCA (11.9-fold), while the lowest increase was for DCA and HDCA (1.6-fold). The percentage of UDCA, CDCA, CA, MCA, and HCA were higher (1.21-1.58-fold), while the percentage of LCA, DCA, HDCA, and MDCA were lower (0.47-0.79 fold) in patients vs. controls.

Unamidated, G-amidated, and T-amidated BA which were 3.3, 5.9, and 9.4-fold were higher in patients than controls. Therefore, the overall % amidation and %G-amidation

did not change or slightly decreased in patients, whereas % T-amidation increased from 8.0 % in controls to 10.8% in patients. Similarly, the concentrations of both sulfated and unsulfated were ~ 6-fold higher in patient; so that the % sulfation of BA was unchanged or slightly lower (0.94-fold) in patients.

The absolute concentrations of mono-, di-, and tri–OH BA were also higher in patients compared with controls, but the % mono-OH decreased (0.8-fold), di-OH remained unchanged, and % tri-OH increased (1.4-fold) due to increasing % CA (1.2-fold), % MCA (1.6-fold), and % HCA (1.5-fold).

Total 12α-OH and non-12α-OH BA were 2.3-fold and 8.5-fold higher in patients, so that the ratio of 12α-OH/ non-12α-OH was 0.5-fold lower, and the % 12α-OH decreased (0.60-fold), while % non-12α-OH BA increased (1.2-fold) in patients. On the other hand, there was no difference in the ratio of CA/CDCA between controls and patients.

Total primary (CDCA, CA, MCA and HCA in all of their forms) and secondary BA (DCA, LCA, UDCA, HDCA, MDCA, Nor-DCA, 12-oxo-CDCA, 3-dehydroCA, 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, and isoDCA in all of their forms) were 8.1-fold and 4.6-fold higher in patients, so that the ratio of primary/ secondary BA was 3.6-fold higher in patients. Therefore, % primary BA was 1.4-fold higher, while % secondary BA was 0.80-fold lower in patients vs. controls.

Hydrophobicity index (HI) of the urinary BA pool was -0.12 in patients vs. -0.01 in controls.

2.3.3. BA profile in low vs. medium- MELD patients

Table 2.6 compares the representative urinary BA concentrations and indices between low- and medium- MELD patients. Total BA concentrations was twice as high in patients with medium- MELD and individual BA concentrations were (1.15 - to 3.9-fold)

higher except for UDCA, HDCA, and MCA in medium vs. low- MELD patients. Percentage CDCA, CA, and HCA were ~1.5-fold higher, while the percentage of all other individual BA were lower in the medium compared with the low- MELD patients.

Unamidated BA concentration was 0.7-fold lower, while G-amidated and T-amidated BA were 1.9 and 3.2-fold higher in the medium- MELD patients. Therefore, % T-amidation was 1.5-fold higher, while there was no difference in the % amidation and % G- amidation between medium and low- MELD patients.

Similarly, the concentrations of both sulfated and unsulfated were 1.3 and 2-fold higher in medium vs. low- MELD. On the other hand, the % sulfation of BA was only 1.07-fold higher, but it was statistically significant.

The absolute concentrations of mono-, di-, and tri–OH BA were also (1.8-2-fold) higher in medium- MELD patients, but the % mono-OH decreased (0.86-fold); while % di- and % tri-OH remained unchanged.

Total 12α-OH and non-12α-OH BA were 2-fold higher in medium vs. low- MELD patients, while that the ratio of 12α-OH/ non-12α-OH was 0.68-fold lower, but was not statistically significant. On the other hand, % 12α-OH decreased (0.77-fold) and % non-12α-OH BA remained unchanged, while the ratio of CA/CDCA was 0.8-fold lower in medium- MELD patients.

Total primary BA were 3.4-fold higher, while total secondary BA were slightly (0.9-fold) lower in medium- MELD patients, so that the ratio of primary/ secondary BA was 2.3-fold higher in medium- MELD patients. Similarly, % primary BA was 1.4-fold higher, while % secondary BA was 0.6-fold lower in medium- MELD patients.

The hydrophobicity index (HI) of the urinary BA pool was -0.05 in medium vs. -0.15 in low- MELD patients.

2.3.4. BA profile in compensated vs. decompensated patients

Patients were categorized according to the presence or absence of hepatic decompensation as diagnosed by the hepatologists. Hepatic decompensation defined as the presence or history of encephalopathy, bleeding varices, ascites, or jaundice, while compensated patients did not have any of these complications. **Table 2.7** compares representative urinary BA concentrations and indices between compensated and decompensated patients. BA indices were compared between compensated vs. decompensated patients using mixed effects models. Total BA was 1.3-fold higher in decompensated compared with compensated patients. All individual BA concentrations were higher in decompensated patients, except UDCA, DCA, MDCA, and MCA. The highest increase was in HCA (3.2-fold), while the lowest increase was for LCA (1.14-fold), which was statistically insignificant.

The percentage of CDCA, HDCA, CA, and HCA were higher (1.3 - 2.1-fold), while the percentage of LCA, UDCA, DCA, and MDCA were lower (0.3 - 0.7-fold) in decompensated vs compensated patients. On the other hand, there was no difference in the percentage of MCA between decompensated and compensated patients.

Unamidated BA concentration was 0.9-fold lower, while G-amidated, and T-amidated BA were 1.3 and 2.1-fold higher in decompensated patients. Therefore, % T-amidation was 1.3-fold higher in decompensated patients, while there was no difference in the % amidation and % G- amidation between decompensated and compensated patients. Similarly, the concentrations of both sulfated and unsulfated were 1.3 and 1.2-

fold higher, but the % sulfation of BA was unchanged or slightly higher (1.03-fold) in decompensated patients.

The absolute concentrations of mono-, di-, and tri–OH BA were also higher in decompensated patients, but the % mono-OH decreased (0.73-fold), % di-OH remained unchanged, and % tri-OH slightly increased (1.13-fold) due to increasing % CA and % HCA, and was not statistically significant.

Total 12α-OH and non-12α-OH BA were 1.14-fold and 1.34-fold higher in decompensated patients, so that the ratio of 12α-OH/ non-12α-OH was 0.7-fold lower, whereas % 12α-OH decreased (0.8-fold) and % non-12α-OH BA remained unchanged. On the other hand, the ratio of CA/CDCA was 0.8-fold lower in decompensated patients.

There was a significant difference in total primary BA, which were two-fold higher, while total secondary BA were 0.8-fold lower in decompensated than compensated patients, so that the ratio of primary/ secondary BA was 2.6-fold higher in decompensated patients. Therefore, % primary BA was 1.5-fold higher, while % secondary BA was 0.56-fold lower in decompensated patients.

The hydrophobicity index (HI) of the urinary BA pool was -0.13 in decompensated vs. -0.18 in compensated patients.

2.3.5. Receiver operating characteristic (ROC) curve analysis

Table 2.8 lists the area under the ROC curve (AUC) for BA concentrations and indices. **Appendix Table B** shows the full list of BA concentrations and indices. Total BA, CDCA, CA, % DCA, % HDCA, % MDCA, total G-Amidated, total unsulfated, total sulfated, total di-OH, total tri-OH, total non-12α-OH, 12α-OH/ non12α-OH, % 12α-OH, % non-12α-OH, total primary, primary/ secondary, % primary, and % secondary produced AUC > 0.7. **Figure 2.1** shows ROC curves of representative BA indices with AUC > 0.7. Potential cut-

off values selected based on the optimum specificity and sensitivity for BA indices with AUC > 0.7 are listed in **Table 2.9**.

2.3.6. Risk analysis: logistic regression analysis

Table 2.10 shows the results of logistic regression analyses for BA indices with ROC (AUC) > 0.7. Logistic regression analysis detects if there is a risk of developing a hepatobiliary disease associated with changes in BA indices. The risk of being diagnosed with a hepatobiliary disease increased with changing levels of all BA indices ($P<0.05$) except (% HDCA and % MDCA). In addition, the odds ratio (OR) from logistic regression analysis quantifies the magnitude of the risk of developing liver diseases per unit (10% and 20% of the normal value) changes in BA indices. For example, for every 20% increase in the % non-12α-OH BA, the likelihood of having a liver disease increases 2.72-folds (Odds ratio [OR]: 2.72; $P<0.05$). In contrast for every 20% increase in the % 12α-OH BA, the likelihood of having a liver disease decreases 0.56-folds (Odds ratio [OR]: 0.56; $P<0.05$).

2.3.7. BA profile in different liver disease subtypes

Table 2.11.A and **Table 2.11.B** compare BA indices with ROC-AUC > 0.7 between controls vs. patients with specific liver disease subtype. Mixed effects models were used to compare disease subtypes individually vs. controls. The goal was to identify BA indices that can serve as diagnostic biomarkers for specific liver disease subtypes.

These specific liver disease subtypes include chronic hepatitis B, hepatitis C, Laennec/alcoholic cirrhosis, primary biliary cholangitis, autoimmune hepatitis, alpha-1-antitrypsin deficiency, primary sclerosing cholangitis, nonalcoholic fatty liver disease/nonalcoholic steatohepatitis (NASH), carcinoma, cryptogenic cirrhosis, polycystic

liver disease, elevated liver function test, and unknown etiologies as diagnosed by the hepatologists.

We have found that most BA indices were significantly different between controls vs. all individual liver disease subtypes. Total BA, total CDCA, total CA, total G-amidated, total unsulfated, total sulfated, total di-OH, total tri-OH, Total non-12α-OH, % non-12α-OH and total primary were higher (1.1- to 39.5-fold) in every liver disease group compared with controls except in polycystic liver disease. % Primary and primary/ secondary were higher (1.1- to 9.27-fold) in all liver disease group compared with controls except in primary biliary cholangitis. % DCA, %HDCA, % 12α-OH, and 12α-OH/ non12α-OH were lower (0.07- to 0.85-fold) in every liver disease group compared with controls. % MDCA and % Secondary were lower in all liver disease group compared with controls except in elevated LFT and primary biliary cholangitis, respectively.

2.3.8. Non-BA parameters

In addition to BA indices, we have also examined other biomarkers currently used in the clinic to evaluate liver functions. These non-BA parameters include AST, ALT, AST/ALT, bilirubin, albumin, INR, protime, creatinine, APRI, and MELD. **Table 2.12** compares the non-BA parameters in controls and patients using mixed effects models. All the non-BA parameters were higher in patients compared to controls except albumin and protime, which were lower in patients. Within the patient population, all non-BA parameters were higher in medium compared to low- MELD patients except albumin, and ALT. The same results also applied to decompensated vs. compensated patients.

The area under the ROC curve (AUC) for non-BA parameters was >0.7 for all of them except creatinine, protime, and AST/ALT ratio. Also, per logistic regression analysis, the risk of being diagnosed with a liver disease increased to various extents with changing levels of all non-BA parameters ($P<0.05$) except creatinine and AST/ ALT. For example,

for every 20% increase in the albumin and protime, the likelihood of having a liver disease decreases 0.28- and 0.85-folds, respectively. In contrast for every 20% increase in the other non-BA parameters, the likelihood of having a liver disease increases 1.13- to 3-folds **(Table 2.13).**

In addition, we have found that most non-BA parameters were significantly different between controls vs. all individual liver disease subtypes (**Table 2.14.A** and **Table 2.14.B)**. Creatinine, INR, ALT, AST, bilirubin, AST/ALT, and MELD were higher in most liver disease group compared with controls. In contrast, albumin and protime were lower in most liver disease group compared with controls.

2.3.9. Association between non-BA parameters and BA indices

Table 2.15 shows the association between non-BA parameters and BA indices using mixed effects models. We have found that all non-BA parameters were significantly associated with most BA concentrations/indices, except creatinine ($p>0.05$).

2.4. Discussion

There were marked differences in total and individual BA concentrations and indices between patients vs. controls. Because there were significant differences in some of the demographics (age, BMI, and gender) between patients and controls (**Table 2.2)**, we wanted to ensure that the differences in the BA profiles are not due to the differences in the demographics. This was demonstrated via: (i) most of BA were not associated with demographic covariates, (ii) the ones that were associated had the same extent of association in the patient and control groups **Table 2.3**.

Patients are usually categorized based on the severity of their disease using the model for end stage liver disease (MELD) system. Patients with high MELD score are usually considered to be at higher risk of developing severe hepatobiliary complications; therefore, may be in a more urgent need for liver transplantation. MELD score was originally developed to predict 3-month mortality of people with end-stage liver disease. [42-46]. The MELD score ranges from 6 to 40 and can be categorized into low- MELD (6–15 score), medium- MELD (16–25), and high MELD (26–40). As there were only four subjects who had a high MELD score in the current study, this group was not included in the statistical analysis. In addition, the compensation status is used to assess the severity of hepatobiliary diseases as diagnosed by the hepatologists. Decompensated patients have severe complications including encephalopathy, ascites, jaundice, or bleeding varices, while compensated patients did not have any of these complications [47]. Accordingly, we have compared the BA profiles between entire patient vs. control populations as well as among the patients with different levels of disease severity.

Absolute total and all individual BA (except MDCA) were higher in patients vs. controls but to different extents, so that ratios (BA indices) increased, decreased, or remained unchanged. The highest increase of individual BA was in UDCA (11.9-fold),

while the lowest increase was for DCA and HDCA (1.6-fold) in patients vs. controls (**Table 2.5**). Among the patient disease-severity subtypes, total BA and most individual BA concentrations were also higher in the medium vs. low- MELD (**Table 2.6**) as well as decompensated vs. compensated (**Table 2.7**). Therefore, total and individual BA concentrations were in the order of medium- MELD> low- MELD> controls and decompensated > compensated > controls.

The percentages of the primary BA CDCA, CA, and HCA were higher, while the percentage of the secondary BA DCA was lower in patients vs. controls. The percentages of these primary BA also increased with the severity of the liver disease (medium-MELD>low- MELD>controls) and (decompensated > compensated> controls), whereas % DCA decreased with the severity of the disease. (**Table 2.6 and Table 2.7**).

The increase in primary BA (CA, CDCA, MCA and HCA in all of their forms) was higher than that of the secondary BA (DCA, LCA, UDCA, HDCA, MDCA, Nor-DCA, 12-oxo-CDCA, 3-dehydroCA, 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, and isoDCA in all of their forms) so that the % primary BA was 1.4 fold higher, while % secondary BA was 0.8 fold lower and the ratio of primary/ secondary BA was 3.6-fold higher in patients (**Table 2.5**). The same trend was also observed in the patients with more severe form of the disease, i.e. medium- MELD vs. low- MELD patients (**Table 2.6**) as well as decompensated vs. compensated (**Table 2.7**).

Cholestatic diseases are associated with impaired bile flow to the intestine, which is expected to translate into reduced transformation of primary BA into secondary BA by intestinal bacteria. Therefore, while all BA concentrations were higher in patients due to the impairment of bile flow, the proportion of secondary BA decreased with the severity of the cholestatic disease, which may reflect the extent of bile flow impairment. In previous

studies, the same results have been observed in patients with cholestatic liver diseases with a marked increase in the proportions of primary BA and a decrease in the proportion of secondary BA [9, 28, 48-50].

Total BA-amidates concentrations and % T-amidation increased in patients, which may be due to the disruption of bile flow to intestine (**Table 2.5**). BA are almost completely conjugated with glycine and taurine after their synthesis in the liver, but partial deamidation of BA occur by intestinal bacteria before being absorbed into the systemic circulation [51-53]. Therefore, less deamidation in the intestine may be associated with lower bile flow during cholestatic diseases, which leads to higher amidated BA in the systemic circulation and eventually urine.

The conjugation of BA with taurine and glycine decreases their pKa, increases their ionization and solubility, enhances their urinary elimination, and decreases their toxicity [38, 54-57]. However, T-amidated BA are generally less cytotoxic than G-amidated BA [56, 58, 59]. The increase in the absolute concentration of T-amidated BA in patients was more profound than that of the G-amidated and the unamidated BA. Therefore, % T-amidation of BA increased, while % G-amidation decreased (**Table 2.5)**. The same trend was observed in the disease severity subtypes, 1.5-fold higher in medium- vs. low-MELD patients, where % T-amidation was 1.3-fold higher in decompensated vs. compensated patients. On the other hand, % G-amidation was unaltered (**Table 2.6** and **Table 2.7)**. The preferential accumulation of T-amidated BA observed in this study can be interpreted as an adaptive compensating response to protect the liver from BA toxicity by increasing elimination of the more toxic G-amidated and unamidated BA compared to the less toxic T-amidated BA [9, 36, 53]. In addition, T-amidated BA are better substrates for the canalicular transporter, BSEP (Bile Salt Export Pump), than G-amidated BA, which in turn are better substrates than unamidated BA [60-62]. Many cholestatic liver diseases

are associated with BSEP mutations that lead to decrease their expression and/or function [63-65]. Therefore, an impairment of the BA transport by BSEP into bile is expected to accumulate T-amidated BA more than G-amidated and unamidated BA in the liver and eventually the spell-over in the systemic blood.

Another possibility for the differences in the BA profiles between patients vs. controls is the differences in the microbiome composition. The relationship between BA profiles and microbiome composition is very complex. While the intestinal bacteria contribute to the BA profile via BA metabolism in the intestine including dehydroxylation, epimerization, desulfation, and deamidation, in turn BA themselves, also contribute to the composition of microbiome via a balance of pro- vs. anti-bacterial effects [66, 67]. In a previous study, levels of fecal BA and the microbiome community structure were compared between patients with cirrhosis and healthy controls. As cirrhosis advanced it was observed that bacterial dysbiosis was associated to low BA levels entering the intestine. Increasing pro-inflammatory and potentially pathogenic taxa, Enterobacteriaceae as cirrhosis progresses was observed in cirrhotic patients with reduced fecal BA levels [67, 68]. Therefore, the nature of the intestinal microbiome plays an important role in the changes we observed in the BA profile in liver patients.

Both sulfated and unsulfated BA were higher in patients, but % sulfation was slightly lower in patients vs. controls (**Table 2.5)**. On the other hand, % sulfation was slightly higher in medium- compared with low-MELD and in decompensated compared with compensated patients (**Table 2.6 and Table 2.7)**. The upregulation of sulfation of BA in patients with liver diseases is thought of as a compensatory response to eliminate and detoxify the accumulated toxic BA [8, 10-13, 30, 69]. However, it is also possible that sulfation activity in these patients may eventually decrease due to exhaustion or defects of these recovery mechanisms. Therefore, while liver insults can be remediated by

upregulating BA sulfation under normal conditions and milder forms of liver diseases, subjects who fail to upregulate this defensive mechanism or exhaust it under more severe forms of the diseases are at higher risk of developing the disease and/or have worse prognosis [36]. Another explanation for the preferential accumulation of BA-sulfates could be related to the inhibition of their canalicular transport into bile by efflux transporters, mainly MRP 2-4 (Multidrug resistance-associated proteins 2-4). These transporters preferentially transport divalent amidated and conjugated (sulfated and glucuronidated) BA from hepatocytes into bile [70-73]. Mrps activity is known to be compromised in various cholestatic liver diseases due downregulation of their expression and/or membrane localization [74-76]. For example, it was shown that Mrp2 failed to anchor to the canalicular membrane in a diverse population of liver patients [77]. Therefore, impaired Mrps function including Mrp2 could play a role in the changes of BA sulfates profile we observed in liver patients. Consequently, when these transporters are compromised in liver diseases, they may lead to the preferential retention of their substrates including BA-sulfates in the liver and systemic circulation.

CYP8B1 catalyzes 12α-hydroxylation of the di-OH CDCA to the tri-OH CA. The CA/CDCA ratio or the ratio of 12α-OH (sum of CA and DCA) to non-12α (sum of all other BA: MDCA, HDCA, CDCA, HCA, UDCA, LCA and MCA) are used as probes to measure CYP8B1 activity [78-80]. The ratio of CA/ CDCA was not different between patients vs. controls, but 12α-OH/ non-12α-OH ratio was 50% lower in patients compared with controls (**Table 2.5)**. Also, both ratios were lower in medium- vs. low-MELD as well as decompensated vs. compensated patients (**Table 2.6** and **Table 2.7).** This indicates that CYP8B1 activity is compromised in liver diseases in general and is further compromised with disease severity. This would be expected as CYP8B1 exclusively takes place in hepatocytes [81, 82]; therefore, compromised as the result of worsening liver condition.

Another explanation can be related to the differences in canalicular transport of the individual BA into bile via BSEP. CDCA has a much higher affinity to BSEP than CA and other 12-α-OH BA [61, 83]. Therefore, when BSEP activity is compromised in the more severe liver diseases, it is expected to lead to the preferential accumulation of its high-affinity substrates including CDCA, which will decrease the CA/CDCA and 12α-OH/ non-12α ratios.

HI was lower in patients vs. controls (**Table 2.5)**, while it was higher in medium- compared with low- MELD (**Table 2.6)** and in decompensated compared with compensated patients (**Table 2.7)**. The term HI was introduced to describe the hydrophobic-hydrophilic balance of individual BA based on their retention time in C18 reversed-phase chromatography. A composite HI can be calculated to describe the hydrophilicity of the entire Bile acids pool based on the HI of individual BA normalized to their relative contribution to the overall BA pool. The higher the composite HI value, the higher is the concentration of the more hydrophobic BA and the more toxic is the BA pool [84-86].

ROC analysis helps compare the accuracy of biomarkers. The higher the AUC under the ROC curve, the greater the overall accuracy of the marker in distinguishing between groups. AUC values in the range of 0.9-1, 0.8-0.9, 0.7-0.8, 0.6-0.7 and 0.5-0.6 were considered excellent, good, fair, poor and fail respectively [87, 88]. Many BA concentrations and BA indices demonstrated AUC > 0.7 suggesting their usefulness as biomarkers for the diagnosis of liver diseases **Table 2.8**. Also, ROC curves are used to determine cut-off values, which quantify the normal ranges of biomarkers. The selection of optimum cut-off values is a trade-off between sensitivity and specificity, where higher cut-off values are associated with higher specificity but lower sensitivity, and vice versa.

Three potential cut-off values for every one of the BA indices with AUC > 0.7 could be used, which achieve a good balance between specificity and sensitivity **Table 2.9**.

Logistic regression analysis help determine the rate of increased or decreased risk of developing a liver disease per a defined change in BA indices. Positive regression coefficients imply that the risk of having a liver disease increases with increasing the values of BA indices, while negative coefficients imply the risk of having a liver disease increases with a decrease in the values of BA indices. we found correlation between the risk of developing a liver disease and many BA indices ($P < 0.05$). The univariate logistic regression associated with a 20% change from the mean value for the absolute BA concentrations ranged from 1.11 to 1.18, whereas it was as high as 2.72 for BA indices (Table 2.10). This suggests that BA indices are more sensitive than absolute BA concentrations in terms of predicting larger magnitudes of the risk of developing a liver disease.

All the above analyses demonstrate that BA indices can serve as a global marker to differentiate the pooled cholestic liver disease population from controls in this study. In addition, we have separated the patients into individual disease groups and performed similar analyses in these groups vs. controls, individually. Most BA indices with ROC-AUC > 0.7 were significantly different between controls vs. most of the individual liver disease subtypes (**Table 2.11.A** and **Table 2.11.B**). In particular, hepatitis C and cirrhosis were the biggest subpopulations in our study, and all global diagnostic BA indices from the pooled patients vs. control analyses ($P <0.05$ and ROC-AUC > 0.7) were also specific diagnostic markers for these two particular liver diseases vs. controls ($P <0.05$).

Similar to, we have compared the currently used non-BA parameters between entire patient vs. control populations as well as the patients with different levels of disease severity vs. controls.

All non-BA parameters (except albumin and protime) were higher in patients vs. controls but to different extents. The highest increase of non-BA parameters was in bilirubin (2.6-fold), while the lowest increase was for creatinine (1.14-fold) in patients vs. controls. Among the patient disease-severity subtypes, all non-BA parameters (except ALT and albumin) were also higher in the medium vs. low- MELD, as well as decompensated vs. compensated patients. The area under the ROC curve for differentiating the patients from controls was >0.7 for all of non-BA parameters except creatinine, protime, and AST/ALT ratio (**Table 2.12**). The univariate logistic regression showed 0.28 to 3-fold risk of developing liver diseases associated with a 20% change from the mean value of the various non- BA parameters (**Table 2.13**). As far the individual liver diseases, many non-BA parameters were also different in these individual diseases vs. controls (**Table 2.14.A** and **Table 2.14.B)**. The non-BA parameters with the highest difference in individual disease subpopulations was bilirubin, with 3.4-fold and 2.35-fold higher values than controls in cirrhotic and hepatitis C patients, respectively (**Table 2.14.A)**. Overall, AST, ALT, and INR showed the biggest differences between patients vs. controls, highest AUC from ROC analysis, and highest OR from logistic regression analyses, while creatinine and AST/ALT ratio were not significant in most of these comparisons (p- value > 0.05) and had a ROC-AUC < 0.7.

By comparing BA indices and non-BA parameters, we have found that all non-BA parameters were significantly associated with most BA concentrations/indices, except creatinine (p>0.05) (**Table 2.15**). In addition, we found that BA indices in general outperformed non-BA parameters as biomarkers for liver diseases on many levels. The

fold difference between patients vs. controls is much higher for BA vs. non-BA. For example: non-BA parameters were 0.76- to 2.52-fold higher in patients than controls (**Table 2.12)**, whereas the BA indices were as high as 11.9 (Total UDCA) and 17.6 (HI) in patients compared controls (**Table 2.5)**. Similarly, the magnitude of change within the MELD groups and compensation status is much higher in BA vs. non-BA. For example: non-BA parameters were 0.78 - to 3.8-fold higher in medium- compared with low- MELD groups (**Table 2.12)**. In contrast, BA indices were as high as 4-fold (total CDCA) higher in medium- compared with low- MELD groups (**Table 2.6)**. When the same comparison is performed on the compensation status, the magnitude of change of non-BA parameters were 0.82 - 2.15 fold higher in decompensated than compensated patients (**Table 2.12)**, whereas BA indices were as high as 3.23-fold (total HCA) in decompensated vs. compensated patients **(Table 2.7)**.

Using ROC analysis 19 BA indices and 6 non-BA parameters had AUC> 0.7, indicating they have high accuracy as biomarkers for cholestatic liver diseases. Also, both changes of BA indices and non-BA parameters were associated with the risk of having a liver disease as determined by the logistic regression analysis. For example, a 20% increase from the mean values of % non-12α-OH was associated with 2.7-fold increase in the risk of having a liver disease (**Table 2.10**), likewise the same change in INR was also associated with a 3-fold increase in the same risk (**Table 2.13**).

Finally, the fold difference between most individual liver disease subtypes vs. controls is much higher for BA vs. non-BA. For example: The highest increase of non-BA parameters was in bilirubin, which was 3.4-fold higher in cirrhotic patients and 2.35-fold higher in patients with hepatitis C than controls (**Table 2.14.A)**, whereas the total CDCA was as high as 11.92 in cirrhotic patients and 10.72 in patients with hepatitis C compared with controls (**Table 2.11.A)**.

2.5. Conclusions

The results of this study demonstrated that total and all individual BA increased in patients with 13 different cholestatic diseases. However, the high inter-individual variability of BA absolute concentrations makes most of them statistically insignificant and prevent their utilization as diagnostic markers. In contrast, BA indices had much lower inter- and intra-individual variability, which allowed their use as diagnostic and prognostic markers for liver diseases.

The increase in the total BA concentration in patients can be attributed to specific changes in the BA pool composition. This increase primarily was a result from the primary BA CDCA, CA, and HCA, while the % of the secondary BA LCA and DCA were lower. Therefore, The % secondary BA was markedly lower while % primary BA was markedly higher in patients, which lead to about 4-fold increase in the primary/ secondary BA ratio. Consequently, the BA pool has drastically shifted in patients from being 37% primary to ~50% primary BA. The increase in T-amidated BA was more profound than that of G-amidated BA, which lead to a marked increase in the % T-amidation in the BA profile of patients. Furthermore, this trend of elevated primary and amidated BA was stronger in medium-MELD and decompensated patients (more sever forms of liver diseases). Because formation of secondary BA as well as deconjugation of amidated BA into their parent BA take place in the intestine, the patterns above can be interpreted as a result of lower bile flow during cholestatic diseases, which leads to higher primary and amidated BA in the systemic circulation and eventually in urine. Also, the changes in the microbiome in liver patients is expected to contribute to this pattern of BA changes. % Sulfation of total BA was slightly decreased in patients. However, it was higher in patients with more severe forms of liver diseases indicating the upregulation of sulfation in these patients as a compensatory response to detoxify the elevated toxic BA. Finally, the increase in non12α-

OH was more profound than that of 12α-OH BA causing to a decrease in the 12α-OH/ non12α-OH ratio, which is a sensitive marker for hepatic CYP8B1. This indicates that CYP8B1 activity is compromised in liver diseases in general and is further compromised with disease severity.

Using mixed effects models, logistic regression, ROC, and pairwise comparisons analyses we have shown that several BA indices outperformed the currently used non-BA markers, currently used in the clinic, as diagnostic markers to differentiate our patient pool as well as individual cholestatic diseases against healthy controls.

2.6. Figures and Tables

Figure legends

Figure 2.1. Receiver operating characteristics (ROC) curves of BA concentrations and indices with AUC > 0.7. The area under the ROC curve (AUC) for differentiating patients from healthy controls. The scale of both the y-axis (sensitivity) and the x-axis (1-specificity) is 0-1. BA indices are higher in patients vs. controls, and the positive actual state was patients except the ones annotated with "*", where BA indices were lower in patients compared to controls. For these BA indices, "1 - AUC" instead of "AUC" was calculated.

Figure 2.1. Receiver operating characteristics (ROC) curves of BA concentrations and indices with AUC > 0.7.

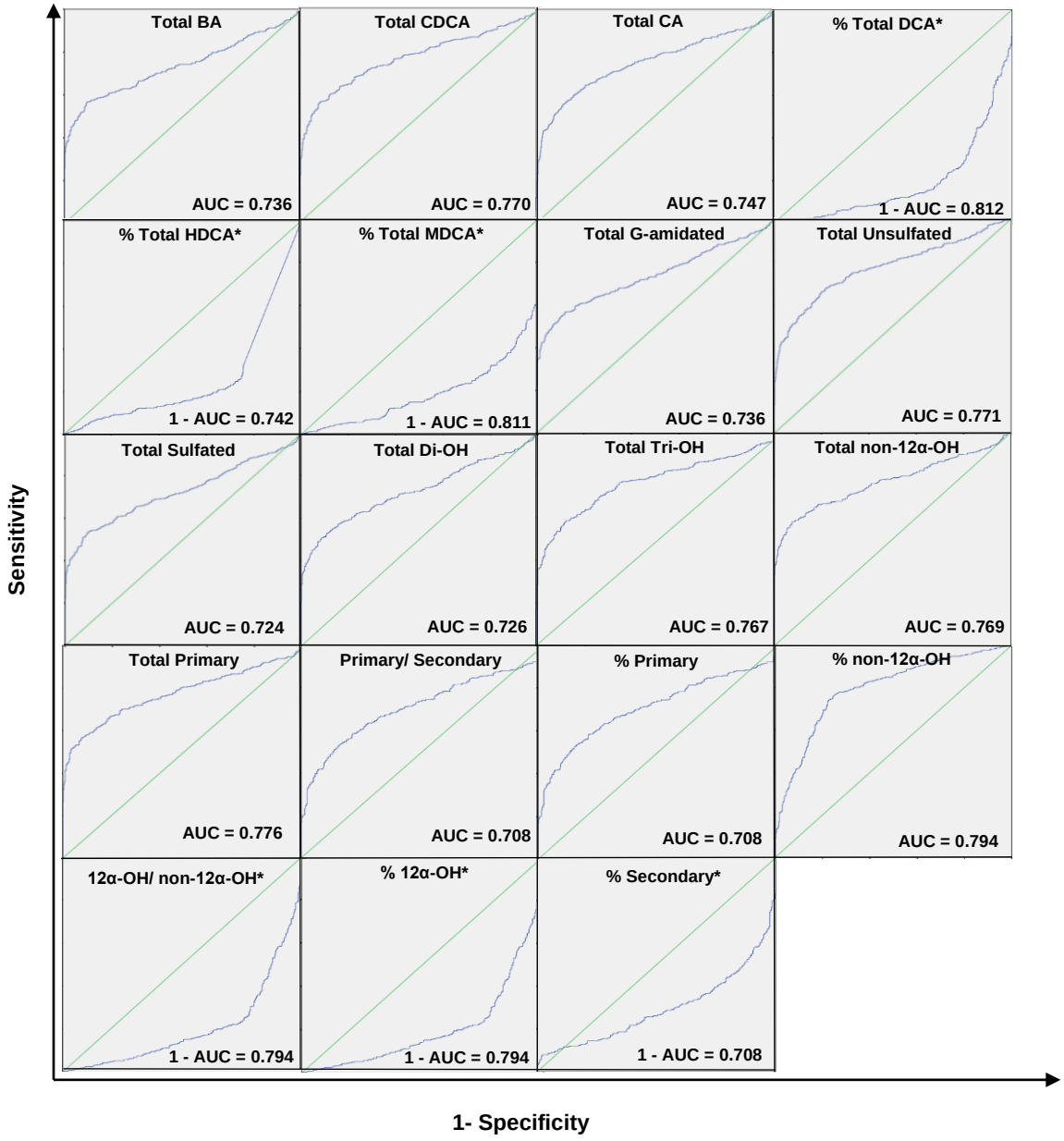


Table 2.1. List of BA indices.

Composition	Hepatic Metabolism	Hydrophilicity	CYP8B1 Activity	Intestinal Contribution
Concentration of individual BA	Total Sulfated	Total Mono-OH	Total 12α-OH	Total Primary
% of individual BA	Total G-amidated	Total Di-OH	Total non-12α-OH	Total Secondary
	Total T-amidated	Total Tri-OH	12α-OH/ non12α-OH	Primary/ Secondary
	% Sulfation	% Mono-OH	CA/ CDCA	% Primary
	% Amidation	% Di-OH	% 12α-OH	% Secondary
	% G-amidation	% Tri-OH	% non-12α-OH	
	% T-amidation	HI		

Table 2.2. Demographics.

	Controls	Patients
N	103	300
Gender *		
Male, Female	32, 71	157, 143
Age *		
Mean ± SEM	44.3 ± 0.64	52.1± 0.54
19-29	17	11
30-41	28	40
42-53	30	92
54-83	28	157
Body Mass Index (BMI) *		
Mean ± SEM	27.5 ± 0.28	30.9 ± 0.32
Normal BMI<25	30	69
Overweight BMI=25-29.9	45	104
Obese BMI≥ 30	28	127
Race		
White	88	247
Black	7	14
Asian	7	13
Hispanic	1	8
Others	0	18

*Significant difference between controls and patients ($p<0.05$)

Table 2.3. Association between demographics and BA concentrations and indices.

BA(μM) / BA indices (%)	Mixed effect with covariates						Interaction (group × covariates)		
	Group	Gender	Age continuous	BMI continuous	Age category	BMI category	Gender	BMI continuous	BMI Category
	P-value								
Total BA	0.000	0.207	0.700	0.067	0.772	0.026	NA	NA	0.136
Total LCA	0.000	0.276	0.191	0.121	0.198	0.178	NA	NA	NA
Total UDCA	0.000	0.047	0.562	0.056	0.613	0.023	0.114	0.219	0.096
Total CDCA	0.000	0.965	0.262	0.113	0.425	0.048	NA	NA	0.199
Total CA	0.005	0.193	0.113	0.791	0.201	0.098	NA	NA	NA
Total MCA	0.000	0.892	0.402	0.534	0.441	0.862	NA	NA	NA
Total HCA	0.003	0.226	0.123	0.831	0.209	0.318	NA	NA	NA
% LCA	0.000	0.047	0.392	0.544	0.733	0.448	0.682	NA	NA
% CDCA	0.000	0.043	0.716	0.201	0.652	0.166	0.112	NA	NA
% DCA	0.000	0.830	0.430	0.400	0.249	0.581	NA	NA	NA
% MCA	0.062	0.166	0.569	0.008	0.595	0.004	NA	0.325	0.312
Total Unamidated	0.000	0.429	0.130	0.090	0.222	0.058	NA	NA	NA
Total G-amidated	0.000	0.102	0.902	0.051	0.950	0.025	NA	0.263	0.136
Total T-amidated	0.002	0.349	0.128	0.532	0.249	0.045	NA	NA	0.245
Total Unsulfated	0.000	0.581	0.581	0.936	0.659	0.353	NA	NA	NA
Total Sulfated	0.000	0.156	0.631	0.051	0.716	0.022	NA	0.263	0.139
Total Mono-OH	0.000	0.276	0.191	0.121	0.329	0.124	NA	NA	NA
Total Di-OH	0.000	0.123	0.721	0.046	0.459	0.060	NA	0.236	NA
Total Tri-OH	0.000	0.333	0.307	0.792	0.563	0.279	NA	NA	NA
% Mono-OH	0.000	0.047	0.392	0.544	0.733	0.448	0.777	NA	NA
% Tri-OH	0.019	0.056	0.638	0.135	0.568	0.043	0.308	NA	0.529
Total 12α-OH	0.000	0.957	0.104	0.384	0.131	0.366	NA	NA	NA
Total non-12α-OH	0.000	0.152	0.893	0.055	0.959	0.019	NA	NA	0.108
12α-OH/ non12α-OH	0.000	0.072	0.549	0.984	0.934	0.598	NA	NA	NA
% 12α-OH	0.000	0.183	0.649	0.868	0.450	0.951	NA	NA	NA
% non-12α-OH	0.000	0.183	0.649	0.868	0.450	0.951	NA	NA	NA
Total Primary	0.000	0.842	0.230	0.232	0.368	0.078	NA	NA	NA
Total Secondary	0.000	0.037	0.615	0.040	0.668	0.023	0.128	0.243	0.132
Primary/ Secondary	0.005	0.278	0.675	0.652	0.704	0.217	NA	NA	NA
% Primary	0.000	0.009	0.610	0.613	0.529	0.782	0.046	NA	NA
% Secondary	0.000	0.009	0.610	0.613	0.529	0.782	0.046	NA	NA
HI	0.000	0.674	0.863	0.990	0.738	0.477	NA	NA	NA

NA: Not applicable because of the lack of significance from the mixed effect analysis.

Table 2.4. Absolute concentrations of major BA in controls and patients.

BA	Unamidated	G-BA	T-BA	Total
	Mean (μM) ± SEM			
Controls				
Unsulfated BA				
LCA	0.000±0.00	0.000±0.00	0.000±0.00	0.001±0.00
UDCA	0.004±0.00	0.033±0.00	0.002±0.00	0.038±0.00
CDCA	0.003±0.00	0.008±0.00	0.002±0.00	0.013±0.00
DCA	0.022±0.00	0.011±0.00	0.002±0.00	0.035±0.00
HDCA	0.01±0.00	0.00±0.00	ND	0.007±0.00
MDCA	0.060±0.01	ND	ND	0.058±0.01
CA	0.179±0.03	0.067±0.00	0.009±0.00	0.255±0.03
MCA	0.028±0.00	0.287±0.02	0.041±0.00	0.356±0.02
HCA	0.008±0.00	0.016±0.00	0.001±0.00	0.026±0.00
Other BA*	0.160±0.01	—	—	0.160±0.01
Total unsulfated	0.464±0.04	0.422±0.02	0.057±0.00	0.943±0.05
Sulfated BA				
LCA	0.010±0.00	0.780±0.04	0.220±0.01	1.010±0.05
UDCA	0.450±0.02	1.040±0.05	0.030±0.00	1.520±0.07
CDCA	0.070±0.01	2.380±0.13	0.060±0.00	2.510±0.13
DCA	0.010±0.00	2.900±0.14	0.220±0.02	3.130±0.16
CA	0.004±0.00	0.056±0.01	0.126±0.01	0.190±0.01
Total sulfated	0.535±0.03	7.170±0.28	0.650±0.03	8.350±0.31
Overall total	1.000±0.05	7.590±0.29	0.710±0.03	9.300±0.33
Patients				
Unsulfated BA				
LCA	0.004±0.00	0.001±0.00	0.0001±0.00	0.005±0.00
UDCA	0.079±0.03	0.410±0.17	0.012±0.00	0.500±0.21
CDCA	0.020±0.00	0.090±0.01	0.100±0.02	0.210±0.03
DCA	0.040±0.00	0.040±0.00	0.010±0.00	0.090±0.01
HDCA	0.010±0.00	0.00±0.00	ND	0.010±0.00
MDCA	0.050±0.01	ND	ND	0.050±0.01
CA	0.240±0.03	0.550±0.07	0.320±0.08	1.120±0.14
MCA	0.120±0.02	1.940±0.29	0.730±0.09	2.790±0.34
HCA	0.010±0.00	0.170±0.02	0.090±0.02	0.270±0.04
Other BA*	0.860±0.13	—	—	0.860±0.13
Total	0.460±0.04	0.42±0.02	0.06±0.00	5.910±0.57
Sulfated BA				
LCA	0.030±0.01	2.230±0.20	0.650±0.06	2.910±0.24
UDCA	1.560±0.23	15.30±2.68	1.230±0.27	18.10±3.08
CDCA	0.190±0.03	18.70±1.79	1.910±0.38	20.80±2.07
DCA	0.040±0.01	4.280±0.54	0.520±0.07	4.840±0.58
CA	0.080±0.01	0.910±0.13	1.030±0.21	2.010±0.31
Total	1.900±0.24	41.40±4.12	5.340±0.74	48.70±4.77
Overall total	3.330±0.33	44.60±4.46	6.610±0.85	54.60±5.20

*Other BA: Nor-DCA, 12-oxo-CDCA, 3-dehydroCA, 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, and isoDCA

ND: not detected

—: not quantified

SEM: standard error of mean

Table 2.5. Representative BA concentrations and indices in controls vs. patients.

BA (μM) / BA indices (%)	Controls		Patients		Patients vs. Controls	
	Mean	SEM	Mean	SEM	Ratio	P-value
Total BA	9.30	0.33	54.6	5.20	5.87	0.000
Total LCA	1.01	0.05	2.92	0.24	2.88	0.000
Total UDCA	1.56	0.07	18.6	3.23	11.9	0.001
Total CDCA	2.52	0.13	21.0	2.09	8.35	0.000
Total DCA	3.16	0.16	4.92	0.58	1.56	0.072
Total HDCA	0.01	0.00	0.01	0.00	1.57	0.051
Total MDCA	0.06	0.01	0.05	0.01	0.90	0.992
Total CA	0.44	0.03	3.13	0.44	7.09	0.003
Total MCA	0.36	0.02	2.79	0.34	7.83	0.000
Total HCA	0.03	0.00	0.27	0.04	10.6	0.001
Other BA*	0.16	0.01	0.86	0.13	5.54	NA
% LCA	11.5%	0.38%	9.20%	0.39%	0.79	0.002
% UDCA	17.7%	0.49%	21.3%	0.88%	1.21	0.138
% CDCA	27.1%	0.65%	36.3%	0.94%	1.34	0.000
% DCA	31.1%	0.68%	14.6%	0.53%	0.47	0.000
% HDCA	0.07%	0.01%	0.04%	0.00%	0.54	0.052
% MDCA	0.64%	0.04%	0.36%	0.05%	0.56	0.135
% CA	5.25%	0.27%	6.27%	0.25%	1.19	0.064
% MCA	4.03%	0.16%	6.39%	0.34%	1.58	0.003
% HCA	0.30%	0.02%	0.45%	0.04%	1.52	0.018
Total Unamidated	1.00	0.05	3.33	0.33	3.34	0.000
Total G-amidated	7.59	0.29	44.6	4.46	5.88	0.000
Total T-amidated	0.71	0.03	6.61	0.85	9.37	0.001
% Amidation	87.7%	0.47%	86.9%	0.65%	0.99	0.053
% G-amidation	79.7%	0.49%	76.0%	0.71%	0.95	0.000
% T-amidation	7.98%	0.26%	10.8%	0.46%	1.35	0.005
Total Unsulfated	0.94	0.05	5.91	0.57	6.26	0.000
Total Sulfated	8.35	0.31	48.7	4.77	5.83	0.000
% Sulfation	88.5%	0.46%	82.9%	0.60%	0.94	0.000
Total Mono-OH	1.01	0.05	2.92	0.24	2.88	0.000
Total Di-OH	7.30	0.29	44.6	4.58	6.11	0.000
Total Tri-OH	0.82	0.04	6.19	0.65	7.52	0.000
% Mono-OH	11.5%	0.38%	9.16%	0.39%	0.79	0.002
% Di-OH	76.6%	0.50%	72.7%	0.65%	0.95	0.001
% Tri-OH	9.58%	0.33%	13.1%	0.43%	1.37	0.000
Total 12α-OH	3.62	0.17	8.35	0.83	2.30	0.001
Total non-12α-OH	5.67	0.20	46.2	4.68	8.15	0.000
12α-OH/ non12α-OH	0.65	0.02	0.33	0.01	0.51	0.000
CA/ CDCA	0.24	0.01	0.24	0.02	1.00	0.625
% 12α-OH	36.7%	0.62%	22.1%	0.54%	0.60	0.000
% non-12α-OH	63.3%	0.62%	77.9%	0.54%	1.23	0.000
Total Primary	3.34	0.15	27.2	2.59	8.15	0.000
Total Secondary	5.95	0.23	27.4	3.52	4.59	0.000
Primary/ Secondary	0.69	0.03	2.52	0.22	3.63	0.000
% Primary	36.7%	0.70%	49.4%	1.06%	1.35	0.000
% Secondary	63.3%	0.70%	50.6%	1.06%	0.80	0.000
HI	-0.01	0.01	-0.12	0.01	17.6	0.000

*Other BA: Nor-DCA, 12-oxo-CDCA, 3-dehydroCA, 6-oxo-LCA, 7-oxo-LCA, 12-oxo-LCA, isoLCA, and isoDCA

NA: not applicable

Table 2.6. Representative BA concentrations and indices in medium- vs. low- MELD patients.

BA (μM) / BA indices (%)	Low- MELD		Medium- MELD		Medium- vs. Low- MELD	
	Mean	SEM	Mean	SEM	Ratio	P-value
Total BA	59.2	7.94	116	24.8	1.96	1.000
Total LCA	3.40	0.35	6.01	1.72	1.77	0.175
Total UDCA	24.4	5.34	18.6	6.30	0.76	0.172
Total CDCA	18.3	2.31	71.4	16.3	3.90	0.000
Total DCA	5.30	0.96	6.08	1.47	1.15	1.000
Total HDCA	0.01	0.00	0.01	0.00	0.61	1.000
Total MDCA	0.05	0.01	0.06	0.01	1.28	1.000
Total CA	2.80	0.48	10.6	4.45	3.79	0.000
Total MCA	3.58	0.57	2.15	0.46	0.60	0.210
Total HCA	0.25	0.04	0.86	0.36	3.48	0.002
% LCA	9.31%	0.53%	7.97%	1.47%	0.86	1.000
% UDCA	23.1%	1.29%	14.3%	2.52%	0.62	1.000
% CDCA	34.7%	1.21%	55.6%	3.17%	1.60	0.000
% DCA	13.8%	0.65%	7.18%	1.33%	0.52	0.005
% HDCA	0.04%	0.01%	0.01%	0.00%	0.32	0.661
% MDCA	0.29%	0.04%	0.13%	0.03%	0.43	1.000
% CA	5.75%	0.30%	8.70%	1.25%	1.51	0.145
% MCA	7.15%	0.48%	3.70%	0.91%	0.52	0.000
% HCA	0.46%	0.07%	0.75%	0.15%	1.61	0.148
Total Unamidated	4.24	0.55	2.87	0.72	0.68	0.062
Total G-amidated	48.4	6.89	92.8	19.7	1.92	1.000
Total T-amidated	6.58	1.04	20.7	7.30	3.15	0.040
% Amidation	86.7%	0.87%	94.4%	1.28%	1.09	0.005
% G-amidation	75.5%	0.96%	77.2%	2.73%	1.02	1.000
% T-amidation	11.2%	0.64%	17.1%	2.15%	1.53	0.002
Total Unsulfated	6.99	0.93	9.04	2.42	1.29	1.000
Total Sulfated	52.3	7.21	107	23.2	2.05	1.000
% Sulfation	82.4%	0.81%	88.3%	1.34%	1.07	0.009
Total Mono-OH	3.40	0.35	6.01	1.72	1.77	0.175
Total Di-OH	48.1	7.01	96.2	20.9	2.00	1.000
Total Tri-OH	6.63	0.90	13.6	4.90	2.06	0.301
% Mono-OH	9.31%	0.53%	7.97%	1.47%	0.86	1.000
% Di-OH	72.0%	0.90%	77.2%	2.14%	1.07	0.058
% Tri-OH	13.4%	0.59%	13.1%	1.40%	0.98	0.274
Total 12α-OH	8.55	1.23	16.8	4.86	1.96	0.053
Total non-12α-OH	50.7	7.21	99.6	21.5	1.96	1.000
12α-OH/ non12α-OH	0.30	0.01	0.20	0.02	0.68	0.135
CA/ CDCA	0.21	0.01	0.17	0.03	0.81	1.000
% 12α-OH	21.0%	0.69%	16.1%	1.44%	0.77	0.008
% non-12α-OH	79.0%	0.69%	83.9%	1.44%	1.06	0.008
Total Primary	25.0	3.08	85.1	19.5	3.41	0.000
Total Secondary	34.3	5.78	31.3	8.05	0.91	0.316
Primary/ Secondary	2.19	0.24	5.02	1.16	2.29	1.000
% Primary	48.1%	1.40%	68.7%	3.10%	1.43	0.014
% Secondary	51.9%	1.40%	31.3%	3.10%	0.60	0.014
HI	-0.15	0.01	-0.05	0.03	0.36	0.189

Table 2.7. Representative BA concentrations and indices in compensated vs. decompensated patients.

BA (μM) / BA indices (%)	Compensated		Decompensated		Decompensated vs. Compensated	
	Mean	SEM	Mean	SEM	Ratio	P-value
Total BA	66.6	10.8	86.9	14.9	1.31	0.160
Total LCA	3.73	0.54	4.26	0.70	1.14	0.547
Total UDCA	27.0	6.49	21.0	9.82	0.78	0.687
Total CDCA	20.4	3.42	45.0	6.28	2.20	0.001
Total DCA	6.85	1.76	4.93	0.73	0.72	0.394
Total HDCA	0.01	0.00	0.02	0.01	1.61	0.430
Total MDCA	0.06	0.01	0.05	0.01	0.86	0.619
Total CA	2.62	0.40	6.28	1.51	2.39	0.024
Total MCA	4.48	0.93	4.07	0.83	0.91	0.864
Total HCA	0.20	0.04	0.64	0.14	3.23	0.002
% LCA	9.00%	0.64%	6.61%	0.64%	0.73	0.020
% UDCA	24.9%	1.97%	12.0%	1.32%	0.48	0.007
% CDCA	33.2%	1.62%	54.74%	2.05%	1.65	0.000
% DCA	14.3%	0.98%	9.17%	1.00%	0.64	0.000
% HDCA	0.02%	0.01%	0.03%	0.01%	1.42	0.532
% MDCA	0.34%	0.14%	0.11%	0.01%	0.33	0.264
% CA	6.07%	0.54%	7.58%	0.48%	1.25	0.262
% MCA	7.26%	0.72%	7.21%	0.82%	0.99	0.542
% HCA	0.35%	0.05%	0.74%	0.08%	2.09	0.005
Total Unamidated	4.35	0.69	3.88	1.04	0.89	0.876
Total G-amidated	56.2	9.81	70.5	12.5	1.25	0.240
Total T-amidated	5.97	0.79	12.6	2.58	2.11	0.010
% Amidation	87.9%	1.15%	93.6%	0.75%	1.06	0.003
% G-amidation	76.5%	1.30%	78.8%	1.23%	1.03	0.161
% T-amidation	11.5%	0.93%	14.8%	1.02%	1.29	0.161
Total Unsulfated	7.84	1.19	9.53	1.85	1.22	0.310
Total Sulfated	58.7	9.97	77.4	13.4	1.32	0.156
% Sulfation	82.7%	1.12%	85.2%	0.99%	1.03	0.054
Total Mono-OH	3.73	0.54	4.26	0.70	1.14	0.547
Total Di-OH	54.4	9.55	70.9	13.1	1.31	0.174
Total Tri-OH	7.30	1.25	11.0	1.96	1.51	0.085
% Mono-OH	9.00%	0.64%	6.61%	0.64%	0.73	0.020
% Di-OH	72.7%	1.14%	76.0%	1.31%	1.05	0.016
% Tri-OH	13.7%	0.92%	15.5%	0.95%	1.13	0.674
Total 12α-OH	10.1	2.08	11.44	1.75	1.14	0.554
Total non-12α-OH	56.5	9.36	75.51	14.0	1.34	0.137
12α-OH/ non12α-OH	0.33	0.02	0.24	0.02	0.71	0.002
CA/ CDCA	0.21	0.02	0.17	0.01	0.79	0.043
% 12α-OH	22.0%	1.07%	17.3%	0.99%	0.79	0.001
% non-12α-OH	78.0%	1.07%	82.7%	0.99%	1.06	0.001
Total Primary	27.7	4.46	56.0	7.59	2.02	0.001
Total Secondary	38.8	7.43	31.0	10.3	0.80	0.874
Primary/ Secondary	2.27	0.44	5.98	0.69	2.64	0.001
% Primary	46.9%	2.05%	70.3%	1.88%	1.50	0.000
% Secondary	53.1%	2.05%	29.7%	1.88%	0.56	0.000
HI	-0.18	0.02	-0.13	0.02	0.75	0.266

Table 2.8. Area under the ROC curve (AUC) of BA concentrations and indices*.

BA (μM) / BA indices (%)	AUC
Total BA	0.736
Total LCA	0.583
Total UDCA	0.662
Total CDCA	0.770
Total DCA	0.514
Total HDCA	0.659
Total MDCA	0.576
Total CA	0.747
Total MCA	0.691
Total HCA	0.623
% LCA	0.672
% UDCA	0.555
% CDCA	0.677
% DCA	0.812
% HDCA	0.742
% MDCA	0.811
% CA	0.554
% MCA	0.540
% HCA	0.579
Total Unamidated	0.592
Total G-amidated	0.736
Total T-amidated	0.690
% Amidation	0.679
% G-amidation	0.590
% T-amidation	0.507
Total Unsulfated	0.771
Total Sulfated	0.724
% Sulfation	0.581
Total Mono-OH	0.583
Total Di-OH	0.726
Total Tri-OH	0.767
% Mono-OH	0.672
% Di-OH	0.502
% Tri-OH	0.603
Total 12α-OH	0.599
Total non-12α-OH	0.769
12α-OH/ non12α-OH	0.794
CA/ CDCA	0.565
% 12α-OH	0.794
% non-12α-OH	0.794
Total Primary	0.776
Total Secondary	0.626
Primary/ Secondary	0.708
% Primary	0.708
% Secondary	0.708
HI	0.664

* AUC from ROC analysis of pooled patients vs. controls.

Table 2.9. ROC analysis of BA concentrations and indices with AUC > 0.7*.

BA (μM) / BA indices (%)	AUC	(Cutoff Value; Sensitivity, Specificity)		
Total BA	0.736	(9.34; 70, 61)	(10.2; 67, 66)	(10.9; 64, 70)
Total CDCA	0.770	(2.66; 72, 70)	(2.74; 71, 71)	(2.88; 70, 73)
Total CA	0.747	(0.43; 70, 69)	(0.44; 69, 69)	(0.45; 69, 70)
% DCA	0.812	(19.5; 81, 70)	(22.1; 75, 75)	(24.3; 70, 81)
% HDCA	0.742	(0.003; 75, 70)	(0.01; 73, 75)	(0.01; 70, 77)
% MDCA	0.811	(0.18; 80, 70)	(0.21; 75, 73)	(0.24; 70, 76)
Total G-amidated	0.736	(7.35; 70, 59)	(8.20; 68, 65)	(8.87; 65, 70)
Total Unsulfated	0.771	(0.96; 71, 70)	(1.00; 71, 71)	(1.04; 70, 73)
Total Sulfated	0.724	(7.71; 70, 56)	(8.72; 68, 64)	(9.63; 65, 70)
Total Di-OH	0.726	(6.90; 70, 59)	(7.51; 67, 64)	(8.39; 65, 70)
Total Tri-OH	0.767	(0.82; 72, 70)	(0.85; 71, 72)	(0.88; 70,74)
Total non-12α-OH	0.769	(6.38; 70, 69)	(6.43; 69, 69)	(6.57; 69, 70)
12α-OH/ non12α-OH	0.794	(0.37; 78, 70)	(0.41; 76, 76)	(0.45; 70, 78)
% 12α-OH	0.794	(27.1; 78, 70)	(29.1; 76, 76)	(30.9; 70, 78)
% non-12α-OH	0.794	(69.1; 78, 70)	(70.9; 76, 76)	(72.9; 70, 78)
Total Primary	0.776	(3.75; 72, 70)	(4.03; 71, 73)	(4.19; 70, 75)
Primary/ Secondary	0.708	(0.60; 70, 58)	(0.66; 67, 64)	(0.74; 64, 70)
% Primary	0.708	(37.4; 70, 58)	(39.9; 67, 64)	(42.6; 64, 70)
% Secondary	0.708	(57.4; 70, 64)	(60.0; 64, 67)	(62.6; 58, 70)

* AUC from ROC analysis of pooled patients vs. controls.

Table 2.10. Univariate logistic regression analysis of BA concentrations and indices*.

BA (μM) / BA indices (%)	B-value (Regression Coefficient)	P-value	Exp(B)-Odds ratio		
			1-unit change	10% change	20% change
Total BA	0.080	0.000	1.08	1.08	1.16
Total CDCA	0.226	0.000	1.25	1.06	1.12
Total CA	1.181	0.000	3.26	1.05	1.11
% DCA	-0.080	0.000	0.92	0.78	0.61
% HDCA	-1.898	0.069	0.15	0.99	0.97
% MDCA	-0.174	0.162	0.84	0.99	0.98
Total G-amidated	0.084	0.000	1.09	1.07	1.14
Total Unsulfated	0.784	0.000	2.19	1.08	1.16
Total Sulfated	0.080	0.000	1.08	1.07	1.14
Total Di-OH	0.094	0.000	1.10	1.07	1.15
Total Tri-OH	0.731	0.000	2.08	1.06	1.13
Total non-12α-OH	0.146	0.000	1.16	1.09	1.18
12α-OH/ non12α-OH	-2.349	0.000	0.10	0.86	0.74
% 12α-OH	-0.079	0.000	0.92	0.75	0.56
% non-12α-OH	0.079	0.000	1.08	1.65	2.72
Total Primary	0.190	0.000	1.21	1.07	1.14
Primary/ Secondary	0.834	0.000	2.30	1.06	1.12
% Primary	0.033	0.000	1.03	1.13	1.27
% Secondary	-0.033	0.000	0.97	0.81	0.66

* BA with ROC-AUC > 0.7 were included in this table.

Table 2.11.a. BA concentrations and indices in controls and patients with specific liver disease subtype*.

BA (μM) / BA indices (%)	Controls		Hepatitis C		Hepatitis B		Laennec Cirrhosis		Primary biliary cholangitis		Primary sclerosing cholangitis		Autoimmune Hepatitis	
	N=103		N=71		N=15		N=105		N=12		N=17		N=26	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Total BA	9.30	0.33	53.3[a]	9.96	13.7[a]	5.23	62.0[a]	9.44	237[a]	69.8	124[a]	27.4	71.9[a]	15.2
Total CDCA	2.52	0.13	27.0[a]	4.89	6.76[a]	4.16	30.0[a]	4.25	28.6[a]	9.99	39.4[a]	10.7	29.2[a]	8.46
Total CA	0.44	0.03	3.05[a]	0.54	1.16[a]	0.70	4.00[a]	1.02	5.07[a]	2.27	6.47[a]	2.25	1.96[a]	0.39
% DCA	31.1%	0.68%	16.2%[a]	1.27%	19.9%[a]	3.26%	13.1%[a]	0.95%	7.99%[a]	2.22%	9.01%[a]	1.86%	15.6%[a]	1.43%
% HDCA	0.07%	0.01%	0.02%[a]	0.00%	0.06%	0.02%	0.02%[a]	0.01%	0.03%	0.02%	0.01%	0.00%	0.02%	0.00%
% MDCA	0.64%	0.04%	0.19%[a]	0.04%	0.38%	0.08%	0.16%[a]	0.03%	0.15%[a]	0.07%	0.18%[a]	0.06%	0.22%[a]	0.05%
Total G-amidated	7.59	0.29	44.8[a]	9.11	11.6[a]	4.52	50.8[a]	8.26	210[a]	60.4	106[a]	24.0	61.8[a]	13.5
Total Unsulfated	0.94	0.05	7.69[a]	1.43	1.62[a]	0.38	7.94[a]	1.21	17.6[a]	8.66	6.21[a]	1.13	6.06[a]	1.44
Total Sulfated	8.35	0.31	45.6[a]	8.73	12.1[a]	4.96	54.1[a]	8.59	219[a]	62.0	117[a]	26.5	65.9[a]	14.7
Total Di-OH	7.30	0.29	41.0[a]	7.98	10.6[a]	4.41	49.05[a]	7.97	214[a]	62.50	111[a]	25.9	60.9[a]	13.7
Total Tri-OH	0.82	0.04	8.61[a]	1.63	1.82[a]	0.80	8.59[a]	1.54	12.48[a]	6.02	9.28[a]	2.44	4.83[a]	0.76
Total non-12α-OH	5.67	0.20	41.8[a]	7.46	10.3[a]	4.52	50.8[a]	7.85	224[a]	66.9	113[a]	26.5	62.3[a]	14.2
12α-OH/ non12α-OH	0.65	0.02	0.37[a]	0.05	0.51	0.08	0.31[a]	0.02	0.15[a]	0.05	0.29[a]	0.05	0.33[a]	0.04
% 12α-OH	36.7%	0.62%	22.8%[a]	1.28%	31.1%	2.9%	21.3%[a]	1.0%	10.8%[a]	2.69%	18.8%[a]	2.6%	22%[a]	1.69%
% non-12α-OH	63.3%	0.62%	77.2%[a]	1.28%	68.9%	2.9%	78.7%[a]	1.0%	89.2%[a]	2.69%	81.2%[a]	2.6%	78.0%[a]	1.69%
Total Primary	3.34	0.15	35.6[a]	6.23	8.58[a]	4.93	38.6[a]	5.47	41.1[a]	15.6	48.6[a]	12.6	34.1[a]	8.91
Primary/ Secondary	0.69	0.03	3.70[a]	0.55	1.52[a]	0.59	4.33[a]	0.60	0.30	0.10	2.88[a]	0.68	2.09[a]	0.60
% Primary	36.7%	0.70%	60.2%[a]	1.99%	47.0%[a]	3.93%	60.9%[a]	1.7%	18.0%[a]	2.97%	51.6%[a]	5.09%	50.1%[a]	2.77%
% Secondary	63.3%	0.70%	39.8%[a]	1.99%	53.0%[a]	3.93%	39.1%[a]	1.7%	82.0%[a]	2.97%	48.4%[a]	5.09%	49.9%[a]	2.77%

*BA with ROC-AUC > 0.7 were included in this table.

[a] Significant difference between each specific liver disease subtype vs. controls (P<0.05).

Table 2.11.b. BA concentrations and indices in controls and patients with specific liver disease subtype*.

BA (μM) / BA indices (%)	α-1 antitrypsin deficiency		NASH		Carcinoma		Cryptogenic cirrhosis		Polycystic liver disease		Elevated LFT		Unknown Etiologys	
	N=5		N=52		N=25		N=11		N=3		N=19		N=5	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Total BA	25.4[a]	6.79	29.8[a]	4.31	90.9[a]	26.7	56.0[a]	16.3	3.34	0.80	106[a]	69.3	203[a]	53.2
Total CDCA	9.14[a]	3.12	13.6[a]	2.33	31.9[a]	8.76	27.7[a]	7.39	1.39	0.72	31.3[a]	18.8	63.5[a]	26.4
Total CA	2.44[a]	0.85	1.65[a]	0.23	3.54[a]	1.11	2.70[a]	0.59	0.07	0.05	5.55[a]	3.09	15.4[a]	8.12
% DCA	18.9%	4.23%	15.7%[a]	1.22%	14.9%[a]	1.43%	7.93%[a]	2.68%	20.6%	4.23%	17.4%[a]	3.07%	8.28%[a]	1.82%
% HDCA	0.03%	1.21%	0.05%	0.01%	0.06%	0.02%	0.04%	0.03%	0.00%	0.00%	0.05%	0.04%	0.01%	0.01%
% MDCA	0.38%	0.16%	0.49%	0.21%	0.16%[a]	0.05%	0.07%[a]	0.01%	0.63%	0.25%	1.34%	0.98%	0.14%	0.10%
Total G-amidated	16.3[a]	4.58	26.0[a]	4.01	78.1[a]	24.5	49.2[a]	14.9	2.89	0.85	86.8[a]	58.2	158[a]	37.53
Total Unsulfated	3.82[a]	1.21	4.64[a]	0.65	13.0[a]	3.16	3.58[a]	0.58	0.20	0.09	13.0[a]	9.22	11.4[a]	2.89
Total Sulfated	21.6[a]	6.11	25.2[a]	4.04	77.9[a]	24.2	52.4[a]	16.0	3.14	0.77	92.9[a]	60.2	191[a]	50.8
Total Di-OH	15.9[a]	4.72	23.2[a]	3.87	71.9[a]	23.0	50.3[a]	16.0	2.83	0.72	91.0[a]	61.0	175[a]	46.1
Total Tri-OH	5.24[a]	1.83	4.27[a]	0.66	13.8[a]	3.62	3.92[a]	0.65	0.16	0.07	10.8[a]	6.50	21.8[a]	8.81
Total non-12α-OH	19.4[a]	5.59	23.8[a]	3.74	75.7[a]	22.1	51.1[a]	16.1	2.56	0.70	94.7[a]	66.1	176[a]	45.6
12α-OH/ non12α-OH	0.40	0.07	0.34[a]	0.02	0.28[a]	0.03	0.22[a]	0.05	0.33	0.07	0.37[a]	0.06	0.20[a]	0.04
% 12α-OH	27.1%	3.18%	23.5%[a]	1.14%	20.6%[a]	1.55%	16.2%[a]	2.88%	24.2%	4.41%	24.8%[a]	2.8%	15.8%[a]	2.84%
% non-12α-OH	72.9%	3.18%	76.5%[a]	1.14%	79.4%[a]	155%	83.8%[a]	2.88%	75.8%	4.41%	75.2%[a]	2.8%	84.2%[a]	2.84%
Total Primary	14.4[a]	4.33	17.9[a]	2.70	45.7[a]	12.1	31.6[a]	7.89	1.56	0.65	42.1[a]	23.9	85.2[a]	34.7
Primary/ Secondary	1.26	0.26	2.28[a]	0.30	2.32[a]	0.40	6.43[a]	2.09	1.00	0.44	1.70[a]	0.43	1.67[a]	0.65
% Primary	50.3%	5.43%	52.8%[a]	2.21%	56.1%[a]	3.25%	68.2%[a]	5.80%	45.4%	11.0%	49.3%[a]	4.6%	40.3%[a]	8.45%
% Secondary	49.7%	5.43%	47.2%[a]	2.21%	43.9%[a]	3.25%	31.8%[a]	5.80%	54.6%	11.0%	50.7%[a]	4.6%	59.7%[a]	8.45%

* BA with ROC-AUC > 0.7 were included in this table.

[a] Significant difference between each specific liver disease subtype vs. controls (P<0.05).

Table 2.12. Summary of non-BA parameters.

Non-BA parameters	Controls		Patients										ROC*
			Pooled		Low- MELD		Medium- MELD		Compensated		Decompensated		AUC
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	
Creatinine (mg/dL)	0.87	0.01	0.99	0.05	0.93	0.07	1.33[b]	0.16	1.05	0.15	1.05	0.06	0.539
Albumin (g/dL)	3.96	0.02	3.61[a]	0.03	3.61	0.03	2.82[b]	0.10	3.69	0.04	3.03[c]	0.06	0.713
INR	0.99	0.01	1.18[a]	0.02	1.11	0.01	1.63[b]	0.10	1.15	0.03	1.36[c]	0.03	0.758
Protime (sec)	13.4	0.10	10.2[a]	0.33	13.6	0.13	19.4[b]	0.98	11.2	0.52	13.7[c]	0.64	0.591
AST (U/L)	22.8	0.34	53.2[a]	2.31	52.1	2.59	79.2[b]	10.4	52.6	3.97	61.7	4.85	0.876
ALT (U/L)	21.0	0.46	51.0[a]	2.60	51.0	3.24	46.0	5.54	49.0	4.09	40.6	3.55	0.825
Bilirubin (mg/dL)	0.62	0.03	1.58[a]	0.09	1.31	0.05	5.02[b]	0.68	1.42	0.12	3.04[c]	0.29	0.804
AST/ALT	1.15	0.01	1.22	0.02	1.21	0.03	1.79[b]	0.09	1.21	0.04	1.61[c]	0.05	0.500
MELD	7.13	0.10	10.3[a]	0.24	9.07	0.16	18.9	0.42	9.54	0.37	14.0[c]	0.46	0.747
APRI	NA	NA	0.93	0.06	1.05	0.07	2.44[b]	0.42	0.94	0.08	1.63[c]	0.18	NA

[a] Significant difference between patients vs. controls ($P<0.05$).

[b] Significant difference between medium- vs. low-MELD groups ($P<0.05$).

[c] Significant difference between decompensated vs. compensated patients ($P<0.05$).

* AUC from ROC analysis of pooled patients vs. controls.

NA: not applied

Table 2.13. Univariate logistic regression analysis of non-BA parameters.

Non-BA parameters	B-value (Regression Coefficient)	P-value	Exp(B)-Odds ratio		
			1 unit change	10% change	20% change
Creatinine (mg/dL)	0.852	0.085	2.34	1.08	1.16
Albumin (g/dL)	-1.616	0.000	0.19	0.53	0.28
INR	5.573	0.000	263	1.74	3.02
Protime (sec)	-0.063	0.002	0.94	0.92	0.85
AST (U/L)	0.141	0.000	1.15	1.38	1.90
ALT (U/L)	0.100	0.000	1.11	1.23	1.52
Bilirubin (mg/dL)	1.775	0.000	5.89	1.12	1.25
AST/ALT	0.542	0.064	1.72	1.06	1.13
MELD	0.470	0.000	1.59	1.39	1.95

Table 2.14.a. Non- BA parameters in controls and patients with specific liver disease subtype.

Non-BA parameters	Controls		Hepatitis C		Hepatitis B		Laennec Cirrhosis		Primary biliary cholangitis		Primary sclerosing cholangitis		Autoimmune Hepatitis	
	N=103		N=71		N=15		N=105		N=12		N=17		N=26	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Creatinine (mg/dL)	0.87	0.17	0.99	0.06	1.08[a]	0.08	1.00[a]	0.05	0.90	0.04	0.96	0.06	0.83	0.03
Albumin (g/dL)	3.96	0.30	3.48[a]	0.06	3.84	0.14	3.37[a]	0.05	3.61[a]	0.09	3.51[a]	0.10	3.42[a]	0.08
INR	0.99	0.20	1.17[a]	0.04	1.17	0.17	1.25[a]	0.03	0.97	0.03	1.16[a]	0.08	1.20[a]	0.09
Protime (sec)	13.4	1.89	8.20[a]	0.70	2.30[a]	1.26	11.4[a]	0.56	10.8[a]	0.93	8.09[a]	1.29	9.30[a]	1.07
AST (U/L)	22.8	6.23	68.4[a]	7.19	29.9[a]	3.59	55.6[a]	4.34	48.4[a]	8.21	59.5[a]	8.38	71.8[a]	9.25
ALT (U/L)	21.0	8.58	57.5[a]	5.58	28.5[a]	2.81	45.6[a]	4.12	46.3[a]	8.57	58.5[a]	9.41	76.4[a]	10.6
Bilirubin (mg/dL)	0.62	0.47	1.46[a]	0.14	0.88	0.17	2.10[a]	0.19	1.14	0.14	1.82[a]	0.44	1.59[a]	0.19
AST/ALT	1.15	0.26	1.24	0.05	1.07	0.07	1.38[a]	0.04	1.24	0.11	1.10	0.08	1.26	0.09
MELD	7.13	1.72	10.1[a]	0.56	13.5[a]	2.04	11.8[a]	0.43	8.03	0.47	10.6[a]	0.91	9.94[a]	0.75
APRI	NA	NA	1.45	0.20	0.24	0.06	1.24	0.12	0.65	0.15	0.79	0.17	0.85	0.12

[a] Significant difference between each specific liver disease subtype vs. controls (P<0.05)

Table 2.14.b. Non- BA parameters in controls and patients with specific liver disease subtype.

Non-BA parameters	α-1 antitrypsin deficiency		NASH		Carcinoma		Cryptogenic cirrhosis		Polycystic liver disease		Elevated LFT		Unknown Etiologys	
	N=5		N=52		N=25		N=11		N=3		N=19		N=5	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Creatinine (mg/dL)	1.01	0.04	0.95	0.04	0.93	0.05	2.47[a]	1.47	0.99	0.14	0.88	0.07	0.81	0.04
Albumin (g/dL)	3.48[a]	0.28	3.64[a]	0.06	3.50[a]	0.09	3.18[a]	0.11	4.07	0.18	3.70[a]	0.14	3.52[a]	0.12
INR	1.45[a]	0.15	1.14[a]	0.02	1.22[a]	0.05	1.36[a]	0.07	1.00	0.00	1.12	0.08	1.24[a]	0.22
Protime (sec)	7.22[a]	2.99	10.1[a]	0.67	10.0[a]	1.07	13.8	1.71	8.13[a]	4.15	8.40[a]	1.57	15.0[a]	2.56
AST (U/L)	40.5[a]	7.90	49.8[a]	4.14	61.0[a]	7.23	58.3[a]	6.88	21.3	5.24	64.4[a]	13.1	71.0[a]	8.08
ALT (U/L)	28.3[a]	4.07	55.2[a]	6.67	51.8[a]	7.20	37.0[a]	3.90	18.3	2.40	72.0[a]	12.7	87.9[a]	19.5
Bilirubin (mg/dL)	1.38[a]	0.18	1.31[a]	0.11	1.43[a]	0.19	2.51[a]	0.56	0.70	0.06	1.39[a]	0.36	1.75[a]	0.39
AST/ALT	1.41	0.14	1.12	0.05	1.27	0.07	1.61[a]	0.12	1.15	0.19	0.96[a]	0.11	1.13[a]	0.15
MELD	12.6[a]	1.74	9.81[a]	0.47	10.3[a]	0.65	13.2[a]	1.36	7.41	1.01	9.37[a]	1.31	10.2[a]	1.74
APRI	0.59	0.27	0.73	0.07	1.43	0.30	1.13	0.22	0.25	0.15	1.04	0.32	0.85	0.13

[a] Significant difference between controls and each specific liver disease subtype ($P<0.05$)

Table 2.15. Association between non-BA parameters and BA concentrations and indices.

Non-BA parameters vs.BA/BA indices	AST	ALT	AST/ ALT	Bilirubin	Albumin	INR	Protime	Creatinine	APRI	MELD
	P-value									
Total BA	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.614	0.000	0.000
Total LCA	0.000	0.001	0.000	0.000	0.000	0.066	0.008	0.900	0.003	0.001
Total UDCA	0.000	0.000	0.088	0.352	0.000	0.312	0.120	0.754	0.258	0.328
Total CDCA	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.643	0.000	0.000
Total DCA	0.000	0.000	0.127	0.134	0.004	0.749	0.925	0.919	0.020	0.946
Total HDCA	0.003	0.008	0.313	0.225	0.257	0.645	0.241	0.599	0.545	0.367
Total MDCA	0.012	0.361	0.006	0.001	0.012	0.781	0.607	0.928	0.561	0.112
Total CA	0.000	0.032	0.000	0.000	0.000	0.000	0.003	0.649	0.007	0.000
Total MCA	0.000	0.000	0.008	0.117	0.000	0.081	0.168	0.517	0.003	0.054
Total HCA	0.000	0.243	0.000	0.000	0.000	0.000	0.000	0.716	0.001	0.000
% LCA	0.000	0.001	0.051	0.003	0.000	0.049	0.996	0.768	0.046	0.000
% UDCA	0.519	0.470	0.126	0.013	0.001	0.001	0.799	0.831	0.008	0.000
% CDCA	0.001	0.379	0.000	0.000	0.000	0.000	0.000	0.574	0.000	0.000
% DCA	0.001	0.141	0.000	0.000	0.000	0.000	0.003	0.001	0.042	0.000
% HDCA	0.044	0.203	0.155	0.126	0.230	0.144	0.877	0.696	0.239	0.052
% MDCA	0.562	0.700	0.395	0.180	0.103	0.013	0.001	0.788	0.258	0.000
% CA	0.004	0.057	0.558	0.000	0.000	0.031	0.813	0.656	0.125	0.000
% MCA	0.126	0.085	0.682	0.093	0.565	0.924	0.880	0.314	0.505	0.548
% HCA	0.095	0.850	0.007	0.000	0.000	0.002	0.027	0.600	0.012	0.000
Total Unamidated	0.000	0.011	0.004	0.037	0.000	0.069	0.084	0.743	0.046	0.009
Total G-amidated	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.650	0.000	0.000
Total T-amidated	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.658	0.000	0.000
% Amidation	0.000	0.009	0.006	0.002	0.000	0.005	0.002	0.157	0.002	0.006
% G-amidation	0.710	0.169	0.265	0.007	0.957	0.910	0.458	0.113	0.582	0.152
% T-amidation	0.000	0.244	0.000	0.000	0.000	0.000	0.002	0.373	0.000	0.000
Total Unsulfated	0.000	0.000	0.000	0.000	0.000	0.002	0.006	0.616	0.000	0.000
Total Sulfated	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.628	0.000	0.000
% Sulfation	0.660	0.842	0.596	0.766	0.255	0.989	0.016	0.276	0.420	0.121
Total Mono-OH	0.000	0.001	0.000	0.000	0.000	0.066	0.008	0.900	0.003	0.001
Total Di-OH	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.644	0.000	0.000
Total Tri-OH	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.533	0.000	0.000
% Mono-OH	0.000	0.001	0.051	0.003	0.000	0.049	0.996	0.768	0.046	0.000
% Di-OH	0.200	0.351	0.207	0.384	0.078	0.721	0.094	0.156	0.448	0.948
% Tri-OH	0.003	0.012	0.761	0.175	0.019	0.221	0.879	0.148	0.109	0.003
Total 12α-OH	0.000	0.000	0.000	0.000	0.000	0.047	0.135	0.778	0.002	0.000
Total non-12α-OH	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.605	0.000	0.000
12α-OH/ non12α-OH	0.194	0.271	0.002	0.017	0.000	0.003	0.001	0.102	0.237	0.000
CA/ CDCA	0.000	0.000	0.423	0.709	0.733	0.074	0.039	0.456	0.008	0.749
% 12α-OH	0.039	0.511	0.000	0.001	0.000	0.000	0.000	0.085	0.142	0.000
% non-12α-OH	0.039	0.511	0.000	0.001	0.000	0.000	0.000	0.085	0.142	0.000
Total Primary	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.592	0.000	0.000
Total Secondary	0.000	0.000	0.030	0.118	0.000	0.309	0.111	0.747	0.101	0.221
Primary/ Secondary	0.014	0.355	0.000	0.002	0.000	0.000	0.000	0.993	0.005	0.000
% Primary	0.000	0.063	0.000	0.000	0.000	0.000	0.001	0.218	0.000	0.000
%Secondary	0.000	0.063	0.000	0.000	0.000	0.000	0.001	0.218	0.000	0.000
HI	0.078	0.067	0.411	0.493	0.026	0.459	0.896	0.584	0.263	0.914

Appendix

Table A. Full list of BA concentrations and indices in controls vs. patients.

BA (μM) / BA indices (%)	Controls		Patients		Patients vs. Controls	
	mean	SEM	Mean	SEM	Ratio	p-value
LCA-S	0.01	0.00	0.03	0.01	4.48	0.00
G-LCA-S	0.78	0.04	2.23	0.20	2.84	0.00
T-LCA-S	0.22	0.01	0.65	0.06	2.96	0.00
UDCA-S	0.45	0.02	1.56	0.23	3.47	0.01
G-UDCA-S	1.04	0.05	15.3	2.68	14.7	0.00
T-UDCA-S	0.03	0.00	1.23	0.27	46.7	0.01
CDCA-S	0.07	0.01	0.19	0.03	2.83	0.00
G-CDCA-S	2.38	0.13	18.7	1.79	7.86	0.00
T-CDCA-S	0.06	0.00	1.91	0.38	34.5	0.00
DCA-S	0.01	0.00	0.04	0.01	6.95	0.04
G-DCA-S	2.90	0.14	4.28	0.54	1.48	0.02
T-DCA-S	0.22	0.02	0.52	0.07	2.35	0.01
CA-S	0.00	0.00	0.08	0.01	19.1	0.00
G-CA-S	0.06	0.01	0.91	0.13	16.1	0.00
T-CA-S	0.13	0.01	1.03	0.21	8.14	0.04
Sulfated BA	8.35	0.31	48.7	4.77	5.83	0.00
Sulfated + Unamidated BA	0.54	0.03	1.90	0.24	3.56	0.00
Sulfated G-BA	7.17	0.28	41.4	4.12	5.78	0.00
Sulfated T-BA	0.65	0.03	5.34	0.74	8.23	0.00
Sulfated U-, G-, T-LCA	1.01	0.05	2.91	0.24	2.88	0.00
Sulfated U-, G-, T-UDCA	1.52	0.07	18.1	3.08	11.9	0.00
Sulfated U-, G-, T-CDCA	2.51	0.13	20.8	2.07	8.31	0.00
Sulfated U-, G-, T-DCA	3.13	0.16	4.84	0.58	1.55	0.01
Sulfated U-, G-, T-CA	0.19	0.01	2.01	0.31	10.8	0.00
Sulfated 12α-OH BA	3.31	0.16	6.85	0.74	2.07	0.00
Sulfated non-12α-OH BA	5.04	0.19	41.8	4.31	8.30	0.00
Sulfated primary BA	2.69	0.13	22.8	2.29	8.48	0.00
Sulfated secondary BA	5.66	0.23	25.8	3.36	4.56	0.00
LCA	0.00	0.00	0.00	0.00	32.9	0.00
G-LCA	0.00	0.00	0.00	0.00	3.28	0.93
T-LCA	0.00	0.00	0.00	0.00	0.49	0.66
UDCA	0.00	0.00	0.08	0.03	22.5	0.03
G-UDCA	0.03	0.00	0.41	0.17	12.5	0.06
T-UDCA	0.00	0.00	0.01	0.00	6.73	0.09
CDCA	0.00	0.00	0.02	0.00	4.65	0.01
G-CDCA	0.01	0.00	0.09	0.01	11.7	0.00
T-CDCA	0.00	0.00	0.10	0.02	62.4	0.01
DCA	0.02	0.00	0.04	0.00	1.71	0.01
G-DCA	0.01	0.00	0.04	0.00	3.81	0.00
T-DCA	0.00	0.00	0.01	0.00	3.70	0.01
HDCA	0.01	0.00	0.01	0.00	1.54	0.08
G-HDCA	0.00	0.00	0.00	0.00	7.29	0.09
MDCA	0.06	0.01	0.05	0.01	0.90	0.82
CA	0.18	0.03	0.24	0.03	1.36	0.09
G-CA	0.07	0.00	0.55	0.07	8.24	0.00
T-CA	0.01	0.00	0.32	0.08	35.5	0.10
MCA	0.03	0.00	0.12	0.02	4.27	0.00
G-MCA	0.29	0.02	1.94	0.29	6.74	0.00
T-MCA	0.04	0.00	0.73	0.09	18.0	0.00
HCA	0.01	0.00	0.01	0.00	0.72	0.46
G-HCA	0.02	0.00	0.17	0.02	10.8	0.00
T-HCA	0.00	0.00	0.09	0.02	70.0	0.07
Nor-DCA	0.00	0.00	0.01	0.00	1.21	0.25
12-oxo-CDCA	0.12	0.01	0.52	0.07	4.43	0.00
3-dehydroCA	0.01	0.00	0.29	0.11	19.7	0.01
6-oxo-LCA	0.00	0.00	0.01	0.00	19.1	0.00
7-oxo-LCA	0.00	0.00	0.02	0.00	6.41	0.00
12-oxo-LCA	0.01	0.00	0.01	0.00	0.67	0.33
isoLCA	0.00	0.00	0.01	0.00	176	0.00
isoDCA	0.00	0.00	0.00	0.00	1.49	0.69
Unsulfated BA	0.94	0.05	5.91	0.57	6.26	0.00
Unsulfated + Unamidated BA	0.46	0.04	1.43	0.16	3.08	0.00
Unsulfated G-BA	0.42	0.02	3.20	0.44	7.59	0.00
Unsulfated T-BA	0.06	0.00	1.27	0.15	22.4	0.00
Unsulfated U-, G-, T-LCA	0.00	0.00	0.01	0.00	7.53	0.05
Unsulfated U-, G-, T-UDCA	0.04	0.00	0.50	0.21	13.2	0.05
Unsulfated U-, G-, T-CDCA	0.01	0.00	0.21	0.03	16.4	0.01

Unsulfated U-, G-, T-DCA	0.03	0.00	0.09	0.01	2.48	0.00
Unsulfated U-, G-, T-HDCA	0.01	0.00	0.01	0.00	1.57	0.08
Unsulfated U-, G-, T-MDCA	0.06	0.01	0.05	0.01	0.90	0.82
Unsulfated U-, G-, T-CA	0.25	0.03	1.12	0.14	4.39	0.01
Unsulfated U-, G-, T-MCA	0.36	0.02	2.79	0.34	7.83	0.00
Unsulfated U-, G-, T-HCA	0.03	0.00	0.27	0.04	10.6	0.00
Unsulfated 12α-OH BA	0.31	0.03	1.50	0.19	4.86	0.00
Unsulfated non-12α-OH BA	0.63	0.03	4.40	0.50	6.94	0.00
Unsulfated primary BA	0.65	0.04	4.39	0.43	6.76	0.00
Unsulfated secondary BA	0.29	0.02	1.52	0.25	5.17	0.00
Total BA	9.30	0.33	54.6	5.20	5.87	0.00
Total Unamidated BA	1.00	0.05	3.33	0.33	3.34	0.00
Total primary unamidated BA	0.29	0.03	0.66	0.06	2.26	0.00
Total secondary unamidated BA	0.71	0.03	2.68	0.30	3.78	0.00
Total primary amidated BA	3.05	0.15	26.6	2.57	8.71	0.00
Total secondary amidated BA	5.24	0.21	24.7	3.29	4.70	0.00
Total G-BA	7.59	0.29	44.6	4.46	5.88	0.00
Total primary G-BA	2.82	0.14	22.4	2.09	7.94	0.00
Total secondary G-BA	4.77	0.20	22.3	3.07	4.66	0.00
Total T-BA	0.71	0.03	6.61	0.85	9.37	0.00
Total primary T-BA	0.23	0.01	4.19	0.68	17.9	0.00
Total secondary T-BA	0.47	0.02	2.42	0.32	5.13	0.00
Total U-, G-, T-LCA	1.01	0.05	2.92	0.24	2.88	0.00
Total U-, G-, T-UDCA	1.56	0.07	18.6	3.23	11.9	0.00
Total U-, G-, T-CDCA	2.52	0.13	21.0	2.09	8.35	0.00
Total U-, G-, T-DCA	3.16	0.16	4.92	0.58	1.56	0.07
Total U-, G-, T-HDCA	0.01	0.00	0.01	0.00	1.57	0.05
Total U-, G-, T-MDCA	0.06	0.01	0.05	0.01	0.90	0.99
Total U-, G-, T-CA	0.44	0.03	3.13	0.44	7.09	0.00
Total U-, G-, T-MCA	0.36	0.02	2.79	0.34	7.83	0.00
Total U-, G-, T-HCA	0.03	0.00	0.27	0.04	10.6	0.00
Total 12α-OH BA	3.62	0.17	8.35	0.83	2.30	0.00
Total non-12α-OH BA	5.67	0.20	46.2	4.68	8.15	0.00
Total primary BA	3.34	0.15	27.2	2.59	8.15	0.00
Total secondary BA	5.95	0.23	27.3	3.52	4.59	0.00
Total Mono-hydroxyl BA	1.01	0.05	2.92	0.24	2.88	0.00
Total Di-hydroxyl BA	7.30	0.29	44.6	4.58	6.11	0.00
Total Tri-hydroxyl BA	0.82	0.04	6.19	0.65	7.52	0.00
%Sulfated BA	88.5%	0.46%	83.0%	0.60%	0.94	0.00
%Sulfation LCA	99.7%	0.05%	99.2%	0.16%	0.99	0.85
%Sulfation UDCA	97.0%	0.16%	97.4%	0.20%	1.00	0.35
%Sulfation CDCA	99.4%	0.04%	98.7%	0.12%	0.99	0.25
%Sulfation DCA	98.3%	0.09%	96.2%	0.35%	0.98	0.01
%Sulfation CA	53.0%	1.22%	52.6%	1.05%	0.99	0.97
%Sulfation on mono-hydroxy	99.7%	0.05%	99.2%	0.16%	0.99	0.85
%Sulfation on dihydroxy	97.6%	0.09%	97.4%	0.16%	1.00	0.23
%Sulfation on tri-hydroxy	26.7%	0.77%	26.5%	0.81%	0.99	0.85
%Sulfation on primary BA	79.2%	0.66%	76.8%	0.68%	0.97	0.93
%Sulfation on secondary BA	93.3%	0.41%	86.9%	0.76%	0.93	0.00
%Sulfation on 12α-OH BA	87.1%	0.85%	75.9%	0.95%	0.87	0.00
%Sulfation on non-12α-OH BA	87.8%	0.43%	84.2%	0.60%	0.96	0.06
% Sulfation on unamidated BA	58.0%	1.02%	49.6%	1.08%	0.85	0.00
%Sulfation on amidated BA	93.6%	0.23%	88.9%	0.44%	0.95	0.00
%Sulfation on G-BA	93.7%	0.22%	90.2%	0.40%	0.96	0.00
%Sulfation on T-BA	91.5%	0.31%	80.9%	0.78%	0.89	0.00
%Unamidated BA	12.3%	0.47%	13.2%	0.65%	1.07	0.01
% G-Amidated BA	79.7%	0.49%	76.0%	0.71%	0.95	0.00
% T-Amidated BA	8.0%	0.26%	10.8%	0.46%	1.35	0.01
% Amidated BA	87.7%	0.47%	86.8%	0.65%	0.99	0.05
%Amidation LCA	98.7%	0.14%	98.1%	0.21%	0.99	0.39
%Amidation UDCA	68.5%	0.72%	78.0%	0.95%	1.14	0.00
%Amidation CDCA	95.8%	0.28%	97.8%	0.20%	1.02	0.00
%Amidation DCA	98.5%	0.12%	95.9%	0.44%	0.97	0.04
%Amidation CA	71.1%	1.18%	77.2%	1.07%	1.09	0.01
%Amidation MCA	88.9%	0.69%	90.8%	0.75%	1.02	0.94
%Amidation HCA	75.4%	1.12%	92.8%	0.81%	1.23	0.00
%Unamidation LCA	1.31%	0.14%	1.90%	0.21%	1.46	0.39
%G-Amidation LCA	75.6%	0.54%	74.7%	0.81%	0.99	0.84
%T-Amidation LCA	23.1%	0.55%	23.4%	0.80%	1.01	0.99

%Unamidation UDCA	31.5%	0.72%	22.0%	0.95%	0.70	0.00
%G-Amidation UDCA	66.2%	0.72%	70.0%	1.17%	1.06	0.98
%T-Amidation UDCA	2.25%	0.09%	8.08%	0.54%	3.60	0.00
%Unamidation CDCA	4.20%	0.28%	2.21%	0.20%	0.53	0.00
%G-Amidation CDCA	93.1%	0.32%	91.8%	0.43%	0.99	0.55
%T-Amidation CDCA	2.72%	0.10%	6.01%	0.40%	2.21	0.02
%Unamidation DCA	1.53%	0.12%	4.08%	0.44%	2.66	0.04
%G-Amidation DCA	91.6%	0.28%	83.0%	0.81%	0.91	0.00
%T-Amidation DCA	6.88%	0.26%	13.0%	0.72%	1.88	0.00
%Unamidation HDCA	98.7%	0.56%	94.0%	0.85%	0.95	0.04
%G-Amidation HDCA	1.29%	0.56%	5.94%	0.85%	4.62	0.03
%Unamidation CA	28.9%	1.18%	22.8%	1.07%	0.79	0.01
%G-Amidation CA	34.1%	0.88%	52.0%	0.93%	1.53	0.00
%T-Amidation CA	37.0%	1.18%	25.2%	0.94%	0.68	0.00
%Unamidation MCA	11.1%	0.69%	9.23%	0.75%	0.83	0.94
%G-Amidation MCA	78.3%	0.65%	69.7%	0.98%	0.89	0.00
%T-Amidation MCA	10.6%	0.35%	21.1%	0.83%	1.98	0.00
%Unamidation HCA	24.6%	1.12%	7.22%	0.81%	0.29	0.00
%G-Amidation HCA	68.9%	1.10%	76.2%	1.06%	1.11	0.00
%T-Amidation HCA	6.50%	0.27%	16.5%	0.87%	2.54	0.00
%Unamidation on primary BA	9.53%	0.57%	6.41%	0.41%	0.67	0.00
%Unamidation on secondary BA	10.3%	0.29%	9.59%	0.46%	0.93	0.00
%Unamidation on 12α-OH	9.41%	0.70%	14.2%	0.76%	1.51	0.00
%Unamidation on non-12α-OH	15.0%	0.48%	12.9%	0.68%	0.86	0.00
%Amidation on monohydroxy	98.7%	0.14%	98.1%	0.21%	0.99	0.39
%Amidation on dihydroxy	90.3%	0.30%	90.8%	0.52%	1.01	0.00
%Amidation on trihydroxy	77.9%	0.97%	83.9%	0.81%	1.08	0.00
%Amidation on primary BA	90.5%	0.57%	93.6%	0.41%	1.03	0.00
%Amidation on secondary BA	85.7%	0.52%	80.0%	0.92%	0.93	0.29
%Amidation on 12α-OH	90.6%	0.70%	85.8%	0.76%	0.95	0.00
%Amidation on non-12α-OH	87.9%	0.35%	92.1%	0.39%	1.05	0.00
%G-amidation on monohydroxy	75.6%	0.54%	74.7%	0.81%	0.99	0.84
%T-amidation on monohydroxy	23.1%	0.55%	23.4%	0.80%	1.01	0.99
%G-amidation on dihydroxy	86.0%	0.35%	83.1%	0.68%	0.97	0.00
%T-amidation on dihydroxy	4.32%	0.17%	7.78%	0.40%	1.80	0.12
%G-amidation on trihydroxy	54.1%	0.79%	61.2%	0.78%	1.13	0.00
%T-amidation on trihydroxy	23.8%	0.74%	22.8%	0.78%	0.96	0.58
%G-amidation on primary BA	82.1%	0.63%	83.2%	0.53%	1.01	0.01
%G-amidation on secondary BA	77.6%	0.53%	68.6%	0.97%	0.88	0.05
%G-amidation on 12α-OH	80.2%	0.71%	68.8%	0.88%	0.86	0.00
%G-amidation on non-12α-OH	78.2%	0.51%	77.4%	0.75%	0.99	0.00
%T-amidation on primary BA	8.38%	0.34%	10.4%	0.46%	1.24	0.51
%T-amidation on secondary BA	8.18%	0.28%	11.4%	0.51%	1.39	0.17
%T-amidation on 12α-OH	10.4%	0.34%	17.0%	0.75%	1.64	0.00
%T-amidation on non-12α-OH	6.81%	0.22%	9.77%	0.42%	1.43	0.19
%Total Mono-hydroxyl BA: Total BA	11.5%	0.38%	9.16%	0.39%	0.79	0.00
%Total Di-hydroxyl BA: Total BA	76.6%	0.50%	72.6%	0.65%	0.95	0.00
%Total Tri-hydroxyl BA: Total BA	9.58%	0.33%	13.1%	0.43%	1.37	0.00
%12α-OH: Total BA	36.7%	0.62%	22.1%	0.54%	0.60	0.00
%non-12α-OH: Total BA	63.3%	0.62%	77.9%	0.54%	1.23	0.00
% Primary BA: Total BA	36.7%	0.70%	49.4%	1.06%	1.35	0.00
%Secondary BA: Total BA	63.3%	0.70%	50.6%	1.06%	0.80	0.00
% Total U-, G-, T-LCA : Total BA	11.5%	0.38%	9.16%	0.39%	0.79	0.00
% Total U-, G-, T-UDCA : Total BA	17.7%	0.49%	21.3%	0.88%	1.21	0.14
% Total U-, G-, T-CDCA : Total BA	27.1%	0.65%	36.3%	0.94%	1.34	0.00

% Total U-, G-, T-DCA : Total BA	31.1%	0.68%	14.6%	0.53%	0.47	0.00
% Total U-, G-, T-HDCA : Total BA	0.07%	0.01%	0.04%	0.00%	0.54	0.05
% Total U-, G-, T-MDCA : Total BA	0.64%	0.04%	0.36%	0.05%	0.56	0.14
% Total U-, G-, T-CA : Total BA	5.25%	0.27%	6.27%	0.25%	1.19	0.06
% Total U-, G-, T-MCA : Total BA	4.03%	0.16%	6.39%	0.34%	1.58	0.00
% Total U-, G-, T-HCA : Total BA	0.30%	0.02%	0.45%	0.04%	1.52	0.02
%LCA: Unamidated BA	0.91%	0.05%	1.41%	0.10%	1.55	0.03
%UDCA: Unamidated BA	50.1%	1.03%	38.1%	1.11%	0.76	0.00
%CDCA: Unamidated BA	6.70%	0.26%	6.85%	0.43%	1.02	0.16
%DCA: Unamidated BA	3.30%	0.13%	4.08%	0.23%	1.23	0.02
%HDCA: Unamidated BA	0.71%	0.07%	0.52%	0.07%	0.73	0.49
%MDCA: Unamidated BA	6.97%	0.39%	4.47%	0.31%	0.64	0.02
%CA: Unamidated BA	13.6%	0.72%	13.0%	0.60%	0.96	0.28
%MCA: Unamidated BA	2.98%	0.16%	4.20%	0.31%	1.41	0.00
%HCA: Unamidated BA	0.93%	0.10%	0.40%	0.14%	0.44	0.02
%G,T-LCA: Amidated BA	13.0%	0.42%	10.9%	0.48%	0.84	0.00
%G,T-UDCA: Amidated BA	14.5%	0.53%	19.4%	0.88%	1.33	0.02
%G,T-CDCA: Amidated BA	29.8%	0.72%	40.5%	0.96%	1.36	0.00
%G,T-DCA: Amidated BA	34.7%	0.73%	16.5%	0.58%	0.47	0.00
%G,T-HDCA: Amidated BA	0.00%	0.00%	0.00%	0.00%	3.43	0.09
%G,T-CA: Amidated BA	3.48%	0.12%	5.46%	0.24%	1.57	0.00
%G,T-MCA: Amidated BA	4.21%	0.18%	6.91%	0.36%	1.64	0.00
%G,T-HCA: Amidated BA	0.23%	0.01%	0.44%	0.03%	1.91	0.03
%G-LCA: G-Amidated BA	11.1%	0.38%	9.98%	0.49%	0.90	0.00
%G-UDCA: G-Amidated BA	15.4%	0.54%	19.0%	0.89%	1.24	0.02
%G-CDCA: G-Amidated BA	31.8%	0.74%	43.6%	1.02%	1.37	0.00
%G-DCA: G-Amidated BA	35.6%	0.75%	16.7%	0.59%	0.47	0.00
%G-HDCA: G-Amidated BA	0.00%	0.00%	0.00%	0.00%	3.61	0.08
%G-CA: G-Amidated BA	1.83%	0.08%	4.23%	0.22%	2.31	0.00
%G-MCA: G-Amidated BA	4.10%	0.17%	6.06%	0.31%	1.48	0.01
%G-HCA: G-Amidated BA	0.23%	0.01%	0.39%	0.03%	1.69	0.09
%T-LCA: T-Amidated BA	32.2%	0.78%	19.9%	0.79%	0.62	0.00
%T-UDCA: T-Amidated BA	5.28%	0.26%	16.5%	0.94%	3.12	0.00
%T-CDCA: T-Amidated BA	10.0%	0.37%	18.8%	0.79%	1.88	0.00
%T-DCA: T-Amidated BA	26.3%	0.67%	15.5%	0.61%	0.59	0.00
%T-CA: T-Amidated BA	20.5%	0.60%	15.5%	0.63%	0.76	0.04
%T-MCA: T-Amidated BA	5.48%	0.25%	13.2%	0.73%	2.40	0.00
%T-HCA: T-Amidated BA	0.24%	0.01%	0.62%	0.06%	2.59	0.02
%Sulfated LCA : Sulfated BA	12.9%	0.42%	11.1%	0.46%	0.86	0.00
%Sulfated UDCA : Sulfated BA	19.7%	0.58%	24.7%	0.93%	1.25	0.25
%Sulfated CDCA : Sulfated BA	30.7%	0.73%	43.5%	1.04%	1.42	0.00
%Sulfated DCA : Sulfated BA	34.2%	0.70%	17.1%	0.59%	0.50	0.00
%Sulfated CA : Sulfated BA	2.42%	0.09%	3.57%	0.16%	1.48	0.00
%Unsulfated LCA : unsulfated BA	0.16%	0.02%	0.18%	0.02%	1.09	0.02
%Unsulfated UDCA : unsulfated BA	4.53%	0.20%	5.12%	0.47%	1.13	0.63
%Unsulfated CDCA : unsulfated BA	1.60%	0.07%	3.12%	0.24%	1.95	0.01
%Unsulfated DCA : unsulfated BA	5.01%	0.18%	2.81%	0.17%	0.56	0.00
%Unsulfated HDCA : unsulfated BA	0.84%	0.07%	0.33%	0.04%	0.39	0.00
%Unsulfated MDCA : unsulfated BA	7.15%	0.34%	2.63%	0.21%	0.37	0.00
%Unsulfated CA : unsulfated BA	22.1%	0.77%	21.62%	0.77%	0.98	0.89
%Unsulfated MCA : unsulfated BA	40.5%	0.90%	38.84%	1.18%	0.96	0.81
%Unsulfated HCA : unsulfated BA	3.39%	0.19%	3.74%	0.28%	1.10	0.66
%Primary unamidated BA:Total BA	3.55%	0.27%	2.61%	0.19%	0.74	0.07
%Secondary unamidated BA:Total BA	8.8%	0.32%	10.5%	0.58%	1.20	0.02

%Primary amidated BA:Total BA	33.1%	0.68%	46.8%	1.06%	1.41	0.00
%Secondary amidated BA:Total BA	54.6%	0.71%	40.0%	0.96%	0.73	0.00
%Primary G-BA:Total BA	30.4%	0.67%	40.8%	0.90%	1.34	0.00
%Secondary G-BA:Total BA	49.3%	0.66%	35.2%	0.93%	0.71	0.00
%Primary T-BA:Total BA	2.73%	0.10%	6.00%	0.35%	2.20	0.00
%SecondaryT-BA:Total BA	5.25%	0.19%	4.81%	0.22%	0.92	0.00
%Unamidated -Unsulfated BA: Total BA	6.03%	0.38%	7.57%	0.48%	1.25	0.53
%Amidated -Unsulfated G-BA : Total BA	4.86%	0.16%	7.04%	0.28%	1.45	0.00
%Amidated -Unsulfated T-BA : Total BA	0.65%	0.03%	2.42%	0.17%	3.73	0.00
%Amidated -Unsulfated BA: Total BA	5.50%	0.19%	9.46%	0.38%	1.72	0.00
% Unamidated: Unsulfated BA	44.5%	1.01%	36.9%	1.20%	0.83	0.00
%Amidated G: Unsulfated BA	48.7%	0.88%	47.8%	0.94%	0.98	0.57
%Amidated T: Unsulfated BA	6.74%	0.24%	15.3%	0.71%	2.26	0.00
% Amidated: Unsulfated BA	55.5%	1.01%	63.1%	1.20%	1.14	0.00
%Unamidated -Sulfated BA: Total BA	6.27%	0.20%	5.58%	0.31%	0.89	0.00
%Amidated -Sulfated G-BA : Total BA	74.9%	0.54%	69.0%	0.76%	0.92	0.57
%Amidated -Sulfated T-BA : Total BA	7.33%	0.24%	8.39%	0.36%	1.15	0.33
%Amidated -Sulfated BA: Total BA	82.2%	0.53%	77.4%	0.72%	0.94	0.29
% Unamidated: BA-Sulfates	7.29%	0.25%	7.33%	0.44%	1.01	0.00
% Amidated G: BA-Sulfates	84.4%	0.35%	82.5%	0.61%	0.98	0.00
%Amidated T: BA-sulfates	8.28%	0.27%	10.1%	0.43%	1.22	0.66
% Amidated: BA-Sulfates	92.7%	0.25%	92.7%	0.44%	1.00	0.00
%Unsulfated LCA: Total BA	0.01%	0.00%	0.03%	0.00%	1.90	0.09
%Unsulfated UDCA: Total BA	0.48%	0.04%	0.52%	0.05%	1.09	0.90
%Unsulfated CDCA: Total BA	0.15%	0.01%	0.42%	0.04%	2.76	0.00
%Unsulfated DCA: Total BA	0.43%	0.02%	0.38%	0.02%	0.87	0.02
%Unsulfated HDCA: Total BA	0.07%	0.01%	0.04%	0.00%	0.54	0.00
%Unsulfated MDCA: Total BA	0.64%	0.04%	0.36%	0.05%	0.56	0.14
%Unsulfated CA: Total BA	3.15%	0.26%	3.36%	0.20%	1.07	0.95
%Unsulfated MCA: Total BA	4.03%	0.16%	6.39%	0.34%	1.58	0.00
%Unsulfated HCA: Total BA	0.30%	0.02%	0.45%	0.04%	1.52	0.12
%Unsulfated Mono-hydroxyl BA: Total BA	0.01%	0.00%	0.03%	0.00%	1.90	0.09
%Unsulfated Di-hydroxyl BA: Total BA	1.78%	0.06%	1.72%	0.09%	0.97	0.10
%Unsulfated Tri-hydroxyl BA: Total BA	7.48%	0.31%	10.2%	0.39%	1.36	0.00
% Unsulfated primary: Total BA	7.63%	0.32%	10.6%	0.40%	1.39	0.00
% Unsulfated Secondary: Total BA	3.91%	0.24%	6.40%	0.43%	1.64	0.05
% Unsulfated 12α-OH: Total BA	3.85%	0.27%	4.94%	0.29%	1.28	0.04
% Unsulfated non 12α-OH: Total BA	7.69%	0.29%	12.1%	0.47%	1.57	0.00
%Sulfated LCA: Total BA	11.5%	0.38%	9.13%	0.39%	0.79	0.00
%Sulfated UDCA: Total BA	17.2%	0.48%	20.8%	0.86%	1.21	0.32
%Sulfated CDCA: Total BA	26.9%	0.64%	35.9%	0.93%	1.33	0.00
%Sulfated DCA: Total BA	30.7%	0.68%	14.2%	0.52%	0.46	0.00
%Sulfated CA: Total BA	2.11%	0.07%	2.91%	0.13%	1.38	0.00
%Sulfated Mono-hydroxyl BA: Total BA	11.5%	0.38%	9.13%	0.39%	0.79	0.00
%Sulfated Di-hydroxyl BA: Total BA	74.8%	0.53%	70.9%	0.67%	0.95	0.46
%Sulfated Tri-hydroxyl BA: Total BA	2.11%	0.07%	2.91%	0.13%	1.38	0.00
% Sulfated primary: Total BA	29.0%	0.64%	38.8%	0.98%	1.34	0.00

% Sulfated Secondary: Total BA	59.4%	0.75%	44.2%	1.02%	0.74	0.00
% Sulfated 12α-OH: Total BA	32.8%	0.69%	17.1%	0.52%	0.52	0.00
% Sulfated non 12α-OH: Total BA	55.6%	0.62%	65.9%	0.71%	1.18	0.00
% Unsulfated Nor-DCA	0.05%	0.00%	0.03%	0.00%	0.55	0.00
% Unsulfated 12-oxo-CDCA	1.71%	0.19%	3.53%	0.35%	2.07	0.33
% Unsulfated 3-dehydroCA	0.21%	0.02%	1.18%	0.16%	5.52	0.00
% Unsulfated 6-oxo-LCA	0.00%	0.00%	0.06%	0.01%	17.2	0.00
% Unsulfated 7-oxo-LCA	0.03%	0.00%	0.09%	0.02%	3.55	0.06
% Unsulfated 12-oxo-LCA	0.24%	0.09%	0.06%	0.01%	0.26	0.13
% Unsulfated isoLCA	0.00%	0.00%	0.11%	0.02%	189	0.02
% Unsulfated isoDCA	0.03%	0.00%	0.02%	0.00%	0.72	0.06
Primary BA/Secondary BA	0.69	0.03	2.52	0.22	3.63	0.00
Total amidated/total unamidated	12.3	0.51	29.8	2.14	2.42	0.00
12α-OH/non12α-OH	0.65	0.02	0.33	0.01	0.51	0.00
CA/CDCA	0.24	0.01	0.24	0.02	1.00	0.63
CA+DCA/CDCA+LCA	1.15	0.04	0.60	0.03	0.52	0.00
Total BA for HI	0.40	0.03	2.92	0.33	7.30	0.00
HI	-0.01	0.01	-0.12	0.01	17.6	0.00

Table B. Area under the ROC curve (AUC) of the full list BA concentrations and indices*.

BA (μM) / BA indices (%)	AUC
LCA-S	0.526
G-LCA-S	0.572
T-LCA-S	0.564
UDCA-S	0.560
G-UDCA-S	0.702
T-UDCA-S	0.664
CDCA-S	0.511
G-CDCA-S	0.770
T-CDCA-S	0.716
DCA-S	0.678
G-DCA-S	0.524
T-DCA-S	0.537
CA-S	0.690
G-CA-S	0.809
T-CA-S	0.594
Sulfated BA	0.724
Sulfated + Unamidated BA	0.512
Sulfated G-BA	0.729
Sulfated T-BA	0.666
Sulfated U-, G-, T-LCA	0.583
Sulfated U-, G-, T-UDCA	0.662
Sulfated U-, G-, T-CDCA	0.770
Sulfated U-, G-, T-DCA	0.515
Sulfated U-, G-, T-CA	0.728
Sulfated 12α-OH BA	0.555
Sulfated non-12α-OH BA	0.765
Sulfated primary BA	0.771
Sulfated secondary BA	0.605
LCA	0.531
G-LCA	0.658
T-LCA	0.640
UDCA	0.519
G-UDCA	0.590
T-UDCA	0.619
CDCA	0.633
G-CDCA	0.690
T-CDCA	0.616
DCA	0.527
G-DCA	0.544
T-DCA	0.611

HDCA	0.657
G-HDCA	0.505
T-HDCA	0.501
MDCA	0.576
CA	0.540
G-CA	0.793
T-CA	0.674
MCA	0.513
G-MCA	0.663
T-MCA	0.711
HCA	0.769
G-HCA	0.646
T-HCA	0.504
Nor-DCA	0.672
12-oxo-CDCA	0.549
3-dehydroCA	0.596
6-oxo-LCA	0.605
7-oxo-LCA	0.581
12-oxo-LCA	0.499
isoLCA	0.517
isoDCA	0.663
Unsulfated BA	0.771
Unsulfated + Unamidated BA	0.638
Unsulfated G-BA	0.760
Unsulfated T-BA	0.746
Unsulfated U-, G-, T-LCA	0.590
Unsulfated U-, G-, T-UDCA	0.603
Unsulfated U-, G-, T-CDCA	0.692
Unsulfated U-, G-, T-DCA	0.540
Unsulfated U-, G-, T-HDCA	0.659
Unsulfated U-, G-, T-MDCA	0.424
Unsulfated U-, G-, T-CA	0.736
Unsulfated U-, G-, T-MCA	0.691
Unsulfated U-, G-, T-HCA	0.623
Unsulfated 12α-OH BA	0.769
Unsulfated non-12α-OH BA	0.745
Unsulfated primary BA	0.759
Unsulfated secondary BA	0.670
Total BA	0.736
Total Unamidated BA	0.592
Total primary unamidated BA	0.593
Total secondary unamidated BA	0.574
Total primary amidated BA	0.780

Total secondary amidated BA	0.613
Total G-BA	0.736
Total primary G-BA	0.782
Total secondary G-BA	0.605
Total T-BA	0.690
Total primary T-BA	0.713
Total secondary T-BA	0.614
Total U-, G-, T-LCA	0.583
Total U-, G-, T-UDCA	0.662
Total U-, G-, T-CDCA	0.770
Total U-, G-, T-DCA	0.514
Total U-, G-, T-HDCA	0.659
Total U-, G-, T-MDCA	0.576
Total U-, G-, T-CA	0.747
Total U-, G-, T-MCA	0.691
Total U-, G-, T-HCA	0.623
Total 12α-OH BA	0.599
Total non-12α-OH BA	0.769
Total primary BA	0.776
Total secondary BA	0.626
Total Mono-hydroxyl BA	0.583
Total Di-hydroxyl BA	0.726
Total Tri-hydroxyl BA	0.767
%Sulfated BA	0.581
%Sulfation LCA	0.610
%Sulfation UDCA	0.600
%Sulfation CDCA	0.552
%Sulfation DCA	0.479
%Sulfation CA	0.514
%Sulfation on mono-hydroxy	0.610
%Sulfation on dihydroxy	0.635
%Sulfation on tri-hydroxy	0.484
%Sulfation on primary BA	0.514
%Sulfation on secondary BA	0.461
%Sulfation on 12α-OH BA	0.696
%Sulfation on non-12α-OH BA	0.506
% Sulfation on unamidated BA	0.621
%Sulfation on amidated BA	0.621
%Sulfation on G-BA	0.573
%Sulfation on T-BA	0.694
%Unamidated BA	0.679
% G-Amidated BA	0.590
% T-Amidated BA	0.507

% Amidated BA	0.679
%Amidation LCA	0.558
%Amidation UDCA	0.856
%Amidation CDCA	0.721
%Amidation DCA	0.529
%Amidation HDCA	0.497
%Amidation CA	0.636
%Amidation MCA	0.671
%Amidation HCA	0.831
%Unamidation LCA	0.442
%G-Amidation LCA	0.545
%T-Amidation LCA	0.448
%Unamidation UDCA	0.856
%G-Amidation UDCA	0.820
%T-Amidation UDCA	0.536
%Unamidation CDCA	0.279
%G-Amidation CDCA	0.568
%T-Amidation CDCA	0.522
%Unamidation DCA	0.471
%G-Amidation DCA	0.377
%T-Amidation DCA	0.581
%Unamidation HDCA	0.506
%G-Amidation HDCA	0.494
%T-Amidation HDCA	0.500
%Unamidation CA	0.636
%G-Amidation CA	0.730
%T-Amidation CA	0.619
%Unamidation MCA	0.671
%G-Amidation MCA	0.600
%T-Amidation MCA	0.643
%Unamidation HCA	0.831
%G-Amidation HCA	0.688
%T-Amidation HCA	0.533
%Unamidation on primary BA	0.688
%Unamidation on secondary BA	0.747
%Unamidation on 12α-OH	0.575
%Unamidation on non-12α-OH	0.780
%Amidation on monohydroxy	0.558
%Amidation on dihydroxy	0.777
%Amidation on trihydroxy	0.657
%Amidation on primary BA	0.688
%Amidation on secondary BA	0.590
%Amidation on 12α-OH	0.575

%Amidation on non-12α-OH	0.826
%G-amidation on monohydroxy	0.545
%T-amidation on monohydroxy	0.448
%G-amidation on dihydroxy	0.679
%T-amidation on dihydroxy	0.511
%G-amidation on trihydroxy	0.614
%T-amidation on trihydroxy	0.477
%G-amidation on primary BA	0.571
%G-amidation on secondary BA	0.517
%G-amidation on 12α-OH	0.663
%G-amidation on non-12α-OH	0.694
%T-amidation on primary BA	0.499
%T-amidation on secondary BA	0.490
%T-amidation on 12α-OH	0.557
%T-amidation on non-12α-OH	0.497
%Total Mono-hydroxyl BA: Total BA	0.672
%Total Di-hydroxyl BA: Total BA	0.502
%Total Tri-hydroxyl BA: Total BA	0.603
%12α-OH: Total BA	0.794
%non-12α-OH: Total BA	0.794
% Primary BA: Total BA	0.708
%Secondary BA: Total BA	0.708
% Total U-, G-, T-LCA : Total BA	0.672
% Total U-, G-, T-UDCA : Total BA	0.555
% Total U-, G-, T-CDCA : Total BA	0.677
% Total U-, G-, T-DCA : Total BA	0.812
% Total U-, G-, T-HDCA : Total BA	0.742
% Total U-, G-, T-MDCA : Total BA	0.811
% Total U-, G-, T-CA : Total BA	0.554
% Total U-, G-, T-MCA : Total BA	0.540
% Total U-, G-, T-HCA : Total BA	0.579
%LCA: Unamidated BA	0.465
%UDCA: Unamidated BA	0.682
%CDCA: Unamidated BA	0.438
%DCA: Unamidated BA	0.442
%HDCA: Unamidated BA	0.689
%MDCA: Unamidated BA	0.661
%CA: Unamidated BA	0.527
%MCA: Unamidated BA	0.471
%HCA: Unamidated BA	0.818
%G,T-LCA: Amidated BA	0.683
%G,T-UDCA: Amidated BA	0.518
%G,T-CDCA: Amidated BA	0.677

%G,T-DCA: Amidated BA	0.828
%G,T-HDCA: Amidated BA	0.496
%G,T-CA: Amidated BA	0.582
%G,T-MCA: Amidated BA	0.534
%G,T-HCA: Amidated BA	0.454
%G-LCA: G-Amidated BA	0.676
%G-UDCA: G-Amidated BA	0.520
%G-CDCA: G-Amidated BA	0.680
%G-DCA: G-Amidated BA	0.832
%G-HDCA: G-Amidated BA	0.495
%G-CA: G-Amidated BA	0.708
%G-MCA: G-Amidated BA	0.504
%G-HCA: G-Amidated BA	0.450
%T-LCA: T-Amidated BA	0.724
%T-UDCA: T-Amidated BA	0.500
%T-CDCA: T-Amidated BA	0.641
%T-DCA: T-Amidated BA	0.732
%T-HDCA: T-Amidated BA	0.501
%T-CA: T-Amidated BA	0.592
%T-MCA: T-Amidated BA	0.620
%T-HCA: T-Amidated BA	0.582
%Sulfated LCA : Sulfated BA	0.653
%Sulfated UDCA : Sulfated BA	0.471
%Sulfated CDCA : Sulfated BA	0.707
%Sulfated DCA : Sulfated BA	0.809
%Sulfated CA : Sulfated BA	0.573
%Unsulfated LCA : unsulfated BA	0.622
%Unsulfated UDCA : unsulfated BA	0.684
%Unsulfated CDCA : unsulfated BA	0.474
%Unsulfated DCA : unsulfated BA	0.741
%Unsulfated HDCA : unsulfated BA	0.753
%Unsulfated MDCA : unsulfated BA	0.817
%Unsulfated CA : unsulfated BA	0.479
%Unsulfated MCA : unsulfated BA	0.522
%Unsulfated HCA : unsulfated BA	0.632
%Primary unamidated BA:Total BA	0.640
%Secondary unamidated BA:Total BA	0.692
%Primary amidated BA:Total BA	0.717
%Secondary amidated BA:Total BA	0.695
%Primary G-BA:Total BA	0.711
%Secondary G-BA:Total BA	0.684
%Primary T-BA:Total BA	0.577
%SecondaryT-BA:Total BA	0.663

%Unamidated -Unsulfated BA: Total BA	0.595
%Amidated -Unsulfated G-BA : Total BA	0.580
%Amidated -Unsulfated T-BA : Total BA	0.632
%Amidated -Unsulfated BA: Total BA	0.625
% Unamidated: Unsulfated BA	0.660
%Amidated G: Unsulfated BA	0.543
%Amidated T: Unsulfated BA	0.631
% Amidated: Unsulfated BA	0.660
%Unamidated -Sulfated BA: Total BA	0.765
%Amidated -Sulfated G-BA : Total BA	0.516
%Amidated -Sulfated T-BA : Total BA	0.576
%Amidated -Sulfated BA: Total BA	0.522
% Unamidated: BA-Sulfates	0.750
% Amidated G: BA-Sulfates	0.670
%Amidated T: BA-sulfates	0.565
% Amidated: BA-Sulfates	0.750
%Unsulfated LCA: Total BA	0.620
%Unsulfated UDCA: Total BA	0.630
%Unsulfated CDCA: Total BA	0.526
%Unsulfated DCA: Total BA	0.673
%Unsulfated HDCA: Total BA	0.742
%Unsulfated MDCA: Total BA	0.811
%Unsulfated CA: Total BA	0.537
%Unsulfated MCA: Total BA	0.540
%Unsulfated HCA: Total BA	0.421
%Unsulfated Mono-hydroxyl BA: Total BA	0.620
%Unsulfated Di-hydroxyl BA: Total BA	0.654
%Unsulfated Tri-hydroxyl BA: Total BA	0.574
% Unsulfated primary: Total BA	0.581
% Unsulfated Secondary: Total BA	0.431
% Unsulfated 12α-OH: Total BA	0.553
% Unsulfated non 12α-OH: Total BA	0.562
%Sulfated LCA: Total BA	0.671
%Sulfated UDCA: Total BA	0.447
%Sulfated CDCA: Total BA	0.676
%Sulfated DCA: Total BA	0.813
%Sulfated CA: Total BA	0.550
%Sulfated Mono-hydroxyl BA: Total BA	0.671
%Sulfated Di-hydroxyl BA: Total BA	0.506
%Sulfated Tri-hydroxyl BA: Total BA	0.550
% Sulfated primary: Total BA	0.680
% Sulfated Secondary: Total BA	0.717
% Sulfated 12α-OH: Total BA	0.799

% Sulfated non 12α-OH: Total BA	0.711
% Unsulfated Nor-DCA	0.763
% Unsulfated 12-oxo-CDCA	0.409
% Unsulfated 3-dehydroCA	0.498
% Unsulfated 6-oxo-LCA	0.600
% Unsulfated 7-oxo-LCA	0.542
% Unsulfated 12-oxo-LCA	0.478
% Unsulfated isoLCA	0.517
% Unsulfated isoDCA	0.684
Primary BA/Secondary BA	0.708
Total amidated/total unamidated	0.678
12α-OH/non12α-OH	0.794
CA/CDCA	0.565
CA+DCA/CDCA+LCA	0.790
Total BA for HI	0.770
HI	0.664

* AUC from ROC analysis of pooled patients vs. controls.

2.7. References

1. Hofmann, A.F., *The continuing importance of bile acids in liver and intestinal disease.* Arch Intern Med, 1999. **159**(22): p. 2647-58.
2. Monte, M.J., et al., *Bile acids: chemistry, physiology, and pathophysiology.* World J Gastroenterol, 2009. **15**(7): p. 804-16.
3. Khurana, S., J.P. Raufman, and T.L. Pallone, *Bile acids regulate cardiovascular function.* Clin Transl Sci, 2011. **4**(3): p. 210-8.
4. Thomas, C., et al., *Targeting bile-acid signalling for metabolic diseases.* Nat Rev Drug Discov, 2008. **7**(8): p. 678-93.
5. Maillette de Buy Wenniger, L. and U. Beuers, *Bile salts and cholestasis.* Dig Liver Dis, 2010. **42**(6): p. 409-18.
6. Pauli-Magnus, C. and P.J. Meier, *Hepatocellular transporters and cholestasis.* J Clin Gastroenterol, 2005. **39**(4 Suppl 2): p. S103-10.
7. Palmer, R.H., *Bile acids, liver injury, and liver disease.* Arch Intern Med, 1972. **130**(4): p. 606-17.
8. Alnouti, Y., *Bile Acid sulfation: a pathway of bile acid elimination and detoxification.* Toxicol Sci, 2009. **108**(2): p. 225-46.
9. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases II. Signature profiles in patients.* Toxicol Sci, 2015. **143**(2): p. 308-18.
10. Makino, I., et al., *Sulfated and nonsulfated bile acids in urine, serum, and bile of patients with hepatobiliary diseases.* Gastroenterology, 1975. **68**(3): p. 545-53.
11. Summerfield, J.A., et al., *Evidence for renal control of urinary excretion of bile acids and bile acid sulphates in the cholestatic syndrome.* Clin Sci Mol Med, 1977. **52**(1): p. 51-65.
12. Takikawa, H., T. Beppu, and Y. Seyama, *Urinary concentrations of bile acid glucuronides and sulfates in hepatobiliary diseases.* Gastroenterol Jpn, 1984. **19**(2): p. 104-9.
13. van Berge Henegouwen, G.P., et al., *Sulphated and unsulphated bile acids in serum, bile, and urine of patients with cholestasis.* Gut, 1976. **17**(11): p. 861-9.
14. LaRusso, N.F., et al., *Primary sclerosing cholangitis: summary of a workshop.* Hepatology, 2006. **44**(3): p. 746-64.
15. Dueland, S., et al., *Regulation of cholesterol and bile acid homoeostasis in bile-obstructed rats.* Biochem J, 1991. **280 (Pt 2)**: p. 373-7.
16. Kawai, H., et al., *Efficacy of urine bile acid as a non-invasive indicator of liver damage in rats.* J Toxicol Sci, 2009. **34**(1): p. 27-38.
17. Buis, C.I., et al., *Altered bile composition after liver transplantation is associated with the development of nonanastomotic biliary strictures.* J Hepatol, 2009. **50**(1): p. 69-79.
18. Geuken, E., et al., *Rapid increase of bile salt secretion is associated with bile duct injury after human liver transplantation.* J Hepatol, 2004. **41**(6): p. 1017-25.
19. Ramaiah, S.K., *A toxicologist guide to the diagnostic interpretation of hepatic biochemical parameters.* Food Chem Toxicol, 2007. **45**(9): p. 1551-7.
20. Ozer, J., et al., *The current state of serum biomarkers of hepatotoxicity.* Toxicology, 2008. **245**(3): p. 194-205.
21. Nusrat, S., et al., *Cirrhosis and its complications: evidence based treatment.* World J Gastroenterol, 2014. **20**(18): p. 5442-60.
22. Huang, W.M., et al., *Intrahepatic cholestasis of pregnancy: detection with urinary bile acid assays.* J Perinat Med, 2007. **35**(6): p. 486-91.
23. Muraji, T., et al., *Urinary sulfated bile acid concentrations in infants with biliary atresia and breast-feeding jaundice.* Pediatr Int, 2003. **45**(3): p. 281-3.

24. Sinakos, E. and K.D. Lindor, *Bile acid profiles in intrahepatic cholestasis of pregnancy: is this the solution to the enigma of intrahepatic cholestasis of pregnancy?* Am J Gastroenterol, 2010. **105**(3): p. 596-8.
25. Ambros-Rudolph, C.M., et al., *The importance of serum bile acid level analysis and treatment with ursodeoxycholic acid in intrahepatic cholestasis of pregnancy: a case series from central Europe.* Arch Dermatol, 2007. **143**(6): p. 757-62.
26. Heikkinen, J., *Serum bile acids in the early diagnosis of intrahepatic cholestasis of pregnancy.* Obstet Gynecol, 1983. **61**(5): p. 581-87.
27. Shinohara, T., et al., *Efficacy of urinary sulfated bile acids for diagnosis of bacterial cholangitis in biliary atresia.* Pediatr Surg Int, 2005. **21**(9): p. 701-4.
28. Trottier, J., et al., *Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing cholangitis: a pilot study.* Dig Liver Dis, 2012. **44**(4): p. 303-10.
29. Simko, V., S. Michael, and R.E. Kelley, *Predictive value of random sample urine bile acids corrected by creatinine in liver disease.* Hepatology, 1987. **7**(1): p. 115-21.
30. Stiehl, A., et al., *Biliary and urinary excretion of sulfated, glucuronidated and tetrahydroxylated bile acids in cirrhotic patients.* Hepatology, 1985. **5**(3): p. 492-5.
31. Dasarathy, S., et al., *Elevated hepatic fatty acid oxidation, high plasma fibroblast growth factor 21, and fasting bile acids in nonalcoholic steatohepatitis.* Eur J Gastroenterol Hepatol, 2011. **23**(5): p. 382-8.
32. Simko, V. and S. Michael, *Urinary bile acids in population screening for inapparent liver disease.* Hepatogastroenterology, 1998. **45**(23): p. 1706-14.
33. Takikawa, H., et al., *Glucuronidated and sulfated bile acids in serum of patients with acute hepatitis.* Dig Dis Sci, 1986. **31**(5): p. 487-91.
34. Bezerra, J.A., et al., *Biliary Atresia: Clinical and Research Challenges for the Twenty-First Century.* Hepatology, 2018. **68**(3): p. 1163-1173.
35. Meng, L.J., et al., *Profiles of bile acids and progesterone metabolites in the urine and serum of women with intrahepatic cholestasis of pregnancy.* J Hepatol, 1997. **27**(2): p. 346-57.
36. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases I. Stability of the baseline profile in healthy subjects.* Toxicol Sci, 2015. **143**(2): p. 296-307.
37. Thakare, R., et al., *Species differences in bile acids I. Plasma and urine bile acid composition.* J Appl Toxicol, 2018. **38**(10): p. 1323-1335.
38. Bathena, S.P., et al., *The profile of bile acids and their sulfate metabolites in human urine and serum.* J Chromatogr B Analyt Technol Biomed Life Sci, 2013. **942-943**: p. 53-62.
39. Huang, J., et al., *Simultaneous characterization of bile acids and their sulfate metabolites in mouse liver, plasma, bile, and urine using LC-MS/MS.* J Pharm Biomed Anal, 2011. **55**(5): p. 1111-9.
40. Heuman', D.M., *Quantitative estimation of the hydrophilichydrophobic balance of mixed bile salt solutions.* Journal of Lipid Research, 1989. **30**.
41. Pagano, M. and K. Gauvreau, *Principles of Biostatistics*. 2nd ed. 2000: Brooks/Cole.
42. Malinchoc, M., et al., *A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts.* Hepatology, 2000. **31**(4): p. 864-71.
43. Kamath, P.S., et al., *A model to predict survival in patients with end-stage liver disease.* Hepatology, 2001. **33**(2): p. 464-70.

44. Murray, K.F., R.L. Carithers, Jr., and Aasld, *AASLD practice guidelines: Evaluation of the patient for liver transplantation.* Hepatology, 2005. **41**(6): p. 1407-32.
45. Zipprich, A., et al., *Incorporating indocyanin green clearance into the Model for End Stage Liver Disease (MELD-ICG) improves prognostic accuracy in intermediate to advanced cirrhosis.* Gut, 2010. **59**(7): p. 963-8.
46. Farnsworth, N., et al., *Child-Turcotte-Pugh versus MELD score as a predictor of outcome after elective and emergent surgery in cirrhotic patients.* Am J Surg, 2004. **188**(5): p. 580-3.
47. Zipprich, A., et al., *Prognostic indicators of survival in patients with compensated and decompensated cirrhosis.* Liver Int, 2012. **32**(9): p. 1407-14.
48. Trottier, J., et al., *Profiling circulating and urinary bile acids in patients with biliary obstruction before and after biliary stenting.* PLoS One, 2011. **6**(7): p. e22094.
49. Modica, S., et al., *Selective activation of nuclear bile acid receptor FXR in the intestine protects mice against cholestasis.* Gastroenterology, 2012. **142**(2): p. 355-65 e1-4.
50. Luo, L., et al., *Assessment of serum bile acid profiles as biomarkers of liver injury and liver disease in humans.* PLoS One, 2018. **13**(3): p. e0193824.
51. Mallory, A., et al., *Patterns of bile acids and microflora in the human small intestine. I. Bile acids.* Gastroenterology, 1973. **64**(1): p. 26-33.
52. Northfield, T.C. and I. McColl, *Postprandial concentrations of free and conjugated bile acids down the length of the normal human small intestine.* Gut, 1973. **14**(7): p. 513-8.
53. Perreault, M., et al., *Role of glucuronidation for hepatic detoxification and urinary elimination of toxic bile acids during biliary obstruction.* PLoS One, 2013. **8**(11): p. e80994.
54. Heuman, D.M., et al., *Conjugates of ursodeoxycholate protect against cytotoxicity of more hydrophobic bile salts: in vitro studies in rat hepatocytes and human erythrocytes.* Hepatology, 1991. **14**(5): p. 920-6.
55. Rolo, A.P., et al., *Bile acids affect liver mitochondrial bioenergetics: possible relevance for cholestasis therapy.* Toxicol Sci, 2000. **57**(1): p. 177-85.
56. Li, Y., et al., *Targeted metabolomics of sulfated bile acids in urine for the diagnosis and grading of intrahepatic cholestasis of pregnancy.* Genes Dis, 2018. **5**(4): p. 358-366.
57. Hofmann, A.F. and L.R. Hagey, *Bile acids: chemistry, pathochemistry, biology, pathobiology, and therapeutics.* Cell Mol Life Sci, 2008. **65**(16): p. 2461-83.
58. Rust, C., et al., *The bile acid taurochenodeoxycholate activates a phosphatidylinositol 3-kinase-dependent survival signaling cascade.* J Biol Chem, 2000. **275**(26): p. 20210-6.
59. Rust, C., et al., *Phosphatidylinositol 3-kinase-dependent signaling modulates taurochenodeoxycholic acid-induced liver injury and cholestasis in perfused rat livers.* Am J Physiol Gastrointest Liver Physiol, 2005. **289**(1): p. G88-94.
60. Gerloff, T., et al., *The sister of P-glycoprotein represents the canalicular bile salt export pump of mammalian liver.* J Biol Chem, 1998. **273**(16): p. 10046-50.
61. Hayashi, H., et al., *Transport by vesicles of glycine- and taurine-conjugated bile salts and taurolithocholate 3-sulfate: a comparison of human BSEP with rat Bsep.* Biochim Biophys Acta, 2005. **1738**(1-3): p. 54-62.
62. Noe, J., B. Stieger, and P.J. Meier, *Functional expression of the canalicular bile salt export pump of human liver.* Gastroenterology, 2002. **123**(5): p. 1659-66.

63. Jansen, P.L., et al., *Hepatocanalicular bile salt export pump deficiency in patients with progressive familial intrahepatic cholestasis.* Gastroenterology, 1999. **117**(6): p. 1370-9.
64. van der Woerd, W.L., R.H. Houwen, and S.F. van de Graaf, *Current and future therapies for inherited cholestatic liver diseases.* World J Gastroenterol, 2017. **23**(5): p. 763-775.
65. Stieger, B., *Role of the bile salt export pump, BSEP, in acquired forms of cholestasis.* Drug Metab Rev, 2010. **42**(3): p. 437-45.
66. Dawson, P.A. and S.J. Karpen, *Intestinal transport and metabolism of bile acids.* J Lipid Res, 2015. **56**(6): p. 1085-99.
67. Ridlon, J.M., et al., *Bile acids and the gut microbiome.* Curr Opin Gastroenterol, 2014. **30**(3): p. 332-8.
68. Kakiyama, G., et al., *Modulation of the fecal bile acid profile by gut microbiota in cirrhosis.* J Hepatol, 2013. **58**(5): p. 949-55.
69. Barrett, K.G., et al., *Upregulation of UGT2B4 Expression by 3'-Phosphoadenosine-5'-Phosphosulfate Synthase Knockdown: Implications for Coordinated Control of Bile Acid Conjugation.* Drug Metab Dispos, 2015. **43**(7): p. 1061-70.
70. Meier, P.J. and B. Stieger, *Bile salt transporters.* Annu Rev Physiol, 2002. **64**: p. 635-61.
71. Hirohashi, T., et al., *ATP-dependent transport of bile salts by rat multidrug resistance-associated protein 3 (Mrp3).* J Biol Chem, 2000. **275**(4): p. 2905-10.
72. Zelcer, N., et al., *Steroid and bile acid conjugates are substrates of human multidrug-resistance protein (MRP) 4 (ATP-binding cassette C4).* Biochem J, 2003. **371**(Pt 2): p. 361-7.
73. Stieger, B., et al., *Drug- and estrogen-induced cholestasis through inhibition of the hepatocellular bile salt export pump (Bsep) of rat liver.* Gastroenterology, 2000. **118**(2): p. 422-30.
74. Kubitz, R., et al., *Regulation of the multidrug resistance protein 2 in the rat liver by lipopolysaccharide and dexamethasone.* Gastroenterology, 1999. **116**(2): p. 401-10.
75. Mottino, A.D., et al., *Altered localization and activity of canalicular Mrp2 in estradiol-17beta-D-glucuronide-induced cholestasis.* Hepatology, 2002. **35**(6): p. 1409-19.
76. Kojima, H., et al., *Changes in the expression and localization of hepatocellular transporters and radixin in primary biliary cirrhosis.* J Hepatol, 2003. **39**(5): p. 693-702.
77. Kojima, H., et al., *Disturbed colocalization of multidrug resistance protein 2 and radixin in human cholestatic liver diseases.* J Gastroenterol Hepatol, 2008. **23**(7 Pt 2): p. e120-8.
78. Pandak, W.M., et al., *Expression of sterol 12alpha-hydroxylase alters bile acid pool composition in primary rat hepatocytes and in vivo.* Gastroenterology, 2001. **120**(7): p. 1801-9.
79. Zhang, M. and J.Y. Chiang, *Transcriptional regulation of the human sterol 12alpha-hydroxylase gene (CYP8B1): roles of heaptocyte nuclear factor 4alpha in mediating bile acid repression.* J Biol Chem, 2001. **276**(45): p. 41690-9.
80. Mork, L.M., et al., *Addition of Dexamethasone Alters the Bile Acid Composition by Inducing CYP8B1 in Primary Cultures of Human Hepatocytes.* J Clin Exp Hepatol, 2016. **6**(2): p. 87-93.
81. Li, T. and U. Apte, *Bile Acid Metabolism and Signaling in Cholestasis, Inflammation, and Cancer.* Adv Pharmacol, 2015. **74**: p. 263-302.

82. Chiang, J.Y., *Negative feedback regulation of bile acid metabolism: impact on liver metabolism and diseases.* Hepatology, 2015. **62**(4): p. 1315-7.

83. Mita, S., et al., *Inhibition of bile acid transport across Na+/taurocholate cotransporting polypeptide (SLC10A1) and bile salt export pump (ABCB 11)-coexpressing LLC-PK1 cells by cholestasis-inducing drugs.* Drug Metab Dispos, 2006. **34**(9): p. 1575-81.
84. Adachi, T., et al., *The involvement of endoplasmic reticulum stress in bile acid-induced hepatocellular injury.* J Clin Biochem Nutr, 2014. **54**(2): p. 129-35.
85. Attili, A.F., et al., *Bile acid-induced liver toxicity: relation to the hydrophobic-hydrophilic balance of bile acids.* Med Hypotheses, 1986. **19**(1): p. 57-69.
86. Watanabe, S. and K. Tsuneyama, *Cattle bile but not bear bile or pig bile induces lipid profile changes and fatty liver injury in mice: mediation by cholic acid.* J Toxicol Sci, 2012. **37**(1): p. 105-21.
87. Fan, J., S. Upadhye, and A. Worster, *Understanding receiver operating characteristic (ROC) curves.* CJEM, 2006. **8**(1): p. 19-20.
88. Pines, J.M., et al., *Evidence-based emergency care: Diagnostic testing and clinical decision rules*. Vol. 83. 2013: John Wiley & Sons.

CHAPTER 3

THE BILE ACID SCORE (BAS): A SURVIVAL MODEL FOR PATIENTS WITH LIVER DISEASES

3.1. Introduction

Cholestatic liver diseases are hepatobiliary diseases associated with a lowering in bile flow due to impairment in bile production or obstruction of bile flow into bile duct [1]. Chronic liver diseases led to over 41,000 deaths in the United States in 2017, making it the 11th leading cause of mortality [2]. The etiology and pathogenesis of most cholestatic diseases are poorly understood and a pharmacological cure for these diseases is not yet available. Most cholestatic diseases progress toward end stage liver failure, which likely requires liver transplantation. Even after liver transplantation, post-surgery complications are common [3], which may require liver re-transplantation.

Biomarkers that are currently used in the clinic for the diagnosis and prognosis of liver diseases are primarily serum liver enzymes such as AST and ALT as well as bilirubin. However, these markers have numerous shortfalls including the lack of specificity to liver or bile duct injuries, where their levels can be elevated in non-hepatobiliary diseases such as hyperthyroidism, adrenal, heart, and muscle disorders. Also, they require severe cell injury before their levels increase in the blood [4, 5]. Multifactorial models with multiple parameters based on these biomarkers are also frequently used and offer advantages compared to the use of their individual biomarker components such as the Child-Turcotte-Pugh (CTP) score. The CTP score, originally developed to predict portosystemic shunt surgery outcomes in cirrhotic patients, formed the basis on which liver disease severity was assessed. However, the usefulness of CTP was limited by a number of inherent problems [6].

More recently, the mayo model for end-stage liver disease (MELD) was developed to predict three-month mortality of patients with end-stage liver disease [7, 8]. MELD is calculated based on serum bilirubin, creatinine, and international normalized ratio (INR), which are related to both liver and renal functions. The MELD score is better than other

prognostic models in patients with end-stage liver disease, such as CTP score. MELD is also currently used in many countries to classify liver diseases' patients awaiting transplantation to identify patients with the highest priority for liver transplant (LT) [9]. Since its implementation, MELD led to a decrease in the number of individuals waiting for liver transplant and decreased mortality on the waiting list without affecting post-transplant survival [9, 10]. Although mainly adopted for use in patients waiting for liver transplant, the MELD score has additionally proved to be an effective predictor of outcome in other circumstances, for example, patients with fulminant hepatic failure or alcoholic hepatitis and patients with cirrhosis going for surgery [10]. However, despite its widespread application, MELD has some limitations. MELD is based on three objective laboratory variables, that are not necessarily liver specific. For example, patients may have an elevated serum creatinine from an underlying kidney disease that unrelated to hepatorenal syndrome. In addition, serum bilirubin can be elevated in cases of hemolysis or sepsis. An elevated INR can also be secondary to warfarin use. Any of these cases can increase the MELD score and overestimate the severity of liver disease [10, 11]. Furthermore, several studies have shown that patients with cholestatic liver diseases may still have high mortality rates despite having low MELD scores [11, 12].

In this report, we have investigated and for the 1st time, the use of BA to build a survival model to predict the prognosis of hepatobiliary diseases. Despite their vital physiological functions, BA are also cytotoxic and can cause hepatic and biliary toxicities. The impediment in bile flow associated with cholestatic liver diseases cause accumulation of BA in the liver and blood. Numerous clinical and preclinical studies have shown up to a 100-fold increase in BA concentrations in urine with various hepatobiliary diseases [13-17].

There is ample evidence from animal and human studies to indicate that BA accumulation in the liver, systemic blood, and extrahepatic tissues can worsen the liver condition that lead to their accumulation, which may contribute to the unfavorable liver disease prognosis. However, the potential use of BA as a marker for liver diseases have never translated into the clinic due to major limitations including the major differences of the physiologic and pathologic effects of the various individual BA as well as the extremely high inter- and intra-individual variability of BA concentrations.

To that regard, we have developed the concept of “BA Indices”, which are ratios calculated from the absolute concentration of individual BA and their metabolites. These ratios quantify in detail, the composition, hydrophilicity, metabolism, formation of secondary BAs, and toxicity of the BA profile [1, 18, 19]. We have shown that BA indices offered numerous advantages over absolute total and individual BA concentrations including low inter- and intra-individual variability and were resistant to covariate influences such as age, gender, BMI, food consumption, and moderate alcohol consumption. Furthermore, we have demonstrated that BA indices outperformed serum liver enzymes such as AST and ALT as biomarkers for the diagnosis of cholestatic liver diseases [1].

In this study, we have extended the application of BA indices to predict the prognosis of liver diseases. This study aims to develop survival models based on BA indices to predict the prognosis of hepatobiliary diseases. The performance of our BA models were compared to non-BA and MELD models in predicting the occurrence of death only and death and/or liver transplant using various statistical approaches.

3.2. Materials and methods

3.2.1. Study participants

New and existing patients of the UNMC hepatology clinic, who were diagnosed with one or multi-hepatobiliary conditions due to chronic hepatitis C (n=64) , hepatitis B (n=15), alcoholic liver disease (n=105), primary biliary cholangitis (PBC) (n=12), primary sclerosing cholangitis (PSC) (n=15), autoimmune hepatitis (n=26), alpha-1-antitrypsin deficiency (n=5), nonalcoholic fatty liver disease/nonalcoholic steatohepatitis (NASH) (n=52), carcinoma (n=25), cryptogenic cirrhosis (n=11), polycystic liver disease (n=5), elevated liver function test (LFT) (n=19), and unknown etiology (n=5), were enrolled in this study. **Table 3.1.** shows a summary of our patient population characteristics. A total of 257 patients (121 female and 136 male) between the ages of 19 and 83 years, who were treated for cholestatic liver diseases in the University of Nebraska Medical Center (UNMC) (Omaha, NE, USA), over the period from November of 2011 - December of 2018, were recruited into the study. All participants were followed up for up to 7 years by collecting urine samples for BA analysis and monitoring non-BA parameters and adverse events including liver transplant, and death from their medical records.

The study was approved by the institutional review board (IRB) at UNMC and written informed consents were provided for all participating subjects. Thirty milliliters of urine samples were collected from patients on their first visit to the hepatology clinic. All urine samples were stored in -80 °C until analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS).

3.2.2. Non-BA parameters

The performance of potential biomarkers from the urinary BA profile was also compared with and existing markers of liver function including alanine transaminase (ALT), aspartate transaminase (AST), serum creatinine, albumin, bilirubin, protime,

international normalized ratio (INR), AST/ ALT ratio, , and AST/ platelet ratio index (APRI). These markers were monitored using the patients' medical records.

3.2.3. Bile acid quantification by liquid chromatography–tandem mass spectrometry

BA concentrations were quantified by LC-MS/MS, as described previously with some modifications [19-21]. Briefly, a Waters ACQUITY ultra performance liquid chromatography (UPLC) system (Waters, Milford, MA, USA) coupled to an Applied Biosystem 4000 Q TRAP® quadrupole linear ion trap hybrid MS with an electrospray ionization (ESI) source (Applied Biosystems, MDS Sciex, Foster City, CA, USA) was used. The following MS source settings were used: temperature, 500°C; ion spray voltage, −4000 V; collision gas pressure, high; curtain gas, 20; gas-1, 35; gas-2 35 (arbitrary units); Q1/Q3 resolution, unit; and interface heater, on. Mobile phase consisted of 7.5 mM ammonium bicarbonate, has been adjusted to pH 9.0 by using ammonium hydroxide (mobile phase A) and 30% acetonitrile in methanol (mobile phase B) at a total flow rate of 0.2 ml/min. The gradient profile was held at 52.5% mobile phase B for 12.75 minutes, increased linearly to 68% in 0.25 minutes, held at 68% for 8.75 minutes, increased linearly to 90% in 0.25 minutes, held at 90% for one minute and finally brought back to 52.5% in 0.25 minutes and then followed by 4.75 minutes re-equilibration (total run time of 28 minutes per sample).

3.2.4. Calculation of BA indices

BA profile in urine was characterized using BA "indices", which describe the composition, hydrophobicity, toxicity, and metabolism of total and individual BA [1, 18, 20]. Briefly, the composition of individual BA was calculated as the ratio of the concentration of individual BA in all of their forms (sulfated, unsulfated, amidated, and unamidated) to the total concentration of BA. The percentage of sulfation of individual BA was calculated

as the ratio of the concentration of sulfated BA to the total concentration of individual BA in all of their forms (amidated, unamidated, sulfated, and unsulfated). The percentage of amidation of individual BA was calculated as the ratio of the concentration of amidated BA to the total concentration of individual BA in all of their forms (unsulfated, sulfated, unamidated and amidated). In addition, percentages of amidation were divided into the percentages of BA existing as glycine-(G) vs. taurine-(T) amidates. The percentages of mono-OH BA (LCA), di-OH BA (HDCA, UDCA, MDCA, DCA, and CDCA) and tri-OH BA (HCA, MCA, and CA) were calculated as the ratio of the concentration of the sum of the respective BA in all their forms to the total concentration of BA. The ratio of primary to secondary BA was calculated as the ratio of the sum of the concentrations of the primary BA: CA, MCA, HCA, and CDCA to the sum of the concentrations of the secondary BA: UDCA, DCA, MDCA, HDCA, and LCA in all their forms. Similarly, the ratio of 12α-OH to non-12α-OH was calculated as the ratio of the sum of the concentrations of CA and DCA to the sum of the concentrations of HDCA, CDCA, LCA, MCA, UDCA, and HCA in all their forms. The hydrophobicity index (HI) of the BA pool was calculated according to the Heuman index, based on the relative contributions of the individual BA to the total BA pool and their His [22].

3.2.5. Preparation of standard solutions and calibration curves

For the preparation of standard solutions and calibration curves, blank matrices were obtained by charcoal stripping as described previously [1, 18-21]. Eleven-point calibration curve was prepared by spiking 10 µL of the appropriate standard solutions and 10 µL of the IS stock ($^{2}H_{4}$-G-CDCA) into 100 µL of the stripped urine matrix. The final concentration of IS was 500 ng/ml and the dynamic range of the standard curves for the various BA analytes was 1-1000 ng/ml.

3.2.6. Sample preparation

Urine sample extraction was described previously [1, 18-21]. Briefly, 100 µl of urine samples were spiked with 10 µl of internal standard (IS), vortexed and loaded on to Supelclean™ LC-18 SPE cartridges, which were pre-conditioned with 4 ml MeOH, and 4 ml H_2O. Samples were then eluted with 4 mL MeOH. Eluates were evaporated under vacuum at room temperature and reconstituted in a 100 µL of 50 % MeOH solution. Ten microliters of reconstituted samples were injected for LC-MS/MS analysis.

3.2.7. MELD Score

MELD was originally developed to predict three-month mortality of patients with end-stage liver disease. MELD is currently used in many countries to classify liver diseases' patients awaiting transplantation to identify patients with the highest priority for liver transplant (LT) [9]. MELD is also used as a predictor of outcomes in other situations, such as patients with cirrhosis going for surgery, and patients with fulminant hepatic failure or alcoholic hepatitis [10]. The MELD score for each patient was calculated according to the original MELD model [7, 8, 23] :

$$\text{MELD score} = 9.57 \times \log_e \text{creatinine}\frac{\text{mg}}{\text{dL}} + 3.78 \times \log_e \text{bilirubin}\frac{\text{mg}}{\text{dL}} + 11.2 \times \log_e \text{INR} + 6.43$$

The following constraints were made: all variables < 1 were bound to 1 to avoid negative scores, the maximum value used for creatinine was 4 mg/dL, all MELD scores exceeding 40 were bound to 40. (6.43) was a constant for liver disease etiology. The MELD score was rounded to the nearest integer and ranged from 6 to 40 [23].

In addition we used the modified MELD model, which takes into account the serum sodium concentration (Na) [23]:

$$\text{MELD Na} = \text{MELD score} - \text{Na} - [\text{MELD score} \times 0.025 \times (140 - \text{Na})] + 140$$

Serum sodium concentration (Na) was bound between 125 and 140 mmol/L. Like the MELD score, the MELD-Na score was rounded to the nearest integer. MELD-Na score provides better calibration and discrimination of the risk of death among candidates for liver transplantation [23].

3.2.8. Statistical analysis

All statistical analysis was performed using the Statistical Product and Service Solutions (SPSS) software, version 25 (IBM corporation, Armonk, NY, USA) and R software, version 3.6.3 (R Foundation for statistical Computing). A p-value of 0.05 was considered significant for all the statistical tests described below.

3.2.8.1. Survival Model Development

Cox proportional hazards regression was used to develop survival models to predict the prognosis of hepatobiliary diseases in terms of progressing specifically into the end points/adverse events of death and/or liver transplant. Models were constructed to predict (i) death only, and (ii) death or liver transplant. We did not develop models to predict liver transplant (LT) only, because patients who died in this study were censored at the time of death even though they could have been candidates for LT later. Therefore, these patients might have needed LT, but they died before then.

For the “death only” models, the only endpoint/adverse event recorded was death at 3 and 5 years. We only had 7 and 17 deaths occurring within earlier time points including 1 and 2 years, respectively, which was not enough to develop survival models. Patients who underwent liver transplant (LT) were censored with the date of transplantation. Patients still alive at the end of each period (3 and 5 years) were considered as censored at that time. The term ‘‘censored’’ indicates that the patient was alive at that date and that

was the end of the follow-up [24]. Patients dropped off, not due to the occurrence of adverse event, i.e. death, before the end of the follow-up period, were censored at the last day they were seen in the clinic.

For the “death or liver transplant” models, we followed the same approach as the “death only” models, with the exception that the endpoint was the occurrence of the adverse events of either death or liver transplant (LT). Patients whom did not have either of the adverse events at the end of each period (3 and 5 years) were censored at that time. Patients dropped off, not due to the occurrence of adverse event, i.e. death or liver transplant, before the end of the follow-up period, were censored at the last day they were seen in the clinic.

Individual BA and non-BA variables were analyzed as possible predictors of survival in a univariate Cox regression analysis. Values of these variables included in the statistical analysis were obtained at the time of patients’ first visits. Significant variables ($P < 0.05$), which were identified from the univariate analysis were included in the multivariate analysis. To build the multivariate model a backward elimination regression method was used to retain the most significant variables with retention criteria of $P < 0.05$.

3.2.8.2. Model performance, Goodness of fit and Validation

Goodness of fit was performed by testing proportional hazards (PH) assumption for each covariate included in the final Cox model and for the global model as a whole using a statistical test and a graphical diagnostic based on Schoenfeld residuals. The Schoenfeld residuals are independent of time. Therefore, a plot that shows a non-random pattern against time is evidence of violation of the PH assumption.

We used the bootstrapping for model validation. The bootstrapping is a resampling technique used to estimate statistics on a population by sampling a dataset with

replacement. Random samples were taken with replacement from our data set, one at a time, to create a series of 1000 new data sets and statistics are calculated by comparing these data sets [25]. The bootstrapping statistics include p-value, bias, standard error, and 95% confidence interval of a bootstrap estimator [26]. The difference between the estimate computed using the original sample and the mean of the bootstrap estimates is a bootstrap estimate of bias. The standard error of an estimator is its standard deviation. It tells us how far your sample estimate deviates from the actual parameter [27, 28].

Receiver operating characteristic curve (ROC) analyses were performed on the scores from the various multivariate Cox models to determine their cut-off values in differentiating patients with vs. without the adverse event. The cut-off values with optimum specificity and sensitivity were selected and the areas under the ROC curve (AUC) values were calculated.

3.2.8.3. Survival Prediction

The average survival probability (S_0 (t)) for a patient with an average score were calculated for different time points. To obtain the probability of survival for t years (S (t)), first the score e.g. (BAS) is calculated, and finally S (t) is calculated using this equation:

$$\text{Survival Probability for t years: } S(t) = S_{\circ}(t)^{\exp(BAS - BAS_{\circ})}$$

Where, BAS_0 is the average score from all patients in this study.

Kaplan-Meier (KM) plots were used to display survival curves. We have divided patients into two categories of high vs. low risk and compared their survival with the Log-rank test and Breslow test [29]. We have tried two cut-offs to define high vs. low risk, the median model score of the population as well as the cut-off values of the model score with optimum specificity and sensitivity based on ROC analysis.

3.2.8.4. Models comparison

We have used multivariate cox regression analyses to build various models for the prediction of the adverse events of (i) death only and (ii) death and/or liver transplant (LT). For both approaches, we developed models that can be divided into the following six categories: (i) BA variables only, (ii) non-BA variables only, (iii) mixed BA and non-BA variables, (iv) original Model for end-stage liver disease (MELD), (v) MELD variables with coefficients from our data set, (vi) original MELD modified with BA and/or non-BA variables.

The performance of the different models in predicting the occurrence of adverse events of death and/or liver transplant within 3- and 5-year periods were compared between the different models using the statistic outcomes from the Bootstrapping, Schoenfeld residuals, areas under the ROC curve (AUC), and Kaplan-Meier analyses.

3.3. Results

3.3.1. Patient population characteristics

Table 3.1 shows a summary of the characteristics of the patient population in our study. The demographic variables were (age, BMI, gender, and race). Subjects were divided into five race groups (White, Black, Asian, Hispanic, and others). During the 7-year follow-up period of 257 patients with cholestatic liver diseases, 27 patients (10.5%) died and 25 patients (9.7%) underwent liver transplantation.

We were interested in predicting the occurrence of adverse events of death and/or liver transplant within 3- and 5-year periods. During a 3-year follow-up period, 21 patients (8.2%) died and 19 patients (7.4%) underwent liver transplantation. While during a 5-year follow-up period, 25 patients (9.7%) died and 21 patients (8.2%) underwent liver transplantation.

3.3.2. Univariate Cox regression analysis

3.3.2.i. Death prediction

Table 3.2 shows the results of univariate Cox regression analyses for death prediction by BA Indices. Cox regression detects the risk of death associated with changes in BA indices. Positive regression coefficients imply that the risk of death increases with increasing the values of BA indices, while negative coefficients imply the risk of death increases with a decrease in the values of BA indices. We found correlation between the risk of death and many BA indices ($P < 0.05$).

The hazard ratio (HR) from Cox regressions analysis quantifies the magnitude of the risk of death per unit change in BA indices. Because BA concentrations and indices have different scales and units, we performed the same calculation per 10% and 20% of the mean value of each variable instead of per absolute unit. For example, for a 20%

increase in the % CDCA, the risk of death increases 1.26-fold (HR: 1.26; $P < 0.05$). Similarly, increasing levels of % CA, % HCA, % T-amidation, % Tri-OH, % non-12α-OH, primary/ secondary, and % primary BA significantly increased the risk of death, whereas decreasing levels of % LCA, % DCA, % G-amidation, % Mono-OH, 12α-OH/ non-12α-OH, % 12α-OH, % secondary BA significantly increased the risk of death.

We performed the same univariate cox regression analysis for demographics and non-BA parameters as well (**Table 3.3**). Notably, the risk of death was significantly higher in males than females from this univariate analysis. Increasing levels of INR, protime, bilirubin, AST/ALT, APRI, and MELD also significantly increased the risk of death, whereas decreasing levels of albumin significantly increased the risk of death.

3.3.2.ii. Death and/or Liver Transplant (LT) prediction

Table 3.4 shows the results of univariate Cox regression analyses for death and/or liver transplant prediction by BA Indices. For death prediction, the only endpoint/adverse event recorded was death, whereas for the death and/or liver transplant prediction, the endpoint/adverse event recorded was the occurrence of either death or liver transplant. Similar to the risk of death only, we found correlation between the risk of death and/or liver transplant and many BA indices ($P < 0.05$). For example, for a 20% increase in the % CDCA, the risk of death and/or liver transplant increases 1.25-fold ([HR]: 1.25; $P < 0.05$). Similarly, increasing levels of % CA, % HCA, total unamidated, % T-amidation, % non-12α-OH, Primary/ Secondary, and % primary BA significantly increased the risk of death and/or liver transplant, whereas decreasing levels of % LCA, % DCA, % mono-OH, 12α-OH/ non-12α-OH, % 12α-OH, and % secondary BA significantly increased the risk of death and/or liver transplant.

We performed the same analysis for demographic and non-BA parameters as well (**Table 3.5).** The risk of death and/or liver transplant was also significantly higher in males than females. Increasing levels of INR, protime, bilirubin, AST/ALT, MELD and APRI significantly increased the risk of death and/or liver transplant, whereas decreasing levels of albumin significantly increased the risk of death and/or liver transplant.

3.3.3. Multivariate Cox regression analysis

3.3.3.i. Death prediction

In multivariate analysis, a backward elimination regression was used to retain the most significant BA variables. The only BA variables retained in the multivariate model were %CDCA and %Tri-OH, which were independently predictive of survival **(Table 3.6.a)**. For example, a 20% increase in the % CDCA and % Tri-OH increases the risk of death by 1.34-fold (HR: 1.34; $P < 0.05$) and 1.14-fold (HR: 1.14; $P < 0.05$), respectively. The BA score (BAS) for individual patients can be calculated from this equation:

$$\text{BA score (BAS) for death} = 0.039 \times \%\,\text{CDCA} + 0.052 \times \%\,\text{Tri OH}$$

For example, for a patient with %CDCA of 20%, and a % Tri-OH of 50%, the BA score (BAS) would be 3.38.

We performed the same multivariate Cox regression analysis for demographics and non-BA parameters as well. For demographic variables, gender was significant in univariate analysis, but did not retain in multivariate analysis when included in the BA model building. In contrast, gender retained in the multivariate analysis for the non-BA model, but with minimal improvement of model goodness of fit and validation (the Bootstrapping, Schoenfeld residuals, areas under the ROC curve (AUC), and Kaplan-Meier analyses). Therefore, we did not include gender in the multivariate Cox models and

AST/ALT ratio was the only significant predictive variable of death **(Table 3.6.b)**. For example, a 20% increase in the AST/ALT, increases the risk of death by 1.36-fold (HR: 1.36; P < 0.05). The non-BA score (non-BAS) for individual patients can be calculated from this equation:

$$\text{non BA score (non BAS) for death} = 1.236 \times \text{AST/ALT}$$

In addition, we used the same methodology to develop other models including: (i) mixed BA and non-BA variables including demographics to test how the performance of a global BA- and non-BA mixed model compares to the BA-only and non-BA-only models (ii) MELD variables with coefficients from our data set to create a model with the original MELD variables, but with model coefficients derived from our data set (iii) original MELD modified with BA and/or non-BA variables including demographics, to test if the performance of the original MELD can be improved by adding significant BA and non-BA parameters from the univariate analysis and vice versa (**Appendix Table A)**. Overall, none of these strategies produced any statistically significant models neither they did improve the BA or non-BA-only model; therefore, were not further evaluated or validated.

3.3.3.ii. Death and/or Liver Transplant (LT) prediction

The only BA variables retained in the multivariate model were % primary and % DCA, which were independently predictive of death and/or liver transplant **(Table 3.7.a)**. For example, a 20% increase in the % primary increases the risk of death and/or liver transplant by 1.23-fold (HR: 1.226; P < 0.05), while 20% increase in the % DCA decreases the risk of death and/or liver transplant by 0.86-fold (HR: 0.857; P < 0.05). The BA score (BAS) for individual patients can be calculated from this equation:

$$\text{BA score (BAS) for death and/or LT} = 0.021 \times \%\ \text{Primary} - 0.049 \times \%\ \text{DCA}$$

We performed the same multivariate Cox regression analysis for demographics and non-BA parameters as well. For the same reason, gender did not retain in multivariate analysis when included in the BA model. In contrast, gender retained in the multivariate analysis for non-BA model, but with minimal improvement of model goodness of fit and validation. Therefore, we did not include gender in the multivariate Cox models and albumin was the only significant predictive variable of death and/or liver transplant **(Table 3.7.b)**. For example, a 20% increase in the albumin, decreases the risk of death and/or liver transplant by 0.39-fold (HR: 0.393; P < 0.05). The non-BA score (non-BAS) for individual patients can be calculated from this equation:

$$\text{non BA score (non BAS) for death and/or LT} = -1.277 \times \text{Albumin } (\frac{\text{g}}{\text{dL}})$$

Similar to the death models, we have developed other non-BA and mixed BA and non-BA models. None of these strategies produced any statistically significant models neither they did improve the BA or non-BA-only model (**Appendix Table B)**; therefore, were not further evaluated or validated.

3.3.4. Model Performance, Goodness of fit and Validation

3.3.4.i. Death prediction

Goodness of fit was performed by testing PH (proportional hazard) assumption for all the covariates of the final Cox model as well as for the global model as a whole, using a statistical test and a graphical diagnostic based on Schoenfeld residuals. The Schoenfeld residuals are independent of time. The plot gives an estimate of the time-dependent coefficient beta (t). Therefore, a graphical diagnostic that shows a non-random pattern against time is evidence of violation of the PH assumption. The PH assumption is supported by a non-significant relationship between residuals and time and is refuted by a significant relationship. The statistical test was not significant for both covariates in the

BA model, which were % CDCA (p-value = 0.854) and %Tri-OH (p-value = 0.970) as well as for the global model as a whole (p-value = 0.974). In addition, from the graphical inspection, there was no pattern with time (**Figure 3.1.a)**. Therefore, we can conclude that the PH assumption was met indicating the model is valid. Similarly, the Schoenfeld residual plots and p-value = 0.199 supported the validity of the non-BA model (**Figure 3.1.b)**.

We also used the bootstrapping validation. We are currently working on building internal and eventually external data sets for model validation. The bootstrapping is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement. Random samples were taken with replacement from our data set, one at a time, to create a series of 1000 new data sets and statistics are calculated by comparing these data sets. Bootstrapping validation results for the BA and non-BA models indicate that our regression coefficients were in the range of the 95% confidence intervals, p-values were statistically significant for each covariate (p-value < 0.05), bias values were very small (0.001 to 0.026) and standard error values were also very small (0.009 to 0.342) (**Table 3.8.a)**. We can conclude that the Bootstrapping validation results supported the validity of the BA and non-BA models.

Figure 3.2 shows the receiver operating characteristics (ROC) curves of the models for death prediction. For 5-year death prediction, the area under the ROC curves (AUC) for BAS, non-BAS, and MELD were 0.740, 0.653, and 0.683, respectively. For 3-year death prediction, the AUC for BAS, non-BAS, and MELD were 0.761, 0.664, and 0.715, respectively. Potential cut-off values selected based on the optimum sensitivity and specificity for different models. The ROC-optimum scores for BA, non-BA, and MELD models for death prediction were 2.71, 1.72, and 10, respectively (**Table 3.9.a**).

3.3.4.ii. Death and/or Liver Transplant (LT) prediction

The statistical test based on Schoenfeld residuals was not significant for both covariates in the BA model which were % DCA and % primary (p-values= 0.322, and 0.494, respectively), as well as for the global model as a whole (p-value = 597). In addition, from the graphical inspection, there was no pattern with time (**Figure 3.3.a)**. Therefore, we can conclude that the PH assumption was met indicating the model is valid. Similarly, the Schoenfeld residual plot and p-value = 0.193 supported the validity of the non-BA model (**Figure 3.3.b)**.

For death and/or LT prediction, bootstrapping validation results for the BA and non-BA models indicate that our regression coefficients were in the range of the 95% confidence intervals, p-values were statistically significant for each covariate (p-value < 0.05), bias values were very small (- 0.014 to 0.001) and standard error values were also very small (0.008 to 0.238) **(Table 3.8.b)**. Therefore, we can conclude that the bootstrapping validation results supported the validity of the BA and non-BA models for death and/or liver transplant prediction.

Figure 3.4 shows the receiver operating characteristics (ROC) curves of the models for death and/or LT prediction. For 5-year, the area under the ROC curves (AUC) for BAS, non-BAS, and MELD were 0.748, 0.743, and 0.763, respectively. For 3-year, the AUC for BAS, non-BAS, and MELD were 0.769, 0.758, and 0.789, respectively. The ROC-optimum scores for BA, non-BA, and MELD models for death and/or LT prediction were 0.76, -4.41, and 10, respectively (**Table 3.9.b**).

3.3.5. Survival Prediction

3.3.5.i. Death prediction

Table 3.10.a presents the estimated survival probability (S_0 (t)) for a patient with an average BA score (BAS_0) of 2.24 (the average BAS from all 257 patients in this study) for different time points. To obtain the survival probability for t years (S (t)), first BAS is calculated, S_0 (t) is identified from **Table 3.10.a,** and finally S (t) is calculated using this equation:

$$\text{Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(BAS - BAS_{\circ})}$$

Where, BAS_0 is the average BA score from all patients in this study; namely 2.24, while BAS is the BA score for that particular patient. For the same example patient discussed above, the probability of surviving for at least 3 years is:

$$\text{Survival Probability for (3)years} = 0.934^{\exp(3.38 - 2.24)} = 0.81 = \%81$$

The relationship between estimated 5- and 3- year survival probability (S(t)) and the BA score in patients with liver disease are shown in **Figure 3.5**.**a**. Survival probability decreases as a function of BA score. For example, the 5-year survival probability for patients with BA scores of 1.2 (25^{th} percentile of the population), 2.1 (50^{th} percentile of the population i.e. median), and 3.1 (75^{th} percentile of the population) are 97%, 93%, and 82%, respectively. Similarly, the 3-year survival probability for patients with the same BA scores above, are 98%, 94%, and 85%, respectively.

Table 3.10.b presents the estimated survival probability (S_0 (t)) for a patient with an average non-BA score (non-BAS_0) of 1.58 for different time points. The survival probability for (t) years is calculated using this equation:

$$\text{Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{non BAS} - \text{non BAS}_{\circ})}$$

The relationship between estimated 5- and 3- year survival probability (S (t)) and the non-BA score in patients with liver disease are shown in **Figure 3.5.b**. For example, the 5-year survival probability for patients with non-BA scores of 1.1 (25^{th} percentile of the population), 1.4 (50^{th} percentile of the population), and 1.9 (75^{th} percentile of the population) are 92%, 90%, and 83%, respectively. Similarly, the 3-year survival probability for patients with the same non-BA scores above, are 95%, 91%, and 86%, respectively.

By the end of the study, up to 7 years monitoring of 257 patients with cholestatic liver diseases, 27 patients (10.5%) have died. The Kaplan-Meier estimator was used to estimate subjects' survival free of adverse events over time. We have tried two cut-off values of the BAS to define high vs. low risk of death: (i) the median of the BAS of the population (2.19) (ii) and the cut-off value of the BAS with optimum specificity and sensitivity based on ROC analysis (2.71) (**Figure 3.6.a**). The estimated mean survival time was 71 months (5.9 years) for the high-risk group and 82 months (6.8 years) for the lower risk group based on the median BAS of 2.19. While the estimated mean survival time was 67 months (5.6 years) for the high-risk group and 80 months (6.7 years) for the lower risk group based on the ROC-optimum BAS (2.71) (**Table 3.11**). The P values of the log rank test and Breslow test were statistically significant (P-value < 0.05) for both cut-offs, indicating the both cut-offs of BAS, can differentiate low vs. high risk of death.

Figure 3.6.b shows the Kaplan Meier survival for the high vs. low risk of death groups based on the median (1.44) and the ROC-optimum (1.72) for the non-BAS. The estimated mean survival time was 74 months (6.2 years) for the high-risk group and 79 months (6.6 years) for the lower risk group based on the median non-BAS of 1.44. The P value from the log rank test and Breslow test were insignificant (p-value > 0.05), indicating the median of non-BAS (1.44) cannot differentiate low vs. high risk of death. While the estimated mean survival time was 71 months (5.9 years) for the high-risk group and 80

months (6.6 years) for the lower risk group based on the ROC-optimum of the non-BAS (1.72) (**Table 3.11**). The P values of the log rank test and Breslow test were < 0.05 for ROC-optimum, indicating the ROC-optimum of non-BAS can differentiate low vs. high risk of death.

Figure 3.6.c shows the Kaplan Meier survival for the high vs. low risk of death groups based on the median (11) and the ROC-optimum (10) for the MELD model. The estimated mean survival time was 74 months (6.2 years) for the high-risk group and 78 months (6.5 years) for the lower risk group based on the median MELD of 11. The P value from the log rank test and Breslow test were insignificant (p-value > 0.05), indicating the median of MELD (11) cannot differentiate low vs. high risk of death. While the estimated mean survival time was 67 months (5.6 years) for the high-risk group and 79 months (6.6 years) for the lower risk group based on the ROC-optimum of the MELD (10) (**Table 3.11**). The P value of the log rank test and Breslow test were statistically significant (p-value < 0.05) for ROC-optimum, indicating the ROC-optimum of MELD, can differentiate low vs. high risk of death.

3.3.5.ii. Death and/or Liver Transplant (LT) prediction

Table 3.12.a presents the estimated liver transplant-free survival probability ($S_0(t)$) for a patient with an average BAS (BAS_0) of 0.43 for different time points. To obtain the liver transplant-free survival probability for t years (S (t)), first BAS is calculated, $S_0(t)$ is identified from **Table 3.12.a,** and finally S (t) is calculated using this equation:

$$\text{Liver transplant free Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(BAS - BAS_{\circ})}$$

The relationship between estimated 5- and 3- year liver transplant-free survival (S (t)) and the BA score in patients with liver disease are shown in **Figure 3.7.a**. For example, the 5-year liver transplant-free survival probability for patients with BA scores of -0.23 (25th

percentile of the population), 0.45 (50th percentile of the population), and 1.21 (75th percentile of the population) are 93%, 86%, and 73%, respectively. Similarly, the 3-year liver transplant-free survival probability for patients with the same BA scores above, are 94%, 87%, and 77%, respectively.

Table 3.12.b presents the estimated liver transplant-free survival probability (S_0 (t)) for a patient with an average non-BA score (non-BAS_0) of - 4.51 for different time points. To obtain the liver transplant-free survival probability for t years (S (t)), first non-BAS is calculated, S_0 (t) is identified from **Table 3.12.b,** and finally S (t) is calculated using this equation:

$$\text{Liver transplant free Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{non BAS} - \text{non BAS}_{\circ})}$$

The relationship between estimated 5- and 3- year liver transplant-free survival probability (S(t)) and the non-BA score in patients with liver disease are shown in **Figure 3.7.b**. For example, the 5-year liver transplant-free survival for patients with non-BAS of -5.10 (25th percentile of the population), -4.72 (50th percentile), and -4.08 (75th percentile) are 89%, 84%, and 70%, respectively. Similarly, the 3-year liver transplant-free survival probability for patients with the same non-BA scores above, are 90%, 85%, and 77%, respectively.

Figure 3.8.a shows the Kaplan Meier liver transplant-free survival for the high vs. low risk of death and/or liver transplant groups based on the median (0.45) and the ROC-optimum (0.76) of the BAS. The estimated mean liver transplant-free survival time was 60 months (4.9 years) for the high-risk group and 79 months (6.6 years) for the lower risk group based on the median BAS (0.45). While the estimated mean liver transplant-free survival time was 56 months (4.6 years) for the high-risk group and 78 months (6.5 years) for the lower risk group based on the ROC-optimum of BAS (0.76) (**Table 3.13**). The P values of the log rank test and Breslow test were statistically significant (P-value < 0.05)

for both cut-offs, indicating the both cut-offs of BAS, can differentiate low vs. high risk of death and/or liver transplant.

Figure 3.8.b shows the Kaplan Meier liver transplant-free survival for the high and low risk groups based on the median (- 4.72) and the ROC-optimum (- 4.41) of non-BAS. The estimated mean liver transplant-free survival time was 62 months (5.2 years) for the high-risk group and 79 months (6.6 years) for the lower risk group based on the median non-BAS (- 4.72). While, the estimated mean liver transplant-free survival time was 60 months (5 years) for the high-risk group and 78 months (6.5 years) for the lower risk group based on the ROC-optimum of non-BAS (- 4.41) (**Table 3.13**). The P values of the log rank test and Breslow test were statistically significant (P-value < 0.05) for both cut-offs, indicating the both cut-offs of non-BAS, can differentiate low vs. high risk of death and/or liver transplant.

For MELD model**, Figure 3.8.c** shows the Kaplan Meier liver transplant-free survival for the high vs. low risk groups based on the median (9) and the ROC-optimum (10) of the MELD score. The estimated mean liver transplant-free survival time was 64 months (5.3 years) for the high-risk group and 78 months (6.5 years) for the lower risk group based on the median MELD of 9. While the estimated mean liver transplant-free survival time was 65 months (5.4 years) for the high-risk group and 75 months (6.3 years) for the lower risk group based on the ROC-optimum of the MELD (10) (**Table 3.13**). The P values of the log rank test and Breslow test were statistically significant (P-value < 0.05) for both cut-offs, indicating the both cut-offs of MELD, can differentiate low vs. high risk of death and/or liver transplant.

3.4. Discussion

Biomarkers that are currently used in the clinic for the diagnosis and prognosis of liver diseases are primarily serum liver enzymes such as AST and ALT as well as bilirubin. However, these markers have numerous shortfalls including the lack of specificity to liver or bile duct injuries, where their levels can be elevated in non-hepatobiliary diseases such as hyperthyroidism, adrenal, heart, and muscle disorders. Also, they require severe cell injury before their levels increase in the blood [4, 5]. Multifactorial models with multiple parameters based on these biomarkers are also frequently used and offer advantages compared to the use of their individual biomarker components such as the Child-Turcotte-Pugh (CTP) score. CTP is calculated from these five variables: encephalopathy grade, severity of ascites, bilirubin (mg/dL), albumin (g/dL), and INR. However, the portosystemic encephalopathy and severity of ascites can be considered as subjective variables, which highly depend on the physician judgement making these diagnosis [6] [30].

More recently, the MELD model was developed to predict three-month mortality of patients with end-stage liver disease [7, 8]. MELD is also currently used in many countries to classify liver diseases' patients awaiting transplantation to identify patients with the highest priority for LT [9]. However, despite its widespread application, MELD has some limitations. MELD is based on three objective laboratory variables, that are not necessarily liver specific. For example, patients may have an elevated serum creatinine from an underlying kidney disease not related to hepatorenal syndrome. In addition, serum bilirubin can be elevated in cases of hemolysis or sepsis. An elevated INR can also be secondary to warfarin use. Any of these cases can increase the MELD score and overestimate the severity of liver disease [10, 11]. Furthermore, several studies have shown that patients with cholestatic liver diseases may still have high mortality rates despite having low MELD scores [11, 12].

BAs have been suggested to be used as biomarkers of cholestatic diseases for decades [14, 18, 31, 32]. However, this was not translated into the clinic, primarily because of the high inter- and intra-individual variability of serum and/or urine BA concentrations due to numerous factors such as food ingestion, diurnal variation, medication intake, gender, alcohol consumption, and obesity [1, 18, 33-36]. To this end, we have developed the concept of BA indices (ratios of individual BA and metabolite concentrations), which provide comprehensive quantification of the composition, hydrophobicity, toxicity, and metabolism of total and individual BA [1, 18-20]. In contrast to the absolute concentrations of BA, we have shown that BA indices calculated from urine or serum have markedly low inter- and intra-individual variability and were more resistant to food intake, gender differences, BMI and age effects [1, 18, 20]. This facilitate the use of BA indices as biomarkers for the diagnosis of hepatobiliary diseases, and we have shown that they outperformed many of the currently used markers [1, 18].

In this study, we have extended the application of BA indices to predict the prognosis of liver diseases. We developed survival models based on BA indices to predict the prognosis of hepatobiliary diseases in terms of progressing into the end points/adverse events of death only and death and/or liver transplant over a 3- and 5-year periods of time. Cox proportional hazards regression was used. Individual BA and non-BA variables including demographics were analyzed as possible predictors of survival in a univariate Cox regression analysis. To build the multivariate model a backward elimination method was used to retain the most significant variables, which were identified from the univariate analysis. In addition to the BA model, we have constructed: (i) non-BA, (ii) mixed BA and non-BA variables to compare with the BA-only and non-BA-only models(iii) MELD variables with coefficients from our data set to create a model with the original MELD variables, but with model coefficients derived from our data set, (iv) original MELD

modified with BA and/or non-BA variables, to test if the performance of the original MELD can be improved by adding significant BA and non-BA parameters from the univariate analysis.

The final multivariate survival models were then validated using the bootstrapping, and goodness of fit was performed by testing the proportional hazards (PH) assumption. Finally, the various models were compared using the above validation criteria and Kaplan-Meier (KM) and receiver operating characteristic (ROC) analyses.

According to the univariate cox regression analysis, % CDCA, % CA, % HCA, % T-amidation, % Tri-OH, % non-12α-OH, Primary/ Secondary, and % primary BA were proportional to the risk of death, whereas % LCA, % DCA, % G-amidation, % Mono-OH, 12α-OH/ non-12α-OH, % 12α-OH, % secondary BA were inversely proportional to the risk of death (**Table 3.2)**. For demographics and non-BA parameters, the risk of death was significantly higher in males than females and increasing levels of INR, protime, bilirubin, AST/ALT, APRI, and MELD significantly increased the risk of death, whereas decreasing levels of albumin significantly increased the risk of death **(Table 3.3)**.

Using the multivariate cox regression analysis, we have constructed these final models for death prediction:

(i) The BA score (BAS) model for death prediction**:**

$$\text{BA score (BAS) for death} = 0.039 \times \%\,\text{CDCA} + 0.052 \times \%\,\text{Tri OH}$$

(ii) The non-BA score model (non-BAS) model for death prediction**:**

$$\text{non BA score (non BAS) for death} = 1.236 \times \text{AST/ALT}$$

BAS in this population ranged from 0-4, while the non-BAS ranged from 0.44-4.98.

Similar models were built for the prediction of death and/or liver transplant (LT)**:**

$$\text{BA score (BAS) for death and/or LT} = 0.021 \times \%\ \text{Primary} - 0.049 \times \%\ \text{DCA}$$

$$\text{non BA score (non BAS) for death and/or LT} = -1.277 \times \text{Albumin}\ (\frac{\text{g}}{\text{dL}})$$

Cholestatic diseases are associated with impaired bile flow to the intestine, which is expected to translate into reduced transformation of primary BA including CDCA and CA into secondary BA by intestinal bacteria. Therefore, accumulation of primary BA in the blood may indicate further impairment in bile flow and worsening of the liver diseases [1, 37-40]. This is in agreement with the BAS model, where increased % CDCA and % Tri-OH BA (primarily consists of CA) were the most significant predictors of liver disease prognosis into death. Another interpretation for the accumulation of CDCA could be related to the fact that CDCA is the best substrate for Bile Salt Export Pump (BSEP), which is responsible for the efflux transport of BA across the canalicular membrane from hepatocytes into bile. Therefore, loss of BSEP function could be associated with the progression of the liver disease [41, 42], which leads to CDCA accumulation in the liver and eventually in the systemic blood. Similarly, the BAS model for death and/or LT predicts that the increase in % primary BA, while the decrease in % DCA (secondary BA) are the most significant predictors of liver disease prognosis into death and/or LT.

None of the other approaches including mixed BA and non-BA variables, MELD variables with coefficients from our data set, or the original MELD modified with BA and/or non-BA variables have resulted in any statistically significant models neither they did improve the above BAS or non-BAS models. Therefore, these models were not further evaluated or validated (**Appendix Table A)**.

Goodness of fit was performed by testing PH (proportional hazard) assumption using a statistical test and a graphical diagnostic based on Schoenfeld residuals. For

death prediction, the PH assumption was met in both BA and non-BA models supporting their validity (**Figure 3.1)**. In addition, we used the bootstrapping method for model validation. Bootstrapping validation results supported the validity of both the BA and non-BA models for death prediction **(Table 3.8.a)**. Also, the PH assumption was met (**Figure 3.3)** and the bootstrapping validation (**Table 3.8.b)** supported the validity for both the BA and non-BA models for death and/or LT prediction. Further validation efforts are also ongoing to build internal and eventually external data sets for more rigorous model validation.

We used ROC analysis to compare the accuracy of our prognostic models. The higher the AUC under the ROC curve, the greater the overall accuracy of the marker in distinguishing between groups. For prognostic models, AUC of 0.9 or greater is rarely seen, AUC between 0.8 and 0.9 indicates excellent diagnostic accuracy, and any AUC over 0.7 may be considered clinically useful [43-45]. Also, ROC curves are used to determine cut-off values which quantify the normal ranges of biomarkers. The selection of optimum cut-off values is a tradeoff between sensitivity and specificity. Accordingly, scores for the BA, non-BA, and MELD models for death prediction of 2.71, 1.72, and 10, respectively, were identified as cut-off values with optimum sensitivity vs. specificity **(Table 3.9.a).** These ROC-optimum cut-off values were used in KM analysis as potential cut-off scores, above which subjects are considered at higher risk of death as described later.

For 5-year death prediction, the AUC for BAS was 0.74 compared to 0.65 for non-BAS and 0.68 for MELD models (**Figure 3.2.a).** Similarly, for 3-year death prediction, the AUC for BAS was 0.76 compared to 0.66 for non-BAS and 0.71 for MELD models (**Figure 3.2.b).** In addition, BAS sensitivity in death prediction (74% vs. 67% and 62%) was 7% and 12% higher than non-BAS and MELD, respectively. BAS specificity was also higher than non-BAS and MELD (68% vs. 66% and 64%). Therefore, ROC analysis show that

BAS is more accurate and results in higher true-positive and true-negative prediction of death compared to both non-BAS and MELD.

Similar results were obtained from the death and/or LT prediction. The 5-year ROC analysis resulted in AUC of 0.74 for both BAS and non-BAS compared to 0.76 for MELD (**Figure 3.4.a).** The 3-year ROC analysis resulted in AUC of 0.76, 0.75, and 0.78 for BAS, non-BAS, and MELD, respectively (**Figure 3.4.b).** ROC sensitivity were 71, 78, and 70%, while specificity were 72, 70, and 66% for BAS, non-BAS, and MELD, respectively (Table 3.9). Therefore, ROC analysis shows all three models have similar accuracy with similar rates of true-positive and negative rates. It has to be noted that liver transplant allocation system is currently based upon the MELD score, where organs allocation is assigned preferentially to patients with the highest MELD scores [7, 46, 47]. Therefore, unlike death prediction, any models to predict LT will be biased in favor of MELD.

The Cox survival model can be used to predict the survival probability at any time point. The survival probability for t years (S (t)) was calculated for every subject using both BAS and non-BAS models, as:

$$\text{Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{BAS} - 2.24)}$$

$$\text{Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{non BAS} - 1.58)}$$

Where S_0 (t) presents the estimated survival probability for a patient with an average BA score of 2.24 or non-BA score of 1.58 for different time points **(Table 3.10)**.

As shown in **Figure 3.5,** both 5- and 3-year survival probabilities decrease as a function of both BA and non-BA scores. For example, the 3-year survival probability for patients with BA scores of 1.2 (25th percentile of the population), 2.1 (50th percentile of the

population i.e. median), and 3.1 (75th percentile of the population) are 98%, 94%, and 85%, respectively. While, the 3-year survival probability for patients with equivalent non-BA scores (25th, 50th, and 75th population percentiles) are 95%, 91%, and 86%, respectively.

The Kaplan-Meier estimator was used to estimate subjects' survival free of adverse event over time. Two cut-off values of the scores to define high vs. low risk of death were proposed: (i) the median of the scores of the population and (ii) the cut-off value of the scores with optimum specificity and sensitivity based on ROC analysis as discussed above. Both median and ROC-optimum cut-offs for BAS (2.19 and 2.71) were able to differentiate low vs. high risk of death. While only the ROC-optimum cut-off for non-BAS (1.72) and the ROC-optimum cut-off for MELD (10) were able to differentiate low vs. high risk of death **(Figure 3.6** and **Table 3.11)**.

Twenty-three patients with high BAS (> the median BAS of 2.19) died vs. four patients with low BAS (< the median BAS of 2.19) for the entire study. Therefore, 19 more patients died with high compared to low BAS. In contrast, nine and five more subjects with high non-BAS and high MELD have died compared to low non-BAS and low MELD, respectively. Also, patients with low BAS lived for an average of 82 months, while patients with high BAS lived for an average of 71 months since their diagnosis with the liver diseases. Therefore, patients with low BAS lived 11 months longer than patients with high BAS. On the other hand, patients with low non-BAS or low MELD (<median score), lived, in average, for only five or four months longer, compared to the high non-BAS or high MELD (high score), respectively (**Table 3.11)**.

Consequently, the shortening of lifespan between patients with high vs. low BAS was 6-7 months more compared to high non-BAS or high MELD (11 months longer

lifespan with low BAS minus 5 or 4 months longer lifespan with low non-BAS or low MELD = 6-7 months). Also, the number of deaths with high BAS is 2-4-fold (19 more patients died with high BAS ÷ 9 or 5 more patients died with high non-BAS or high-MELD = 2-4) higher than that with high non-BAS or high MELD. Therefore, it can be concluded that in this patient population, patients with high BAS are at a much higher risk of death compared to patients with high MELD or high non-BAS.

Similarly, for the prediction of death and/or LT, the liver transplant-free survival probability for t years (S (t)) was calculated for every subject as:

$$\text{Liver transplant free Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{BAS} - 0.43)}$$

$$\text{Liver transplant free Survival Probability for (t) years: } S(t) = S_{\circ}(t)^{\exp(\text{non BAS} - (-4.51))}$$

Where S_0 (t) presents the estimated survival probability for a patient with an average BA score of 0.43 or non-BA score of -4.51 for different time points **(Table 3.12)**.

As shown in **Figure 3.7,** both 5- and 3-year liver transplant-free survival probabilities decrease as a function of both BA and non-BA scores. According to the KM analysis, both median and ROC-optimum cut-offs for BAS (0.45 and 0.76), non-BAS (-4.72 and -4.41), and MELD scores (9 and 10) were able to differentiate low vs. high risk of death and/or LT (**Figure 3.8 and Table 3.13**).

Forty-three patients with high (>median) BAS and non-BAS died and/or had LT vs. 9 patients with low (<median) BAS and non-BAS for the entire study. Therefore, 34 more patients died and/or had LT with high compared to low BAS and non-BAS. In contrast, 31 more subjects with high MELD died and/or had LT compared to low MELD. Also, patients with low BAS lived without the need for LT for an average of 79 months, after which they either died or had LT, while patients with high BAS lived without the need for LT for an

average of 60 months since their diagnosis with the liver diseases. Therefore, patients with low BAS lived without need for LT 19 months longer than patients with high BAS. On the other hand, patients with low non-BAS or low MELD (<median score), lived without need for LT, in average, for only 17 or 14 months longer, compared to the high non-BAS or high MELD (high score), respectively **(Table 3.13)**.

Consequently, patients with high BAS lived without need for LT 2-5 months less than patients with high non-BAS or high MELD. Therefore, it can be concluded that in this patient population, patients with high BAS are at a slightly higher risk of death and/or LT compared to patients with high MELD or high non-BAS.

3.5. Conclusions

We have developed and validated a survival model (the bile-acids score (BAS) model) based on BA indices to predict the prognosis of cholestatic liver diseases. Our results demonstrate that the BAS model is more accurate and results in higher true-positive and true-negative prediction of death compared to both non-BAS and MELD models. Both 5- and 3-year survival probabilities markedly decreased as a function of BAS. Moreover, patients with high BAS had a 4-fold higher rate of death and lived for an average of 11 months shorter than subjects with low BAS. The increased risk of death with high vs. low BAS was also 2-4-fold higher and the shortening of lifespan was 6-7-month lower compared to MELD or non-BAS. Similarly, we have shown the use of BAS to predict the survival of patients with and without LT. Therefore, BAS could be used to define the most seriously ill patients, who need earlier intervention such as LT. This will help provide guidance for timely care for liver patients.

3.6. Figures and Tables

Figure Legends

Figure 3.1. Schoenfeld residual plots for death prediction by the BAS and non-BAS models. The solid line is a smoothing spline fit to the plot, with the dashed lines representing a ± 2-standard-error band around the fit. The global Schoenfeld Test p-value: (a) = 0.974 for BAS, and (b) = 0.199 for non-BAS.

Figure 3.2. Receiver operating characteristics (ROC) curves of BAS, non-BAS, and MELD for death prediction. The area under the ROC curves (AUC) for BAS, non-BAS, and MELD for (a) 5-year, (b) 3-year death prediction.

Figure 3.3. Schoenfeld residual plots for death and/or liver transplant prediction by BAS and non-BAS models. The solid line is a smoothing spline fit to the plot, with the dashed lines representing a ± 2-standard-error band around the fit. The global Schoenfeld Test p-value:(a) = 0.597 for BAS, and (b) = 0.193 for non-BAS.

Figure 3.4. Receiver operating characteristics (ROC) curves of BAS, non-BAS, and MELD for death and/or liver transplant prediction. The area under the ROC curves (AUC) for BAS, non-BAS, and MELD for (a) 5-year, (b) 3-year death and/or liver transplant prediction.

Figure 3.5. Estimated 5- and 3-year survival (S(t)) from the BAS and non-BAS models. The relationship between estimated 5- and 3- year survival probability (S(t)) as a function of (a) BAS, (b) non-BAS. Q1, Q2, and Q3 are 25^{th}, 50^{th}, and 75^{th} percentiles of the population, respectively.

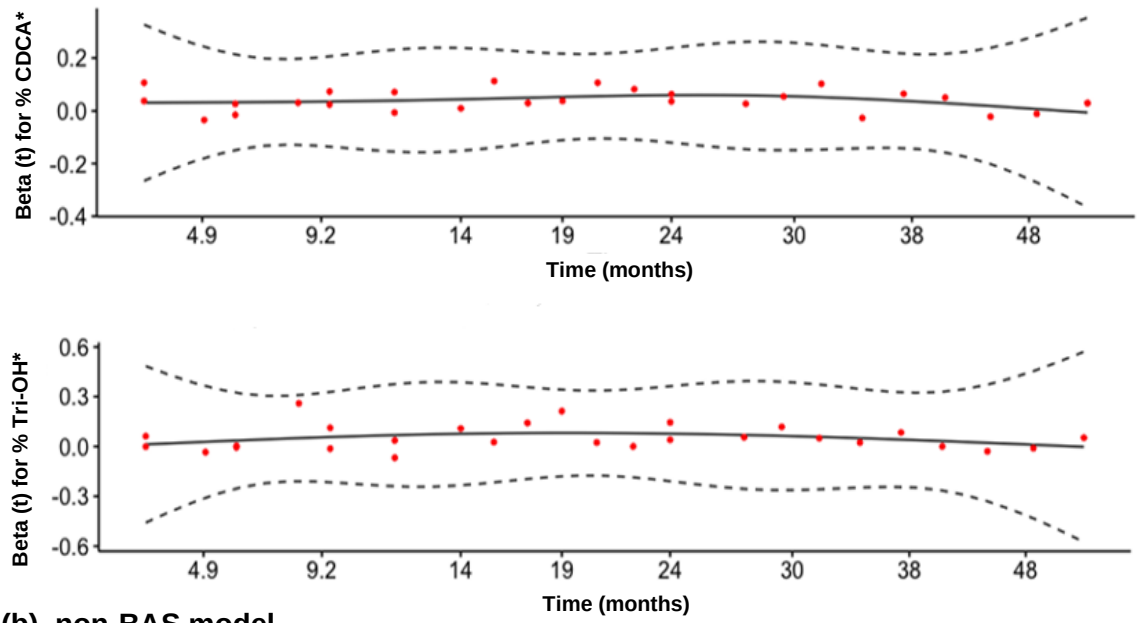
Figure 3.6. Kaplan-Meier survival plots for high vs. low BAS, non-BAS, and MELD models. The median and ROC-optimum cutoff values of the (i) BAS, (ii) non-BAS, and (ii) MELD were used to define high vs. low risk of death. "*" indicates P-values < 0.05 from the Log rank and Breslow tests.

Figure 3.7. Estimated 5- and 3-year liver transplant-free survival (S(t)) from the BAS and non-BAS models. The relationship between estimated 5- and 3- year liver transplant-free survival probability (S(t)) as a function of (a) BAS, (b) non-BAS. Q1, Q2, and Q3 are 25^{th}, 50^{th}, and 75^{th} percentiles of the population, respectively.

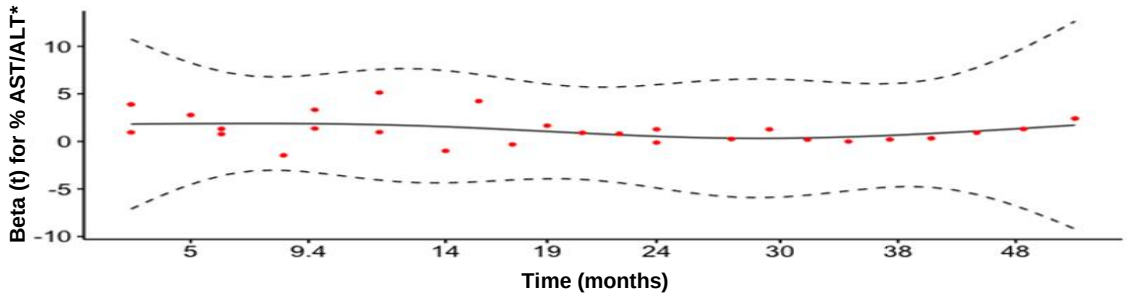
Figure 3.8. Kaplan-Meier liver transplant-free survival plots for high vs. low BAS, non-BAS, and MELD models. The median and ROC-optimum cutoff values of the (i) BAS, (ii) non-BAS, and (ii) MELD were used to define high vs. low risk of death and/or liver transplant. "*" indicates P-values < 0.05 from the Log rank and Breslow tests.

Figure 3.1. Schoenfeld residual plots for death prediction by the BAS and non-BAS models.

(a) BAS model



(b) non-BAS model

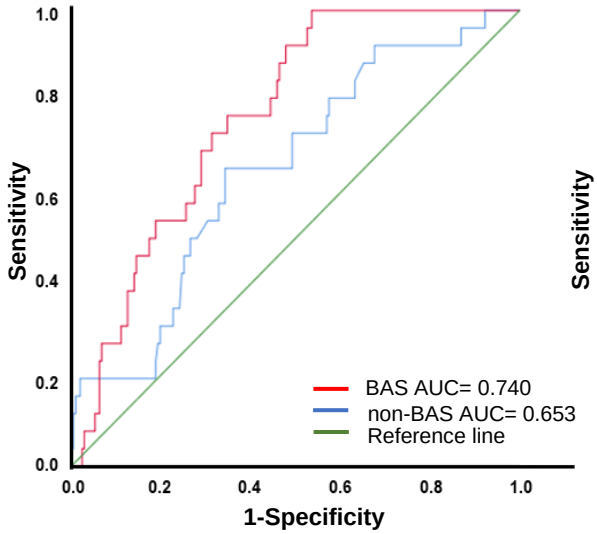


*Variable Schoenfeld Test p-values = 0.85, 0.97, 0.19 for %CDCA, %Tri-OH, and AST/ALT, respectively

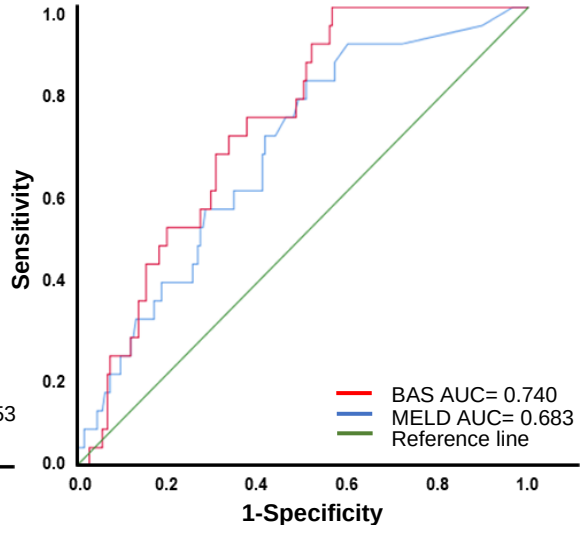
Figure 3.2. Receiver operating characteristics (ROC) curves of BAS, non-BAS, and MELD for death prediction.

(a) 5-year death prediction:

(i) BAS vs. non-BAS

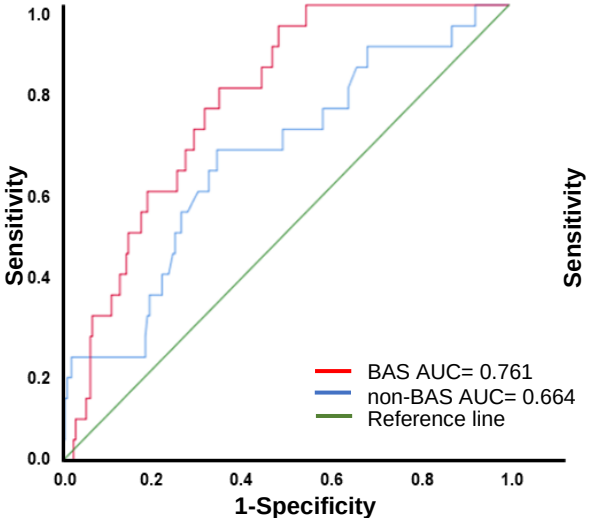


(ii) BAS vs. MELD



(b) 3-year death prediction:

(i) BAS vs. non-BAS



(ii) BAS vs. MELD

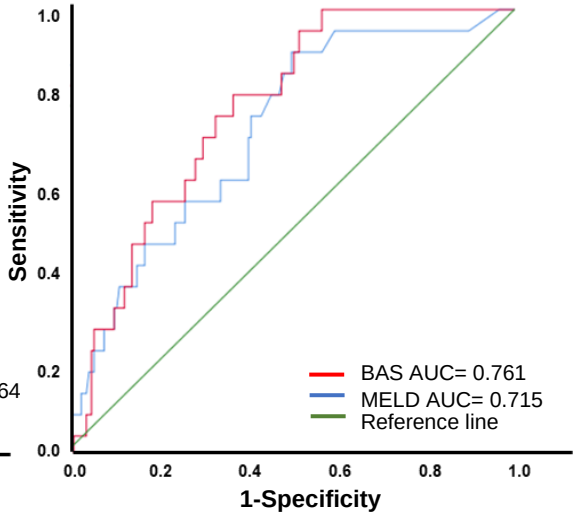
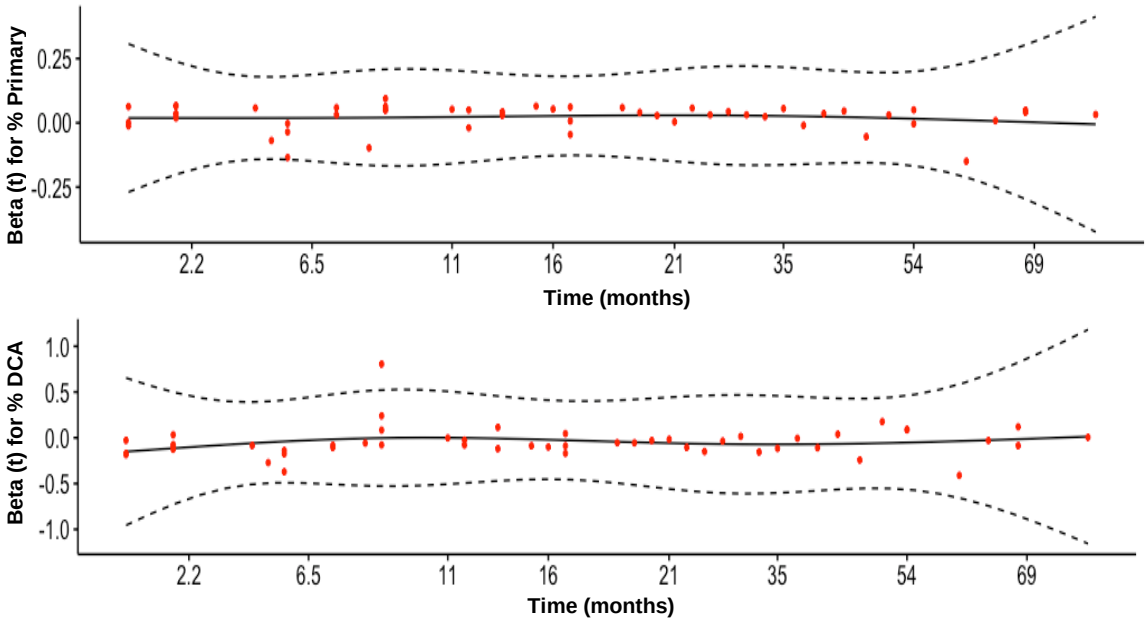
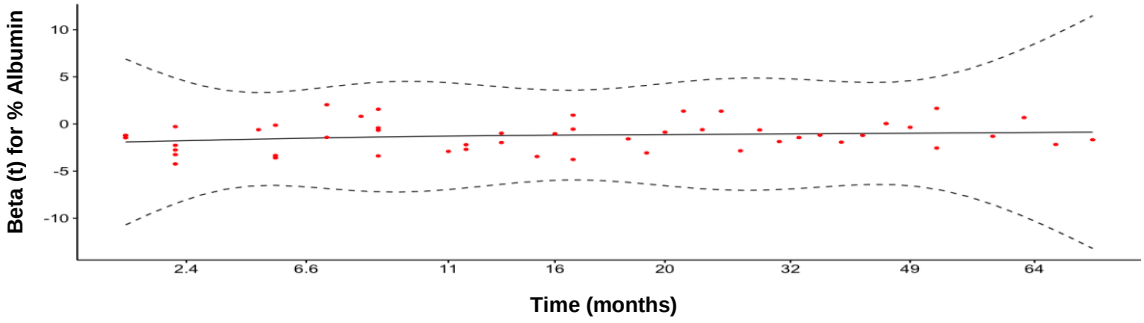


Figure 3.3. Schoenfeld residual plots for death and/or liver transplant prediction by BAS and non-BAS models.

(a) BAS model



(b) non-BAS model

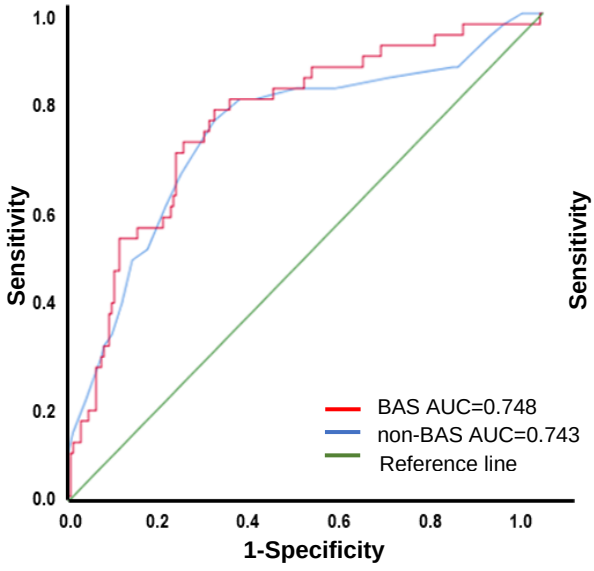


*Variable Schoenfeld Test p-values = 0.49, 0.32, 0.19 for %Primary, %DCA, and %Albumin, respectively

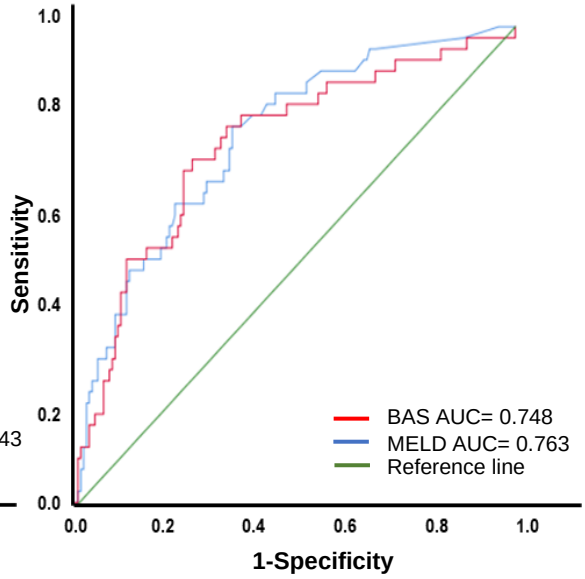
Figure 3.4. Receiver operating characteristics (ROC) curves of BAS, non-BAS, and MELD for death and/or liver transplant prediction.

(a) 5-year death and/or liver transplant prediction:

(i) BAS vs. non-BAS

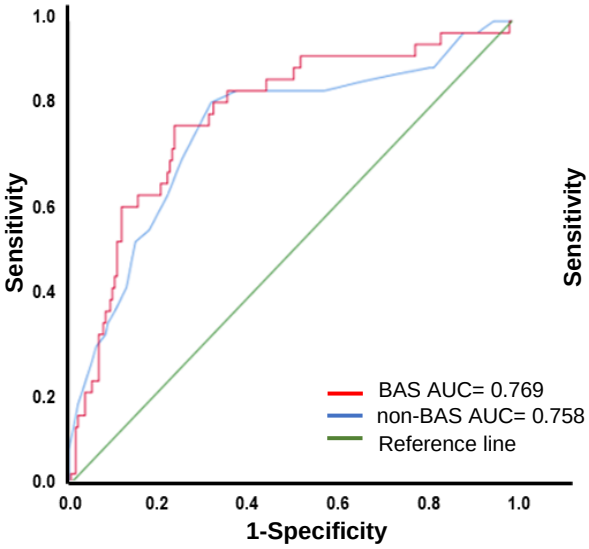


(ii) BAS vs. MELD



(b) 3-year death and/or liver transplant prediction:

(i) BAS vs. non-BAS



(ii) BAS vs. MELD

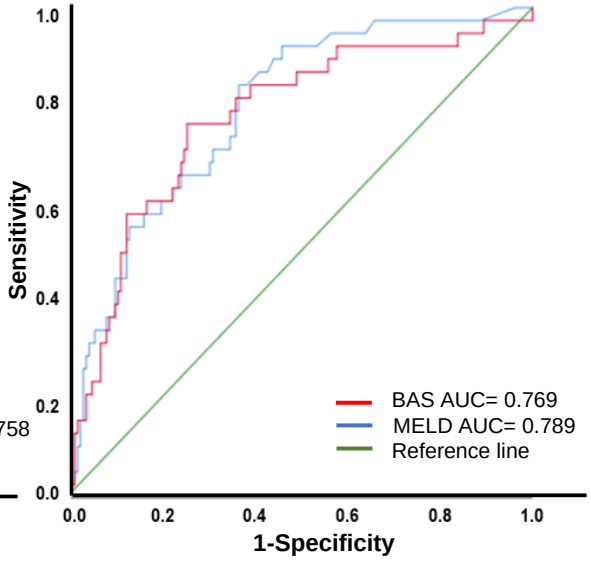
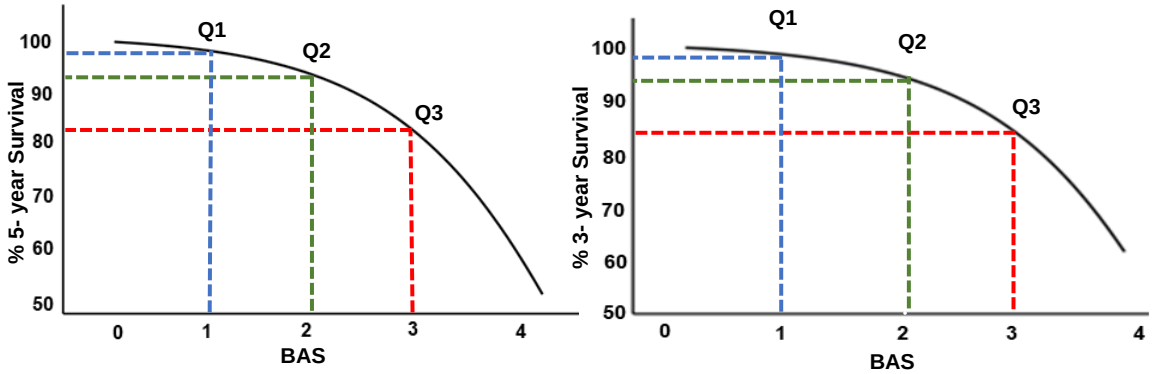


Figure 3.5. Estimated 5- and 3-year survival (S(t)) from the BAS and non-BAS models.

(a) BAS



(b) non-BAS

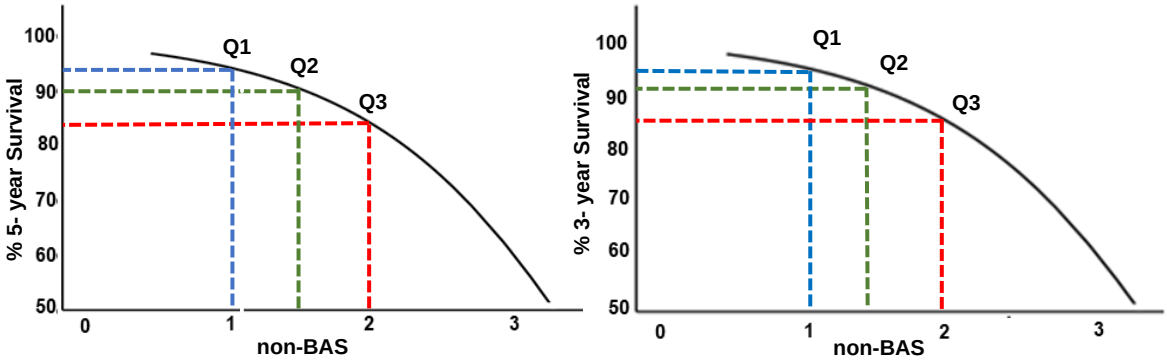


Figure 3.6. Kaplan-Meier survival plots for high vs. low BAS, non-BAS, and MELD models.

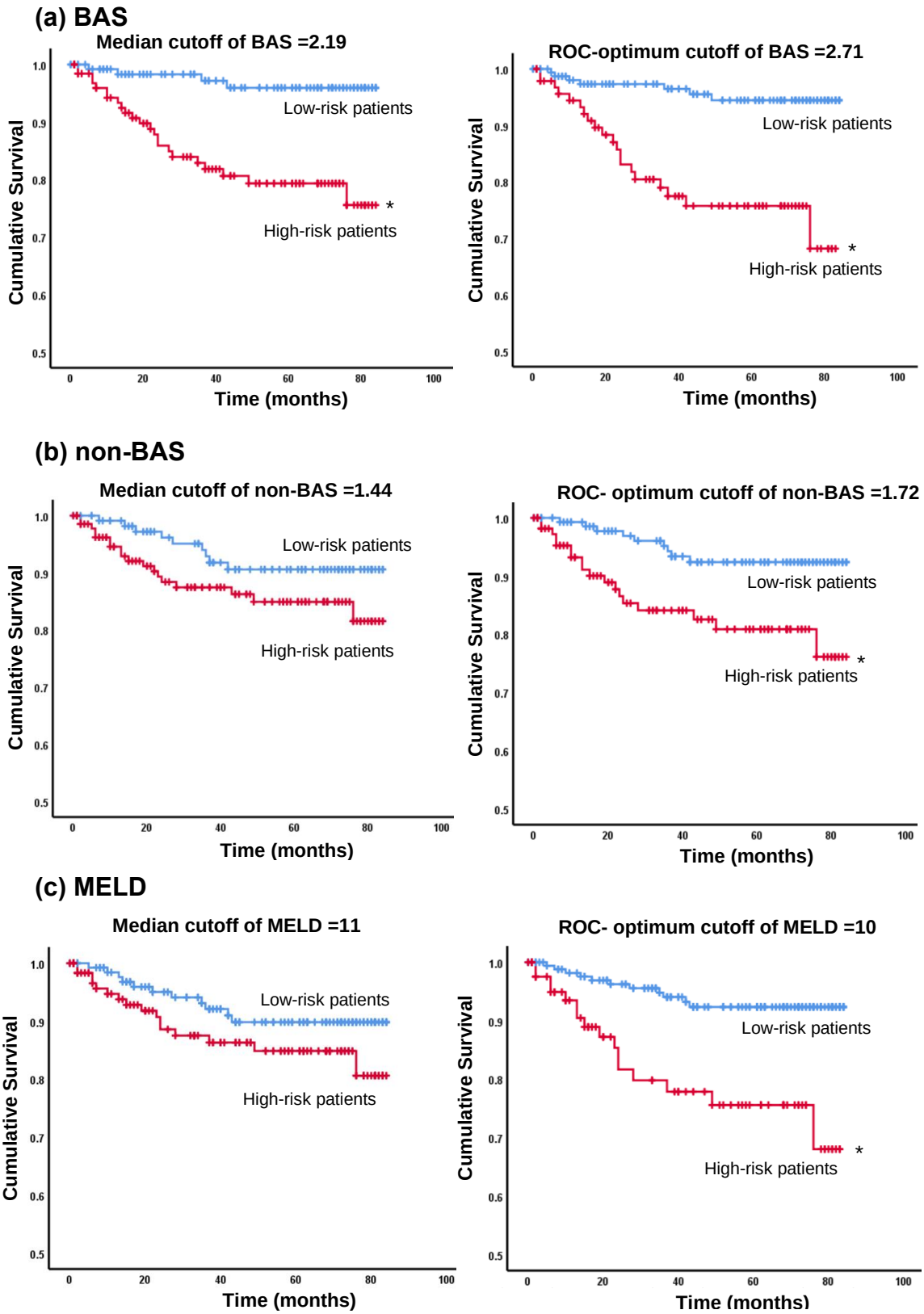
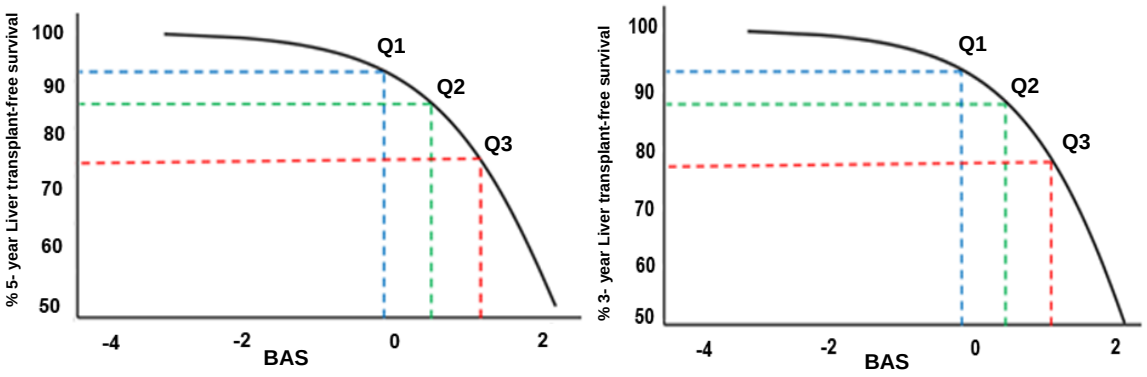


Figure 3.7. Estimated 5- and 3-year liver transplant-free survival (S(t)) from the BAS and non-BAS models.

(a) BAS



(b) non-BAS

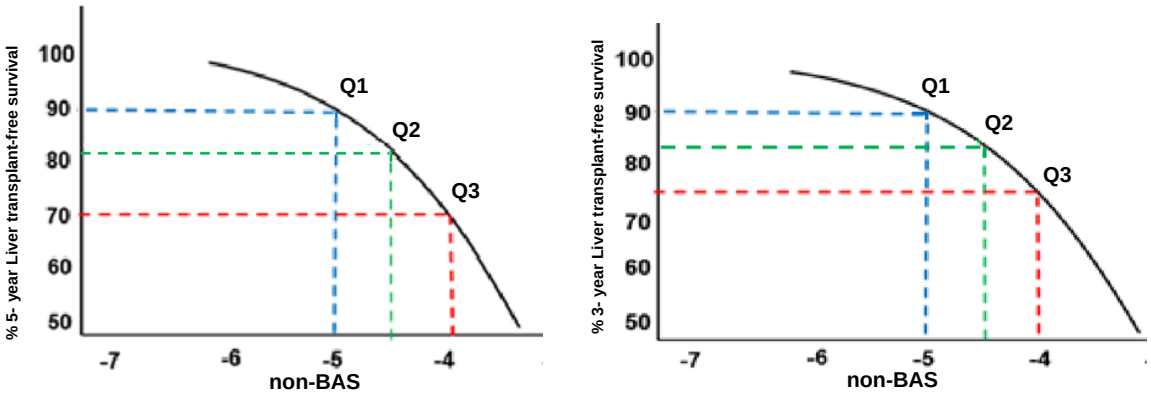


Figure 3.8. Kaplan-Meier liver transplant-free survival plots for high vs. low BAS, non-BAS, and MELD models.

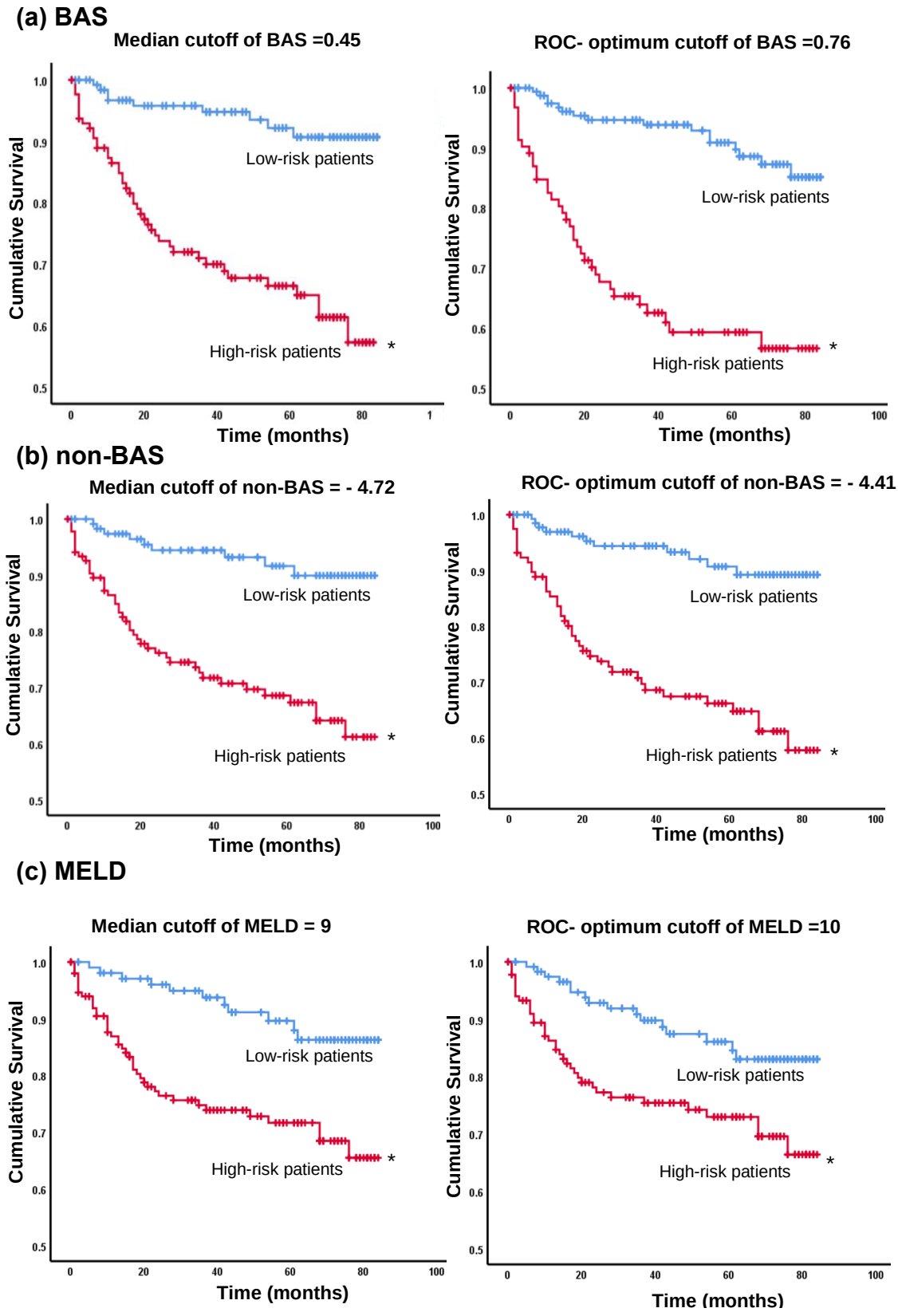


Table 3.1. Patient population characteristics.

	Patients	Death	Liver Transplant (LT)
N	257	27	25
Gender			
Male	136	21	17
Female	121	6	8
Age (years)			
Mean ± SEM	52.2 ± 0.71	55.9 ± 1.88	52.9 ± 2.1
Body Mass Index (BMI)			
Mean ± SEM	30.7 ± 0.45	29.65 ± 1.19	29.11 ± 0.45
Race			
White	217	26	24
Black	11	0	0
Asian	7	0	0
Hispanic	4	0	1
Others	18	1	0
Non-BA parameters (Mean ± SEM)			
Creatinine (mg/dL)	1.02 ± 0.09		
Albumin (g/dL)	3.53 ± 0.04		
INR	1.19 ± 0.02		
Protime (sec)	12.01 ± 0.42		
AST (U/L)	59.9 ± 4.07		
ALT (U/L)	54.9 ± 4.26		
Bilirubin (mg/dL)	1.75 ± 0.15		
AST/ALT	1.28 ± 0.04		
MELD	10.6 ± 0.34		
APRI	1.15 ± 0.11		

Table 3.2. Univariate Cox regression analyses for death prediction by BA Indices.

BA (μM)/ BA indices (%)	B-value (Regression Coefficient)	P-value	Hazard ratio (HR): Exp (B)		
			1 unit change	10% change	20% change
Total BA	-0.001	0.683	1.00	0.99	0.99
Total LCA	-0.063	0.331	0.94	0.98	0.96
Total UDCA	-0.005	0.477	1.00	0.99	0.98
Total CDCA	0.002	0.617	1.00	1.01	1.01
Total DCA	-0.047	0.266	0.95	0.97	0.94
Total HDCA	-13.76	0.424	0.00	0.98	0.97
Total MDCA	-5.021	0.347	0.01	0.98	0.95
Total CA	-0.005	0.783	1.00	1.00	1.00
Total MCA	-0.006	0.793	0.99	1.00	1.00
Total HCA	0.061	0.631	1.06	1.00	1.00
% LCA	-0.101	0.016	0.90	0.92	0.84
% UDCA	-0.027	0.070	0.97	0.94	0.89
% CDCA	0.031	0.000	1.03	1.12	1.26
% DCA	-0.092	0.001	0.91	0.87	0.76
% HDCA	-11.28	0.170	0.00	0.97	0.93
% MDCA	-0.325	0.599	0.72	0.99	0.98
% CA	0.067	0.003	1.07	1.04	1.09
% MCA	0.020	0.322	1.02	1.01	1.02
% HCA	0.454	0.015	1.57	1.02	1.04
Total Unamidated	0.007	0.804	1.01	1.00	1.00
Total G-amidated	-0.001	0.667	1.00	0.99	0.99
Total T-amidated	-0.002	0.799	1.00	1.00	1.00
% Amidation	0.012	0.577	1.01	1.11	1.24
% G-amidation	-0.025	0.032	0.98	0.82	0.68
% T-amidation	0.039	0.001	1.04	1.04	1.09
Total Unsulfated	-0.001	0.968	1.00	1.00	1.00
Total Sulfated	-0.001	0.659	1.00	0.99	0.99
% Sulfation	-0.023	0.081	0.98	0.82	0.68
Total Mono-OH	-0.063	0.331	0.94	0.98	0.96
Total Di-OH	-0.001	0.703	1.00	0.99	0.99
Total Tri-OH	-0.004	0.766	1.00	1.00	0.99
% Mono-OH	-0.101	0.016	0.90	0.92	0.84
% Di-OH	-0.004	0.761	1.00	0.97	0.94
% Tri-OH	0.034	0.011	1.03	1.04	1.09
Total 12α-OH	-0.012	0.465	0.99	0.99	0.98
Total non-12α-OH	-0.001	0.762	1.00	0.99	0.99
12α-OH/ non12α-OH	-2.837	0.020	0.06	0.91	0.83
CA/ CDCA	-0.099	0.828	0.91	1.00	1.00
% 12α-OH	-0.043	0.019	0.96	0.91	0.83
% non-12α-OH	0.043	0.019	1.04	1.40	1.95
Total Primary	0.001	0.769	1.00	1.00	1.01
Total Secondary	-0.006	0.397	0.99	0.98	0.96
Primary/ Secondary	0.037	0.016	1.04	1.01	1.02
% Primary	0.041	0.000	1.04	1.23	1.51
% Secondary	-0.041	0.000	0.96	0.82	0.67
HI	-0.092	0.912	0.91	1.00	1.00

Table 3.3. Univariate Cox regression analyses for death prediction by non-BA parameters and demographics.

Demographics and Non-BA parameters	B-value (Regression Coefficient)	P-value	Hazard ratio (HR): Exp (B)		
			1 unit change	10% change	20% change
Gender	1.251	0.007	3.49	-	-
Age (year)	0.029	0.093	1.03	1.16	1.35
BMI	-0.025	0.390	0.98	0.93	0.86
Race	*	0.950	*	*	*
Creatinine (mg/dL)	0.030	0.737	1.03	1.00	1.01
Albumin (g/dL)	-1.189	0.000	0.30	0.65	0.43
INR	0.781	0.013	2.19	1.10	1.20
Protime (sec)	0.073	0.002	1.08	1.09	1.19
AST (U/L)	0.002	0.443	1.00	1.01	1.02
ALT (U/L)	-0.003	0.437	1.00	0.98	0.97
Bilirubin (mg/dL)	0.096	0.035	1.10	1.02	1.03
AST/ALT	1.236	0.000	3.44	1.16	1.36
MELD	0.104	0.000	1.11	1.11	1.24
APRI	0.267	0.000	1.31	1.03	1.06

*Race is a categorical variable, which has five race groups. There are five values for the regression coefficient and HR, one for each race group, which are not shown, because race was not statistically significant in univariate Cox regression analysis.

Table 3.4. Univariate Cox regression analyses for death and/or liver transplant prediction by BA Indices.

BA (μM)/ BA indices (%)	B-value (Regression Coefficient)	P-value	Hazard ratio (HR): Exp (B)		
			1 unit change	10% change	20% change
Total BA	0.001	0.202	1.00	1.01	1.01
Total LCA	0.008	0.694	1.01	1.00	1.00
Total UDCA	0.002	0.185	1.00	1.00	1.01
Total CDCA	0.003	0.112	1.00	1.01	1.01
Total DCA	-0.007	0.569	0.99	1.00	0.99
Total HDCA	-1.326	0.714	0.27	1.00	1.00
Total MDCA	0.370	0.802	1.45	1.00	1.00
Total CA	0.000	0.978	1.00	1.00	1.00
Total MCA	0.002	0.848	1.00	1.00	1.00
Total HCA	0.062	0.481	1.06	1.00	1.00
% LCA	-0.085	0.003	0.92	0.92	0.86
% UDCA	-0.017	0.064	0.98	0.96	0.93
% CDCA	0.031	0.000	1.03	1.12	1.25
% DCA	-0.074	0.000	0.93	0.89	0.79
% HDCA	-4.848	0.207	0.01	0.98	0.97
% MDCA	-0.366	0.431	0.69	0.99	0.97
% CA	0.051	0.004	1.05	1.03	1.06
% MCA	-0.012	0.524	0.99	0.99	0.98
% HCA	0.455	0.000	1.58	1.02	1.04
Total Unamidated	0.022	0.030	1.00	1.01	1.01
Total G-amidated	0.001	0.196	1.00	1.00	1.01
Total T-amidated	0.002	0.745	1.00	1.00	1.00
% Amidation	0.016	0.315	1.02	1.15	1.33
% G-amidation	-0.015	0.093	0.99	0.89	0.79
% T-amidation	0.030	0.002	1.03	1.03	1.07
Total Unsulfated	0.007	0.283	1.01	1.00	1.01
Total Sulfated	0.001	0.205	1.00	1.00	1.01
% Sulfation	-0.002	0.890	1.00	0.98	0.97
Total Mono-OH	0.008	0.694	1.01	1.00	1.00
Total Di-OH	0.001	0.145	1.00	1.00	1.01
Total Tri-OH	0.001	0.902	1.00	1.00	1.00
% Mono-OH	-0.085	0.003	0.92	0.92	0.86
% Di-OH	0.015	0.168	1.02	1.12	1.25
% Tri-OH	0.015	0.190	1.02	1.02	1.04
Total 12α-OH	-0.003	0.643	1.00	1.00	0.99
Total non-12α-OH	0.001	0.104	1.00	1.00	1.01
12α-OH/ non12α-OH	-2.412	0.004	0.09	0.92	0.85
CA/ CDCA	-0.478	0.458	0.62	0.99	0.98
% 12α-OH	-0.043	0.001	0.96	0.91	0.82
% non-12α-OH	0.043	0.001	1.04	1.39	1.94
Total Primary	0.002	0.216	1.00	1.01	1.01
Total Secondary	0.001	0.307	1.00	1.00	1.01
Primary/ Secondary	0.035	0.001	1.04	1.01	1.02
% Primary	0.033	0.000	1.03	1.17	1.38
% Secondary	-0.033	0.000	0.97	0.84	0.71
HI	0.408	0.508	1.50	1.00	0.99

Table 3.5. Univariate Cox regression analyses for death and/or liver transplant prediction by non-BA parameters and demographics.

Demographics and Non-BA parameters	B-value (Regression Coefficient)	P-value	Hazard ratio (HR): Exp (B)		
			1 unit change	10% change	20% change
Gender	0.982	0.002	2.67	-	-
Age (year)	0.017	0.151	1.02	1.09	1.19
BMI	-0.029	0.174	0.97	0.91	0.84
Race	*	0.806	*	*	*
Creatinine (mg/dL)	0.031	0.650	1.03	1.00	1.01
Albumin (g/dL)	-1.277	0.000	0.28	0.63	0.39
INR	0.838	0.000	2.31	1.10	1.21
Protime (sec)	0.071	0.000	1.07	1.08	1.17
AST (U/L)	0.001	0.416	1.00	1.01	1.01
ALT (U/L)	-0.003	0.254	1.00	0.98	0.97
Bilirubin (mg/dL)	0.100	0.002	1.11	1.01	1.03
AST/ALT	1.098	0.000	3.00	1.14	1.30
MELD	0.113	0.000	1.12	1.11	1.24
APRI	0.208	0.000	1.23	1.02	1.04

*Race is a categorical variable, which has five race groups. There are five values for the Regression Coefficient and HR, one for each race group, which are not shown, because race was not statistically significant in univariate Cox regression analysis.

Table 3.6. Multivariate Cox regression analysis for death prediction.

(a) The BA score model (BAS)

BA indices (μM)	B-value (Regression Coefficient)	Standard Error	P-value	Hazard ratio (HR): Exp (B)		
				1 unit change	10% change	20% change
% CDCA	0.039	0.010	0.000	1.040	1.159	1.344
% Tri-OH	0.052	0.016	0.001	1.053	1.069	1.142

*Using the regression coefficients from this table, the BA score (BAS) equation was:

$$\text{BA score} = 0.039 \times \%\,\text{CDCA} + 0.052 \times \%\,\text{Tri OH}$$

(b) The non-BA score model (non-BAS)

Non-BA parameters	B-value (Regression Coefficient)	Standard Error	P-value	Hazard ratio (HR): Exp (B)		
				1 unit change	10% change	20% change
AST/ALT	1.236	0.303	0.000	3.442	1.165	1.357

*Using the regression coefficients from this table, the non-BA score equation was:

$$\text{non BA score} = 1.236 \times \text{AST/ALT}$$

Table 3.7. Multivariate Cox regression analysis for death and/or liver transplant prediction.

(a)The BA score model (BAS)

BA indices (μM)	B-value (Regression Coefficient)	Standard Error	P-value	Hazard ratio (HR): Exp (B)		
				1 unit change	10% change	20% change
% Primary	0.021	0.007	0.003	1.021	1.107	1.226
% DCA	-0.049	0.020	0.013	0.952	0.926	0.857

*Using the regression coefficients from this table, the BA score equation was:

$$\text{BA score} = \ 0.021 \ \times \% \ \text{Primary} - \ 0.049 \ \times \% \ \text{DCA}$$

(b)The non-BA score model (non-BAS)

Non-BA parameters	B-value (Regression Coefficient)	Standard Error	P-value	Hazard ratio (HR): Exp (B)		
				1 unit change	10% change	20% change
Albumin (g/dL)	-1.277	0.222	0.000	0.279	0.627	0.393

*Using the regression coefficients from this table, the non-BA score equation was:

$$\text{non BA score} = \ -1.277 \ \times \text{Albumin} \ (\frac{\text{g}}{\text{dL}})$$

Table 3.8. Bootstrapping validation.

(a) For death prediction by BAS and non-BAS models

Variables	Regression Coefficient	Bias	Standard Error	P-value	95% Confidence Interval	
					Lower	Upper
BAS						
% CDCA	0.039	0.001	0.009	0.001	0.023	0.059
% Tri-OH	0.052	0.001	0.019	0.002	0.016	0.089
non-BAS						
AST/ALT	1.236	0.026	0.342	0.001	0.606	1.992

(b) For death and/or liver transplant prediction by BAS and non-BAS models

Variables	Regression Coefficient	Bias	Standard Error	P-value	95% Confidence Interval	
					Lower	Upper
BAS						
% Primary	0.021	0.001	0.008	0.006	0.008	0.039
% DCA	- 0.049	- 0.001	0.025	0.041	- 0.102	- 0.004
non-BAS						
Albumin (g/dL)	- 1.277	- 0.014	0.238	0.001	- 1.772	- 0.824

Table 3.9. ROC analysis of BAS, non-BAS, and MELD models

(a) Death prediction

Models	AUC (5-year)	AUC (3-year)	(Cutoff Value; Sensitivity, Specificity)
BAS	0.740	0.761	(2.71; 74, 68)
non-BAS	0.653	0.664	(1.72; 67, 66)
MELD	0.683	0.715	(10; 62, 64)

(b) Death and/or liver transplant prediction

Models	AUC (5-year)	AUC (3-year)	(Cutoff Value; Sensitivity, Specificity)
BAS	0.748	0.769	(0.76; 71, 72)
non-BAS	0.743	0.758	(-4.41; 78, 70)
MELD	0.763	0.789	(10; 70, 66)

Table 3.10. Estimated survival probability ($S_0(t)$) for death prediction.

(a) The BA score model (BAS)

t (months)	5	7	14	24	36	60	76
S_0 (t)	0.993	0.985	0.971	0.948	0.934	0.916	0.901

(b) The non-BA score model (non-BAS)

t (months)	5	7	14	24	36	60	76
S_0 (t)	0.989	0.978	0.958	0.924	0.902	0.876	0.855

Table 3.11. Kaplan-Meier analysis for survival.

Cutoff	Total N	N of events	Estimated mean (months)	Standard error	95% Confidence Interval
BAS					
Median cutoff of 2.19					
Low risk < 2.19	128	4	81.68	1.14	79.44-83.93
High risk > 2.19	129	23	70.72	2.5	65.81-75.62
ROC-optimum cutoff of 2.71					
Low risk < 2.71	162	7	80.8	1.19	78.48-83.13
High risk > 2.71	95	20	67.33	3.07	61.30-73.35
non-BAS					
Median cutoff of 1.44					
Low risk < 1.44	118	9	78.68	1.70	75.34-82.02
High risk > 1.44	139	18	73.97	2.21	69.64-78.29
ROC-optimum cutoff of 1.72					
Low risk < 1.72	145	9	79.70	1.39	76.98-82.42
High risk > 1.72	112	18	71.25	2.74	65.88-76.62
MELD					
Median cutoff of 11					
Low risk < 11	133	11	78.06	1.71	74.71-81.42
High risk > 11	124	16	73.91	2.35	69.29-78.52
ROC-optimum cutoff of 10					
Low risk < 10	173	11	79.49	1.32	76.91-82.07
High risk > 10	84	16	67.16	3.50	60.30-74.02

Table 3.12. Estimated liver transplant-free survival probability ($S_0(t)$) for death and/or liver transplant prediction.

(a) The BA score model (BAS)

t (months)	5	7	14	24	36	60	76
S_0 (t)	0.973	0.959	0.929	0.889	0.873	0.844	0.800

(b) The non-BA score model (non-BAS)

t (months)	5	7	14	24	36	60	76
S_0 (t)	0.972	0.954	0.918	0.870	0.851	0.817	0.776

Table 3.13. Kaplan-Meier analysis for liver transplant-free survival.

Cutoff	Total N	N of events	Estimated mean (months)	Standard error	95% Confidence Interval
BAS					
Median cutoff of 0.45					
Low risk < 0.45	128	9	79.22	1.56	76.16-82.28
High risk > 0.45	129	43	59.75	2.93	54.01-65.49
ROC-optimum cutoff of 0.76					
Low risk < 0.76	163	16	77.96	1.49	75.03-80.88
High risk > 0.76	94	36	55.72	3.64	48.04-62.31
non-BAS					
Median cutoff of -4.72					
Low risk < -4.72	120	9	78.79	1.69	75.49-82.09
High risk > -4.72	137	43	61.97	2.83	56.42-67.52
ROC-optimum cutoff of -4.41					
Low risk < -4.41	138	11	78.42	1.63	75.23-81.62
High risk > -4.41	119	41	59.94	3.10	53.86-66.02
MELD					
Median cutoff of 9					
Low risk < 9	105	11	77.67	1.84	74.05-81.28
High risk > 9	152	42	63.98	2.69	58.71-69.24
ROC-optimum cutoff of 10					
Low risk < 10	121	16	75.39	2.03	71.41-79.36
High risk > 10	136	36	64.62	2.81	59.12-70.11

Appendix

Table A. Developing other survival models for death prediction.

Other Models	Cox (P-value)	Bootstrapping (P-value)
Mixed BA and non-BA		
%CDCA	0.006	0.004
% G-amidation	0.015	0.018
AST/ALT	0.031	0.072
MELD variables with coefficients from our data set		
Creatinine	0.677	NA
INR	0.150	NA
Bilirubin	0.614	NA
Serum Na	0.058	NA
Original MELD modified with BA variables		
%CDCA	0.008	NA
%Tri-OH	0.129	NA
MELD	0.271	NA
Original MELD modified with non-BA variables		
AST/ALT	0.016	NA
MELD	0.253	NA
Original MELD modified with BA and non-BA variables		
%CDCA	0.017	NA
% G-amidation	0.029	NA
AST/ALT	0.040	NA
MELD	0.963	NA

NA: Not applicable. Bootstrapping was not performed because P-values of model parameters were not significant (P-value > 0.05)

Table B. Developing other liver transplant-free survival models for death and/or liver transplant prediction.

Other Models	Cox (P-value)	Bootstrapping (P-value)
Mixed BA and non-BA		
% 12α-OH	0.030	0.060
Albumin	0.011	0.026
Gender	0.020	0.037
MELD	0.004	0.010
MELD variables with coefficients from our data set		
Creatinine	0.588	NA
INR	0.015	NA
Bilirubin	0.243	NA
Serum Na	0.088	NA
Original MELD modified with BA variables		
%DCA	0.074	NA
%primary BA	0.358	NA
MELD	0.005	NA
Original MELD modified with non-BA variables		
Albumin	0.002	0.003
MELD	0.018	0.068
Original MELD modified with BA and non-BA variables		
% 12α-OH	0.030	0.060
Albumin	0.011	0.026
Gender	0.020	0.037
MELD	0.004	0.010

NA: Not applicable. Bootstrapping was not performed because P-values of model parameters were not significant (P-value > 0.05).

3.7. References

1. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases II. Signature profiles in patients.* Toxicol Sci, 2015. **143**(2): p. 308-18.
2. Kochanek, K.D., et al., *Deaths: final data for 2017.* 2019.
3. Bolondi, G., et al., *Predictive factors of short term outcome after liver transplantation: A review.* World J Gastroenterol, 2016. **22**(26): p. 5936-49.
4. Ramaiah, S.K., *A toxicologist guide to the diagnostic interpretation of hepatic biochemical parameters.* Food Chem Toxicol, 2007. **45**(9): p. 1551-7.
5. Ozer, J., et al., *The current state of serum biomarkers of hepatotoxicity.* Toxicology, 2008. **245**(3): p. 194-205.
6. Boin, I.F., et al., *Liver transplant recipients mortality on the waiting list: long-term comparison to Child-Pugh classification and MELD.* Transplant Proc, 2004. **36**(4): p. 920-2.
7. Kamath, P.S., et al., *A model to predict survival in patients with end-stage liver disease.* Hepatology, 2001. **33**(2): p. 464-70.
8. Wiesner, R., et al., *Model for end-stage liver disease (MELD) and allocation of donor livers.* Gastroenterology, 2003. **124**(1): p. 91-6.
9. Kamath, P.S., W.R. Kim, and G. Advanced Liver Disease Study, *The model for end-stage liver disease (MELD).* Hepatology, 2007. **45**(3): p. 797-805.
10. Lau, T. and J. Ahmad, *Clinical applications of the Model for End-Stage Liver Disease (MELD) in hepatic medicine.* Hepat Med, 2013. **5**: p. 1-10.
11. Cholongitas, E., et al., *MELD is not enough--enough of MELD?* J Hepatol, 2005. **42**(4): p. 475-7; author reply 478-9.
12. Freeman, R.B., *MELD: the holy grail of organ allocation?* J Hepatol, 2005. **42**(1): p. 16-20.
13. Alnouti, Y., *Bile Acid sulfation: a pathway of bile acid elimination and detoxification.* Toxicol Sci, 2009. **108**(2): p. 225-46.
14. Makino, I., et al., *Sulfated and nonsulfated bile acids in urine, serum, and bile of patients with hepatobiliary diseases.* Gastroenterology, 1975. **68**(3): p. 545-53.
15. Summerfield, J.A., et al., *Evidence for renal control of urinary excretion of bile acids and bile acid sulphates in the cholestatic syndrome.* Clin Sci Mol Med, 1977. **52**(1): p. 51-65.
16. Takikawa, H., T. Beppu, and Y. Seyama, *Urinary concentrations of bile acid glucuronides and sulfates in hepatobiliary diseases.* Gastroenterol Jpn, 1984. **19**(2): p. 104-9.
17. van Berge Henegouwen, G.P., et al., *Sulphated and unsulphated bile acids in serum, bile, and urine of patients with cholestasis.* Gut, 1976. **17**(11): p. 861-9.
18. Bathena, S.P., et al., *Urinary bile acids as biomarkers for liver diseases I. Stability of the baseline profile in healthy subjects.* Toxicol Sci, 2015. **143**(2): p. 296-307.
19. Thakare, R., et al., *Species differences in bile acids I. Plasma and urine bile acid composition.* J Appl Toxicol, 2018. **38**(10): p. 1323-1335.
20. Bathena, S.P., et al., *The profile of bile acids and their sulfate metabolites in human urine and serum.* J Chromatogr B Analyt Technol Biomed Life Sci, 2013. **942-943**: p. 53-62.
21. Huang, J., et al., *Simultaneous characterization of bile acids and their sulfate metabolites in mouse liver, plasma, bile, and urine using LC-MS/MS.* J Pharm Biomed Anal, 2011. **55**(5): p. 1111-9.
22. Heuman, D.M., *Quantitative estimation of the hydrophilic-hydrophobic balance of mixed bile salt solutions.* J Lipid Res, 1989. **30**(5): p. 719-30.

23. Kim, W.R., et al., *Hyponatremia and mortality among patients on the liver-transplant waiting list.* N Engl J Med, 2008. **359**(10): p. 1018-26.
24. Malinchoc, M., et al., *A model to predict poor survival in patients undergoing transjugular intrahepatic portosystemic shunts.* Hepatology, 2000. **31**(4): p. 864-71.
25. Blackstone, E.H., *Breaking down barriers: helpful breakthrough statistical methods you need to understand better.* J Thorac Cardiovasc Surg, 2001. **122**(3): p. 430-9.
26. Hesterberg, T.C., *What Teachers Should Know About the Bootstrap: Resampling in the Undergraduate Statistics Curriculum.* Am Stat, 2015. **69**(4): p. 371-386.
27. Rousselet, G.A. and R.R. Wilcox, *Reaction times and other skewed distributions: problems with the mean and the median.* Meta-Psychology, 2019.
28. Efron, B. and R. Tibshirani, *Improvements on cross-validation: the 632+ bootstrap method.* Journal of the American Statistical Association, 1997. **92**(438): p. 548-560.
29. Schoenfeld, D., *Partial residuals for the proportional hazards regression model.* Biometrika, 1982. **69**(1): p. 239-241.
30. Garcia-Tsao, G., *The Child-Turcotte Classification: From Gestalt to Sophisticated Statistics and Back.* Dig Dis Sci, 2016. **61**(11): p. 3102-3104.
31. Huang, W.M., et al., *Intrahepatic cholestasis of pregnancy: detection with urinary bile acid assays.* J Perinat Med, 2007. **35**(6): p. 486-91.
32. Simko, V., S. Michael, and R.E. Kelley, *Predictive value of random sample urine bile acids corrected by creatinine in liver disease.* Hepatology, 1987. **7**(1): p. 115-21.
33. Galman, C., B. Angelin, and M. Rudling, *Bile acid synthesis in humans has a rapid diurnal variation that is asynchronous with cholesterol synthesis.* Gastroenterology, 2005. **129**(5): p. 1445-53.
34. Steiner, C., et al., *Bile acid metabolites in serum: intraindividual variation and associations with coronary heart disease, metabolic syndrome and diabetes mellitus.* PLoS One, 2011. **6**(11): p. e25006.
35. Trottier, J., et al., *Profile of serum bile acids in noncholestatic volunteers: gender-related differences in response to fenofibrate.* Clin Pharmacol Ther, 2011. **90**(2): p. 279-86.
36. Glicksman, C., et al., *Postprandial plasma bile acid responses in normal weight and obese subjects.* Ann Clin Biochem, 2010. **47**(Pt 5): p. 482-4.
37. Trottier, J., et al., *Metabolomic profiling of 17 bile acids in serum from patients with primary biliary cirrhosis and primary sclerosing cholangitis: a pilot study.* Dig Liver Dis, 2012. **44**(4): p. 303-10.
38. Trottier, J., et al., *Profiling circulating and urinary bile acids in patients with biliary obstruction before and after biliary stenting.* PLoS One, 2011. **6**(7): p. e22094.
39. Modica, S., et al., *Selective activation of nuclear bile acid receptor FXR in the intestine protects mice against cholestasis.* Gastroenterology, 2012. **142**(2): p. 355-65 e1-4.
40. Luo, L., et al., *Assessment of serum bile acid profiles as biomarkers of liver injury and liver disease in humans.* PLoS One, 2018. **13**(3): p. e0193824.
41. Hayashi, H., et al., *Transport by vesicles of glycine- and taurine-conjugated bile salts and taurolithocholate 3-sulfate: a comparison of human BSEP with rat Bsep.* Biochim Biophys Acta, 2005. **1738**(1-3): p. 54-62.
42. Mita, S., et al., *Inhibition of bile acid transport across Na+/taurocholate cotransporting polypeptide (SLC10A1) and bile salt export pump (ABCB 11)-*

coexpressing LLC-PK1 cells by cholestasis-inducing drugs. Drug Metab Dispos, 2006. **34**(9): p. 1575-81.
43. Fan, J., S. Upadhye, and A. Worster, *Understanding receiver operating characteristic (ROC) curves.* CJEM, 2006. **8**(1): p. 19-20.
44. Pines, J.M., et al., *Evidence-based emergency care: Diagnostic testing and clinical decision rules*. Vol. 83. 2013: John Wiley & Sons.
45. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. **143**(1): p. 29-36.
46. Wiesner, R.H., et al., *MELD and PELD: application of survival models to liver allocation.* Liver Transpl, 2001. **7**(7): p. 567-80.
47. Freeman, R.B., Jr., et al., *The new liver allocation system: moving toward evidence-based transplantation policy.* Liver Transpl, 2002. **8**(9): p. 851-8.