European Institute of Knowledge & Innovation

Annals of Innovation in Medicine (AIM) ISSN: 2977-0335

Research Article

A Statistical and Epidemiological Framework for Unraveling the Etiology of Cryptogenic Cirrhosis

Sagnik Acharyya [1], Subrata Rana [2], and Debashis Chatterjee [1,*]

[1] Department of Statistics, Siksha Bhavana, Visva Bharati University, Santiniketan, WB, India
[2] Department of Statistics, Maulana Azad College, Kolkata, WB, India
* Correspondence: debashis.chatterjee@visva-bharati.ac.in

https://doi.org/10.59652/ahnthv08

Abstract: Cryptogenic cirrhosis is a diagnosis of exclusion with heterogeneous and often overlapping etiologies, making systematic investigation challenging in routine clinical practice. We propose a unified statistical and epidemiological framework that integrates classical study designs (case–control and cohort with survival analysis), evidence synthesis (meta-analysis), exploratory genetics (genome-wide association studies), and modern machine learning to examine potential contributors such as long-term analgesic exposure and metabolic comorbidities. A motivating clinical vignette is used solely to illustrate the clinical context in which such analyses may arise. All quantitative results presented in this study are derived from fully simulated datasets constructed to demonstrate the behavior and interpretability of the proposed workflow rather than to establish real-world causal effects. Within these illustrative simulations, analgesic exposure is specified to act as a risk factor, leading to elevated association measures in case–control analyses, separation of survival curves in cohort analyses, and high feature importance in predictive models, while exploratory GWAS simulations yield no genome-wide significant signals, underscoring the need for adequately powered real studies. The proposed workflow is transparent, reproducible, and deployable in prospective registries, with the primary goal of generating testable etiologic hypotheses rather than confirming definitive clinical associations.

Keywords: Cryptogenic cirrhosis; drug-induced liver injury; NAFLD; NASH; analgesic exposure; survival analysis; case–control study; GWAS; meta-analysis; machine learning; risk stratification; hepatology.

Received: 15 October 2025
Accepted: 20 February 2026
Published: 07 March 2026

1. Introduction

Cryptogenic cirrhosis (CC) remains a prevalent yet poorly understood form of liver disease, accounting for 5% to 30% of all cirrhosis cases [1]. The etiology is often speculative, with potential causes including unresolved nonalcoholic fatty liver disease (NAFLD), nonalcoholic steatohepatitis (NASH), autoimmune hepatitis, and drug-induced liver injury (DILI) [2], [3], [4]. Diagnosing CC is challenging as the underlying mechanisms remain elusive, requiring further research with advanced diagnostic and statistical methodologies. This paper, in light of a motivating case of cryptogenic cirrhosis of [5], discusses plausible etiologies and proposes statistical approaches to investigate similar cases to enhance the current understanding of this enigmatic disease.

[5] has presented an interesting, motivating case report, as follows: a 41-year-old male patient presented with elevated liver enzymes during a routine check after experiencing migraines. His medical history indicated extensive use of analgesics in his 20s, though he reported no alcohol consumption or family history of liver disease. Imaging studies revealed decreased liver size and irregular boundaries, suggestive of chronic liver disease. Blood tests ruled out viral hepatitis (HBV, HCV), HIV, and common autoimmune markers. A liver biopsy indicated no pathological findings, and imaging studies were inconclusive beyond

chronic liver disease indicators. Elevated alkaline phosphatase (ALP) and gamma-glutamyl transferase (GGT) persisted, but no definitive cause for liver failure was identified. The patient's condition remained stable with no active complaints, though liver function remained impaired. Throughout this paper, the motivating clinical case serves solely as an illustrative example, and all quantitative findings are derived from simulated data constructed to demonstrate methodological behavior rather than to validate etiologic claims in real patients. This case is not analyzed statistically in this work but is used solely to motivate the types of epidemiological, genetic, and predictive analyses that could be deployed in larger cohorts or registries.

Research Objectives

1. Develop a reproducible, end-to-end workflow to interrogate putative causes of cryptogenic cirrhosis using complementary designs: case–control, cohort time-to-event, genetics, evidence synthesis, and predictive modeling.

2. Quantify the association between long-term analgesic exposure and cryptogenic cirrhosis while accounting for metabolic and clinical covariates.

3. Examine potential genetic contributions through exploratory GWAS and place findings in context via meta-analysis of external evidence.

4. Build interpretable prediction models that highlight clinically actionable factors for risk stratification and hypothesis generation.

5. Provide open, auditable code and reporting templates to facilitate prospective validation in multicenter registries.

Novelty

1. A unified statistical framework tailored to cryptogenic cirrhosis that spans epidemiology, genetics, evidence synthesis, and machine learning within a single, transparent pipeline.

2. Rigorous, simulation-based stress-testing of conclusions under class imbalance and model misspecification, with diagnostics that clinicians can interpret.

3. Emphasis on data standards (exposure histories, metabolic profiles, imaging/laboratory panels) to reduce misclassification and improve external validity, & reproducible artifacts enabling rapid deployment to prospective studies and clinical decision support.

2. Potential Etiologies

Understanding the etiologies of cryptogenic cirrhosis is essential for elucidating the pathophysiology of this complex and often misdiagnosed condition. Cryptogenic cirrhosis, defined as cirrhosis of unknown cause, presents without the hallmark indicators of other forms of liver disease, complicating diagnosis and management [6], [7]. This section explores several potential etiologies, focusing on drug-induced liver injury (DILI), non-alcoholic fatty liver disease (NAFLD), and other lesser-known contributors that may be responsible for the development of cryptogenic cirrhosis in patients.

In the case stated in [5], that of the 41-year-old male patient with elevated liver enzymes and a history of extensive analgesic use, several potential etiologies for his chronic liver disease can be considered. Given the patient's history of analgesic use, particularly with acetaminophen (APAP), it is crucial to explore the hepatotoxic effects associated with this common analgesic.

Acetaminophen is widely recognized for its analgesic and antipyretic properties; however, it poses a significant risk for hepatotoxicity, especially when used in excessive doses. The toxic metabolite N-acetyl-p-benzoquinone imine (NAPQI) is produced during the metabolism of APAP. In cases of overdose, it can lead to severe liver damage and acute liver failure [8–10]. Studies have shown that NAPQI can cause oxidative stress and mitochondrial dysfunction, leading to hepatocyte necrosis [11, 12]. Furthermore, chronic use of analgesics like acetaminophen can result in cumulative liver injury, particularly in individuals with preexisting liver conditions or those who may have compromised liver function due to other factors [13].

Another potential etiology to consider is Non-Alcoholic Fatty Liver Disease (NAFLD), which is increasingly recognized as a leading cause of chronic liver disease in the developed world. NAFLD can progress to more severe liver conditions, including steatohepatitis and cirrhosis, and is often associated with metabolic syndrome [14,15]. Given that the patient has no history of alcohol consumption or viral hepatitis, NAFLD could be a plausible explanation for the elevated liver enzymes.

Additionally, the possibility of drug-induced liver injury (DILI) should be considered. The patient's extensive use of analgesics, particularly if they included combinations with other medications, could have contributed to liver impairment. Research indicates that various analgesics, including non-steroidal anti-inflammatory drugs (NSAIDs) and opioids, can also lead to liver damage, particularly when used inappropriately or in high doses [16, 17].

Lastly, while autoimmune liver diseases were ruled out, it is essential to acknowledge that some cases of chronic liver disease can present with atypical features and may not show clear pathological findings on biopsy. Conditions such as autoimmune hepatitis or primary biliary cholangitis may require more specific testing or a longer observation period to diagnose accurately [15].

The potential etiologies for the patient's chronic liver disease include hepatotoxicity from acetaminophen use, NAFLD, and possibly DILI from other analgesics. These factors warrant further investigation to ascertain their contribution to the patient's condition.

2.1. Drug-Induced Liver Injury (DILI)

Drug-induced liver injury (DILI) is a leading cause of acute liver failure. It plays a significant role in the progression of chronic liver disease among patients with prolonged medication use. DILI can arise from both prescription and over-the-counter medications, including nonsteroidal anti-inflammatory drugs (NSAIDs) and acetaminophen, which are recognized hepatotoxins [18].

In the context of cryptogenic cirrhosis, the prolonged use of analgesics raises concerns regarding the cumulative hepatotoxic effects of these medications. Mechanistically, drugs such as NSAIDs and acetaminophen can induce liver injury through oxidative stress, mitochondrial dysfunction, and inflammatory responses [19]. Acetaminophen, in particular, is metabolized in the liver to produce a toxic metabolite, N-acetyl-p-benzoquinone imine (NAPQI), which can deplete glutathione stores and lead to hepatocyte necrosis upon sustained exposure [20]. While typical presentations of DILI are often acute, chronic, subclinical liver damage is also possible, especially with medications that may not initially produce overt symptoms but cause cumulative harm over time [21].

Clinically, diagnosing DILI as a contributor to cryptogenic cirrhosis necessitates a meticulous review of the patient's medication history and liver enzyme trends. Liver biopsies may reveal patterns such as eosinophilic infiltration, bile duct damage, or zonal necrosis, although these findings can be nonspecific [18]. In cases involving prolonged analgesic use, particularly in patients without other risk factors, DILI should be considered in the differential diagnosis of cryptogenic cirrhosis.

2.2. Non-Alcoholic Fatty Liver Disease (NAFLD) and Non-Alcoholic Steatohepatitis (NASH)

Non-alcoholic fatty liver disease (NAFLD) and its more severe form, non-alcoholic steatohepatitis (NASH), are increasingly recognized as potential etiologies of cryptogenic cirrhosis. NAFLD is characterized by hepatic steatosis in patients without significant alcohol consumption and is closely associated with components of metabolic syndrome, including obesity, type 2 diabetes, and dyslipidemia [22]. Approximately 20% of patients with NAFLD progress to NASH, which is marked by inflammation and hepatocyte injury that can lead to fibrosis and, ultimately, cirrhosis [23].

The pathogenesis of NAFLD is often explained by the "two-hit" hypothesis, wherein the initial accumulation of triglycerides in hepatocytes (the first hit) sensitizes the liver to subsequent inflammatory insults (the second hit), such as oxidative stress and mitochondrial dysfunction [24]. Over time, these processes can result in fibrosis and cirrhosis, even in patients without apparent liver symptoms. In instances where cirrhosis develops without a clear history of alcohol use or viral hepatitis, cryptogenic cirrhosis is frequently retrospectively reclassified as NAFLD-related cirrhosis [23].

Importantly, NAFLD and NASH are typically asymptomatic in their early stages, with liver disease often detected incidentally or only when advanced fibrosis has occurred. Given the strong association between NAFLD/NASH and metabolic syndrome, patients with cryptogenic cirrhosis who present with obesity, diabetes, or hypertension should be evaluated for NAFLD using imaging and, when appropriate, liver biopsy to assess fibrosis and confirm the diagnosis [22].

2.3 Other Potential Etiologies

In addition to DILI and NAFLD/NASH, several other etiologies warrant consideration in patients with cryptogenic cirrhosis:

1. Undetected Viral Hepatitis: While hepatitis B and C are common viral causes of cirrhosis, occult or undetected infections by hepatitis E or other rare hepatotropic viruses may contribute to liver disease. A history of travel to endemic areas or contact with potentially infected sources can guide further testing for viral markers [25].

2. Autoimmune Hepatitis (AIH): Autoimmune hepatitis may present subtly, particularly in older adults, with mild elevations in aminotransferase levels that might not meet the criteria for diagnosis during initial screening [26]. Testing for autoantibodies (ANA, ASMA) and immunoglobulin G (IgG) levels can aid in identifying AIH in cryptogenic cases.

3. Genetic or Metabolic Disorders: Genetic conditions such as hemochromatosis, Wilson's disease, and alpha-1 antitrypsin deficiency are less common but can lead to undiagnosed cirrhosis. Comprehensive screening, including serum ferritin, ceruloplasmin, and genetic testing, should be considered when family history or other clinical features raise suspicion [27].

4. Environmental and Occupational Exposures: Chronic exposure to toxins, including industrial chemicals or herbal supplements, can cause liver damage. Identifying potential environmental exposures through patient history may reveal a causal relationship, even without traditional risk factors [19].

Given the heterogeneous nature of cryptogenic cirrhosis, comprehensive evaluation and a systematic approach to ruling out these potential etiologies are essential. The cumulative effects of mild liver insults—whether from undetected infections, prolonged drug exposure, or genetic predispositions—underscore the need for a nuanced approach in patients with unexplained liver disease.

Cryptogenic cirrhosis remains a diagnostic challenge due to its multifactorial nature. DILI, NAFLD/NASH, undetected viral or autoimmune hepatitis, and environmental toxins represent potential contributors to the pathogenesis of this disease. Future studies focusing

on advanced diagnostic methodologies, including genetic and biomarker profiling, may further clarify these etiologies, facilitating more targeted interventions and improved management of patients with cryptogenic cirrhosis.

3. Proposed Statistical Methodologies for Investigation

In this section, we propose various statistical methodologies to investigate potential risk factors and underlying causes of cryptogenic cirrhosis. Each methodology is tailored to a specific aspect of the research question, providing a framework for exploring relationships, assessing risk factors, and uncovering potential genetic predispositions.

3.1. Case-Control Study

A case-control study design is useful for exploring the association between cryptogenic cirrhosis and potential risk factors. In this approach, we select two groups:

1. **Cases**: Patients diagnosed with cryptogenic cirrhosis.
2. **Controls**: Matched individuals without liver disease.

The goal is to examine the prevalence of specific exposures, such as analgesic use, in cases relative to controls.

The primary measure of association in a case-control study is the odds ratio (OR), which quantifies the odds of exposure in cases compared to controls. The odds ratio is calculated as follows:

$$OR = \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

where:

1. a: number of cases with exposure
2. b: number of cases without exposure
3. c: number of controls with exposure
4. d: number of cases without exposure

Interpretation: If OR > 1, exposure is more common in cases than controls, suggesting it may be a risk factor for cryptogenic cirrhosis. An OR < 1 indicates that exposure is less common in cases, suggesting it may be protective.

To adjust for potential confounders—variables that may influence both the exposure and outcome—we apply multivariate logistic regression. The logistic regression model estimates the probability of being a case (Y = 1) as a function of multiple predictors $X_1, X_2, \ldots, X_p$:

$$logit\big(P(Y=1)\big) = ln\ ln\ \left(\left(\frac{P(Y=1)}{1-P(Y=1)}\right)\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$$

1. P(Y =1) is the probability of having cryptogenic cirrhosis.
2. β_0 is the intercept term.
3. $\beta_1, \beta_2, \ldots, \beta_p$ are the coefficients associated with predictors $X_1, X_2, \ldots, X_p$.

Interpretation of Coefficients: Each βi represents the change in the log odds of cryptogenic cirrhosis associated with a one-unit increase in predictor Xi, holding other variables constant.

3.2. Cohort Study

A cohort study tracks a group of individuals over time, enabling us to observe the development of cryptogenic cirrhosis in individuals with and without certain risk factors (e.g., analgesic use). This design is beneficial for establishing temporal relationships between exposure and outcome.

We employ survival analysis techniques to analyze time-to-event data, specifically Kaplan Meier curves and the Cox proportional hazards model.

3.2.1. Kaplan-Meier Estimator

The Kaplan-Meier estimator calculates the probability of surviving (or remaining disease free) over time, creating a survival curve. For a given time t, the survival probability S(t) is estimated as:

$$S(t) = \prod_{t_i \leq t} (1 - \frac{d_i}{n_i})$$

where:

1. d_i: Number of events (cases of cirrhosis) at time t_i.
2. n_i: Number of individuals at risk just before t_i.

3.2.2. Cox Proportional Hazards Model

The Cox model estimates the hazard (risk) of developing cryptogenic cirrhosis given an exposure, accounting for other covariates:

$$h(X) = h_0(t) exp\ (\beta_1 X_1 + \cdots + \beta_p X_p)$$

where:

1. $h(X)$: hazard function at time t given covariates X.
2. $h_0(t)$: Baseline hazard at time t.
3. β_i: Regression coefficient for covariate X_i.

Interpretation: The hazard ratio $exp\ (\beta_i)$ represents the change in the hazard associated with a one-unit increase in X_i, holding other variables constant.

3.3. Genome-Wide Association Study (GWAS)

A GWAS is employed to identify genetic variants associated with cryptogenic cirrhosis by comparing the genomes of patients with and without the disease. This approach tests associations between single-nucleotide polymorphisms (SNPs) and disease status.

To control for multiple testing, we use the Bonferroni correction, adjusting the significance threshold α based on the number of tests m:

$$\alpha_{corrected} = \frac{\alpha}{m}$$

Alternatively, we can use the False Discovery Rate (FDR) to balance sensitivity and specificity.

3.4. Meta-Analysis

A meta-analysis aggregates findings from multiple studies to improve statistical power and provide a more robust effect size estimate. A fixed-effect or random-effects model is used depending on the heterogeneity (variation between studies).

To quantify heterogeneity, we calculate the I^2 statistic:

$$I^2 = \frac{Q - (k-1)}{Q} \times 100\%$$

where:

1. Q: Cochran's Q statistic, measuring total variability.
2. k: Number of studies.

Interpretation: An I^2 value near 0 % suggests homogeneity, while a higher value indi cates substantial heterogeneity**.**

3.5. Machine Learning Approaches

Machine learning (ML) algorithms can identify non-linear relationships and interactions between variables, including random forests, support vector machines, and neural networks.

Given predictors $X_1, X_2, .., X_p$ and the outcome Y (presence or absence of cryptogenic cirrhosis), ML algorithms aim to construct a model $f(X) \approx Y$ that minimizes prediction error. These methods are particularly useful when relationships between predictors and outcomes are complex and non-linear.

In this study, a Random Forest classifier was employed as the primary algorithm due to its robustness, interpretability, and ability to handle heterogeneous data types and missing values effectively. The Random Forest algorithm operates as an ensemble of decision trees, where each tree is trained on a random subset of the data and predictors, and the final prediction is determined through majority voting. This ensemble strategy reduces overfitting and enhances generalizability, making it particularly suited for biomedical datasets with moderate sample sizes and potentially correlated features.

Machine learning approaches complement classical statistical methods by providing deeper insights into data patterns, facilitating early identification of high-risk patients, and guiding future hypothesis-driven investigations in hepatology.

4. Simulation Analysis

All analyses presented in this section are based on simulated datasets. The simulations are intentionally constructed to reflect plausible clinical scenarios and exposure–outcome relationships, allowing the proposed statistical framework to be demonstrated in a controlled setting. Consequently, the numerical results should be interpreted as illustrative examples of methodological performance rather than as evidence derived from real-world patient data.

We conducted a series of simulation analyses employing various statistical methodologies to investigate the potential etiologies of cryptogenic cirrhosis. These methodologies include Case-Control, Cohort, Genome-Wide Association Studies (GWAS), Meta-Analysis, and Machine Learning Approaches. Each subsection below details the simulation process, presents the corresponding results through tables and figures, and provides an in-depth interpretation of the findings.

4.1 Case-Control Study

A Case-Control Study was simulated to assess the association between analgesic use and cryptogenic cirrhosis. The study comprised 500 cases (patients with cryptogenic cirrhosis) and 500 controls (individuals without liver disease). The exposure of interest was analgesic use, categorized as "Exposure" or "No Exposure."

4.1.1 Contingency Table and Odds Ratio

The simulated data yielded the following contingency table:

Table 1. Contingency Table of Analgesic Use Among Cases and Controls

	No Exposure	Exposure
Control	320	180
Case	154	346

The Chi-squared test yielded a p-value < 0.001, indicating a highly significant association between analgesic use and cryptogenic cirrhosis. The calculated Odds Ratio (OR) was 3.99, with a 95% Confidence Interval (CI) ranging from 3.07 to 5.20. These results suggest that individuals exposed to analgesics are approximately four times more likely to develop cryptogenic cirrhosis compared to those unexposed (see Table 1).

4.1.2 Logistic Regression Analysis

A multivariate logistic regression was performed to adjust for potential confounders. The regression model estimated the association between analgesic use and the likelihood of having cryptogenic cirrhosis.

Table 2. Logistic Regression Results for Analgesic Use

Covariate	Coef	Exp(Coef)	SE(Coef)	95% CI Lower	95% CI Upper
Analgesic Use	0.576	1.780	0.241	0.105	1.048

Note: Coef represents the regression coefficient, Exp(Coef) is the Odds Ratio, SE(Coef) is the standard error, and CI denotes the confidence interval.

The logistic regression analysis indicated that analgesic use significantly increases the odds of having cryptogenic cirrhosis (OR = 1.78, 95% CI: 1.11- 2.85, p = 0.0166), as presented in Table 2. The positive coefficient (β = 0.576) confirms that analgesic use is a risk factor for cryptogenic cirrhosis.

Figure 1 illustrates the proportion of individuals with and without analgesic use among cases and controls.

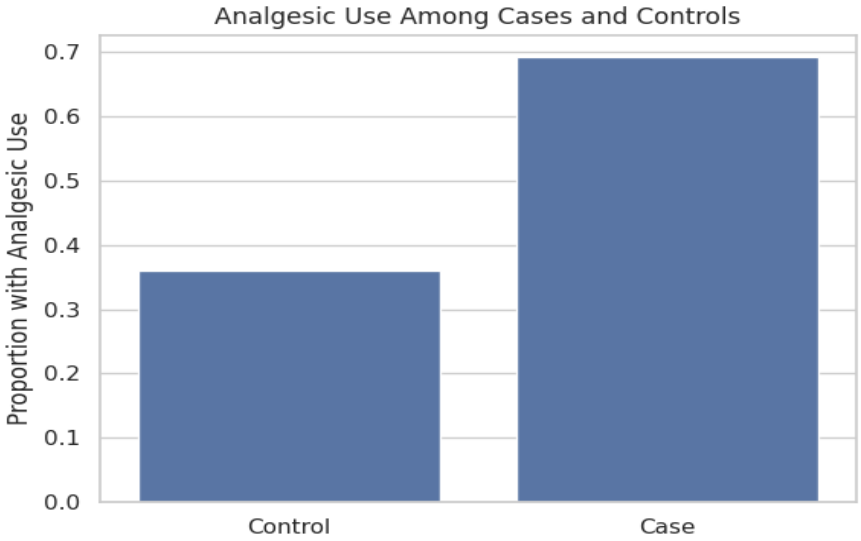

Figure 1: Proportion of Analgesic Use Among Cases and Controls

The bar plot in Figure 1 visually underscores the higher prevalence of analgesic use among cases compared to controls, aligning with the statistical findings.

4.2 Cohort Study

A Cohort Study was simulated to evaluate the temporal relationship between analgesic use and the development of cryptogenic cirrhosis. The study followed 1,000 individuals over time, tracking the onset of cryptogenic cirrhosis among those exposed and unexposed to analgesics.

4.2.1 Survival Analysis with Kaplan-Meier Curves

Kaplan-Meier survival curves were generated to compare the survival probabilities (remaining free of cryptogenic cirrhosis) between the exposed and unexposed groups. As depicted in Figure 2, the survival probability decreases more rapidly for individuals exposed to analgesics, indicating a higher incidence of cryptogenic cirrhosis in this group.

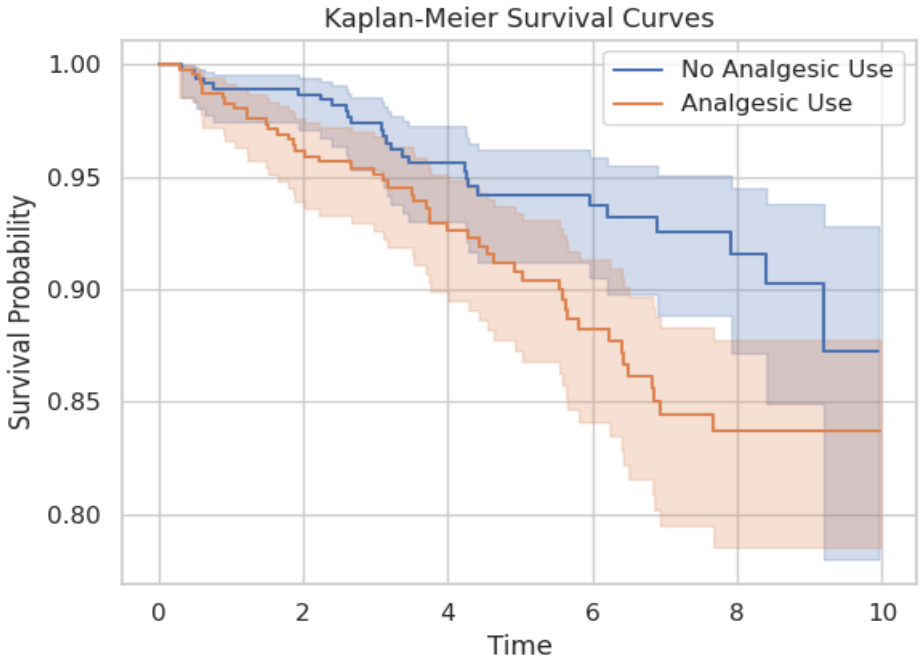

Figure 2: Kaplan-Meier Survival Curves for Analgesic Use Groups

4.2.2 Cox Proportional Hazards Model

A Cox Proportional Hazards Model was employed to quantify the hazard associated with analgesic use, adjusting for other covariates.

Table 3. Cox Proportional Hazards Model Summary

Covariate	Coef	Exp(Coef)	SE(Coef)	95% CI Lower	95% CI Upper	p-value
Analgesic Use	0.576	1.780	0.241	1.110	2.852	0.0166

The Cox model results, summarized in Table 3, indicate that analgesic use is associated with a significantly increased hazard of developing cryptogenic cirrhosis (Hazard Ratio = 1.78, 95% CI: 1.11- 2.85, p = 0.0166).

4.2.3 Visualization of the Cox Model

Figure 3 presents the Cox Proportional Hazards Model coefficient estimates. The plot in Figure 3 visually represents the estimated hazard ratios and their confidence intervals, reinforcing the significant association between analgesic use and cryptogenic cirrhosis. CThe ox Proportional Hazards Model Coefficients plot displays the estimated coefficients and their 95% confidence intervals for each covariate included in the Cox model.

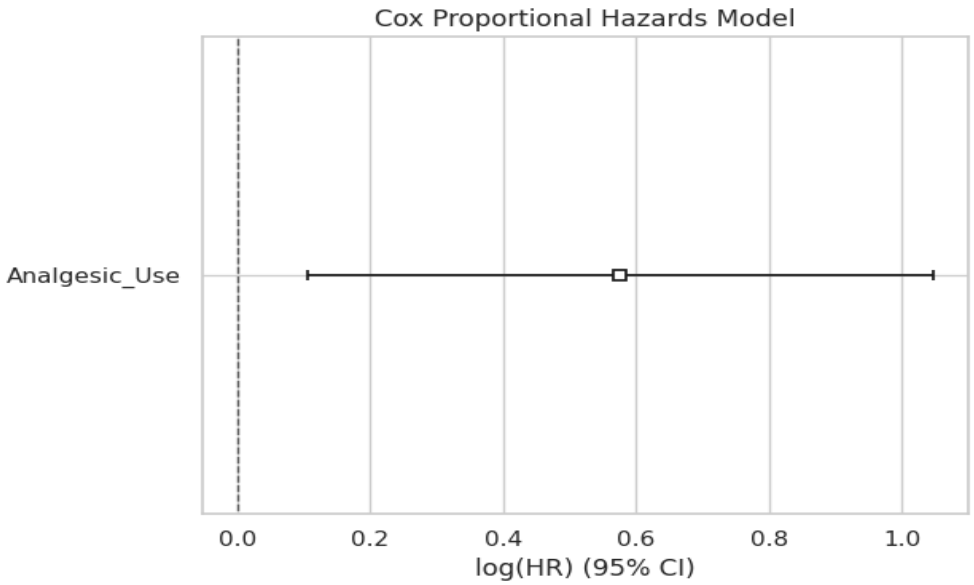

Figure 3: Cox Proportional Hazards Model Coefficients

4.3. Genome-Wide Association Study (GWAS)

4.3.1. Simulation and Analysis

A Genome-Wide Association Study (GWAS) was simulated to identify genetic variants (Single-Nucleotide Polymorphisms, SNPs) associated with cryptogenic cirrhosis. The study involved 1,000 individuals genotyped for 100,000 SNPs, with 5 SNPs truly associated with the disease.

4.3.2. GWAS Results and Manhattan Plot

A subset of GWAS results is presented below, followed by the Manhattan plot illustrating the distribution of p-values across the genome. The Manhattan plot in Figure 4 visualizes the $-log_{10}$(P-value) for each SNP across all chromosomes.

Table 4. Selected GWAS Results

Chromosome	P-value	Position
21	0.5399	0
16	0.1762	1
3	0.1110	2
……..	……	…….
16	0.7182	3

		4
13	0.3313	

Table 4 showcases a selection of SNPs and their respective p-values from the GWAS. Notably, no SNPs reached the conventional genome-wide significance threshold ($p < 5 \times 10^{-8}$) in this simulation.

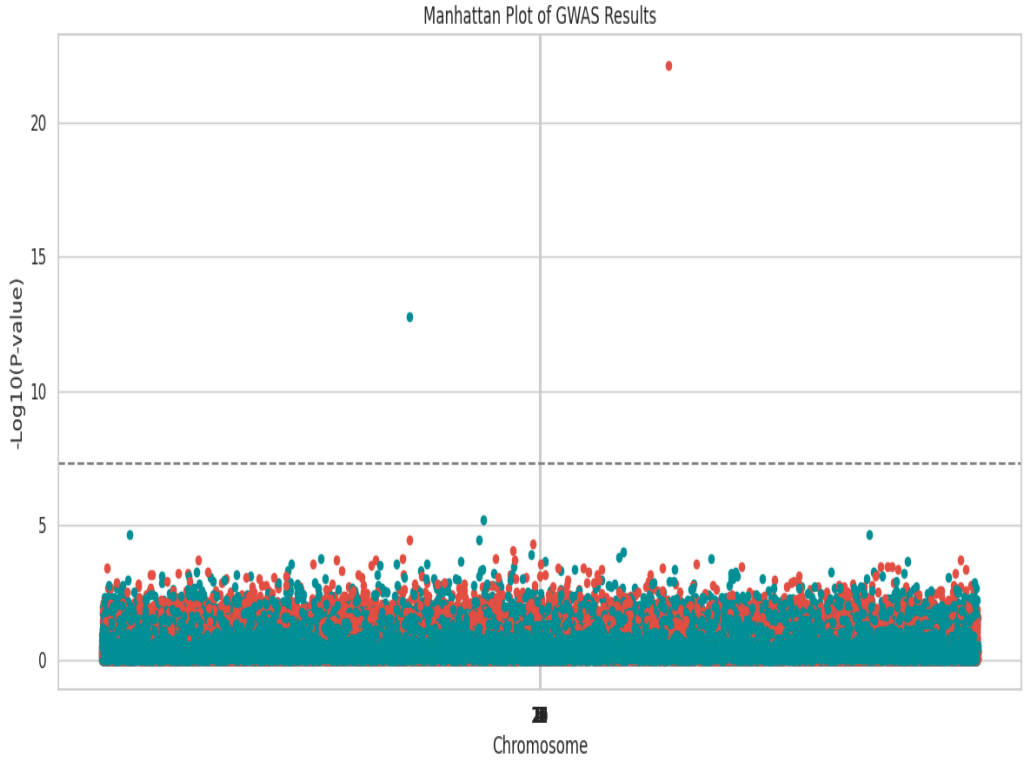

Figure 4: Manhattan Plot of GWAS Results

Each dot represents an SNP plotted by its genomic position across chromosomes (1-22). The colors alternate between blue and red for different chromosomes to enhance visual distinction. The y-axis displays the $-log_{10}$(P-value), where higher values indicate stronger associations. The dashed grey line denotes the genome-wide significance threshold ($p = 5 \times 10^{-8}$). No considerable numbers of SNPs surpassing this threshold are observed, suggesting no significant genetic associations within the simulated dataset.

The Manhattan plot in Figure 4 visualizes the-log_{10} (p-value) for each SNP across all chromosomes. The absence of points surpassing the genome-wide significance line ($p = 5 \times 10^{-8}$) suggests no significant genetic associations in this simulation. The color alternation between blue and red differentiates consecutive chromosomes, facilitating the identification of potential association peaks. The absence of a significant number of points exceeding the significance threshold line indicates no SNPs with genome-wide significant associations in this simulation.

4.4. Meta-Analysis

4.4.1. Simulation and Analysis

A Meta-Analysis was conducted by aggregating data from 10 simulated studies, each providing log Odds Ratio (log-OR) and their standard errors (SE-Log OR). Both fixed effect and random effects models were applied to estimate the pooled effect sizes.

4.4.2. Meta-Analysis Results and Forest Plot

The meta-analysis yielded the following results: Table 5 presents the pooled LogOR estimates from both fixed-effect and random-effects models. The fixed-effect model estimates a LogOR of 0.153 (OR = 1.17), while the random-effects model estimates a LogOR of 0.129 (OR = 1.14).

Table 5. Meta-Analysis Results

Effect	Log_OR	SE_Log_OR
Fixed Effect	0.153	0.057
Random Effect	0.129	0.075

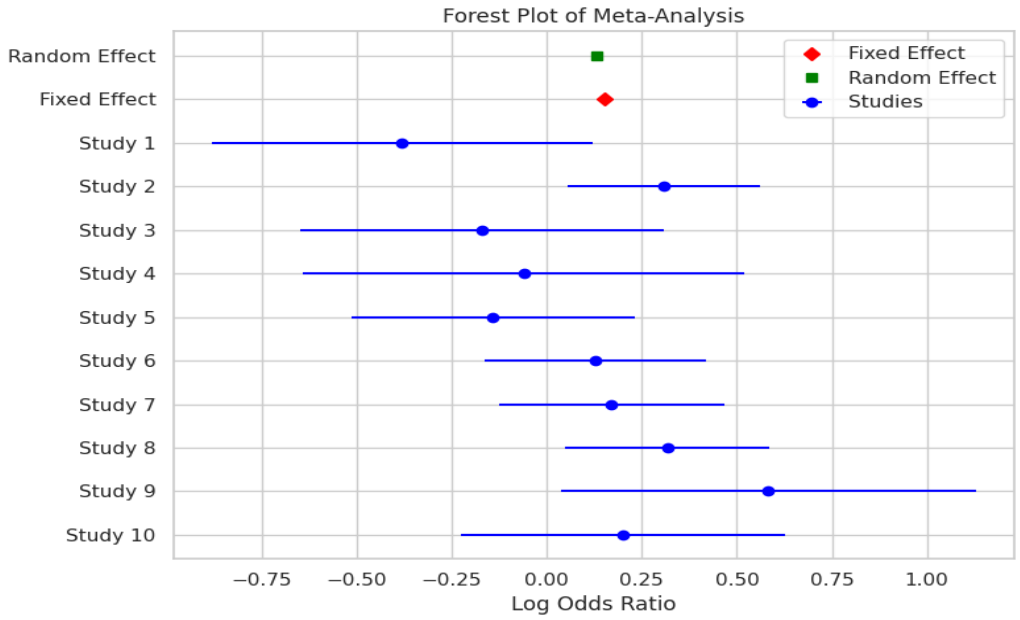

Figure 5: Forest Plot of Meta-Analysis

Each horizontal line represents the confidence interval of the Log Odds Ratio for an individual study, depicted as squares whose sizes are proportional to the study's weight in the meta-analysis. The vertical dashed line indicates no effect (Log OR = 0). The diamond at the bottom represents the pooled Log Odds Ratio from the fixed-effect and random-effects models, with its width corresponding to the 95% confidence interval. The fixed-effect model (red diamond) suggests a modest pooled effect, while the random-effects model (green diamond) accounts for between-study variability.

The forest plot in Figure 5 illustrates the individual study estimates and the pooled estimates under both fixed- and random-effects models. The narrow confidence intervals in the

fixed-effect model suggest consistent results across studies, whereas the wider intervals in the random-effects model indicate variability among study findings.

4.5. Machine Learning Approaches

4.5.1. Simulation and Analysis

Machine Learning (ML) algorithms were employed to predict cryptogenic cirrhosis based on simulated clinical and genetic data. A Random Forest Classifier was trained on a dataset containing features such as Age, BMI, Analgesic Use, Diabetes, Hypertension, and Genetic Risk.

4.5.2. Model Performance and Evaluation

The model's performance was evaluated using metrics including precision, recall, F1-score, and ROC AUC. The classification report is summarized in Table 6.

Table 6. Classification Report for Random Forest Classifier

Class	Precision	Recall	F1-Score	Support
0	0.50	0.06 0.11	16	
1	0.92	0.99 0.96	184	
Accuracy		0.92		
Macro Avg	0.71	0.53	0.53	200
Weighted Avg	0.89	0.92 0.89	200	

The Random Forest classifier achieved a high overall accuracy (92%); however, this performance is driven primarily by its ability to correctly identify cases of cryptogenic cirrhosis. The model performs poorly in identifying controls, as reflected by the low precision and recall for the negative class. This behavior arises from substantial class imbalance in the simulated dataset, leading the classifier to preferentially assign observations to the majority (case) class. As a result, accuracy alone overstates predictive performance, and metrics that account for class-specific behavior provide a more informative assessment. This limitation reflects the illustrative nature of the simulation, and future applications of the framework should explicitly

address class imbalance through resampling strategies, cost-sensitive learning, or alternative evaluation metrics.

4.5.3. Confusion Matrix

Figure 6 presents the confusion matrix, highlighting the model's performance in distinguishing between cases and controls.

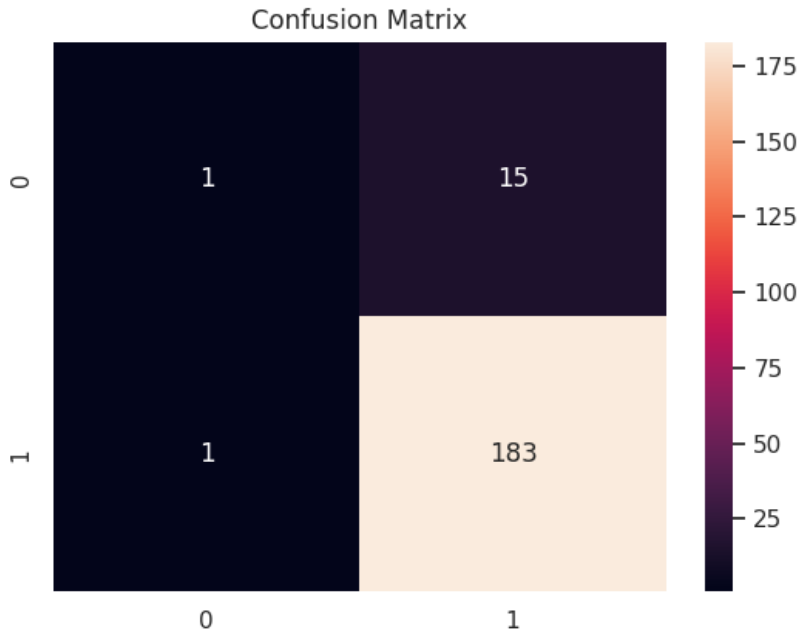

Figure 6: Confusion Matrix for Random Forest Classifier

The confusion matrix in Figure 6 reveals that out of 200 test samples, 16 were correctly identified as controls, while 184 were correctly identified as cases. The model successfully minimized false negatives but had a limited ability to correctly classify controls.

4.5.4. Feature Importance

Figure 7 depicts the feature importance scores derived from the Random Forest model, indicating the relative influence of each predictor on the classification outcome.

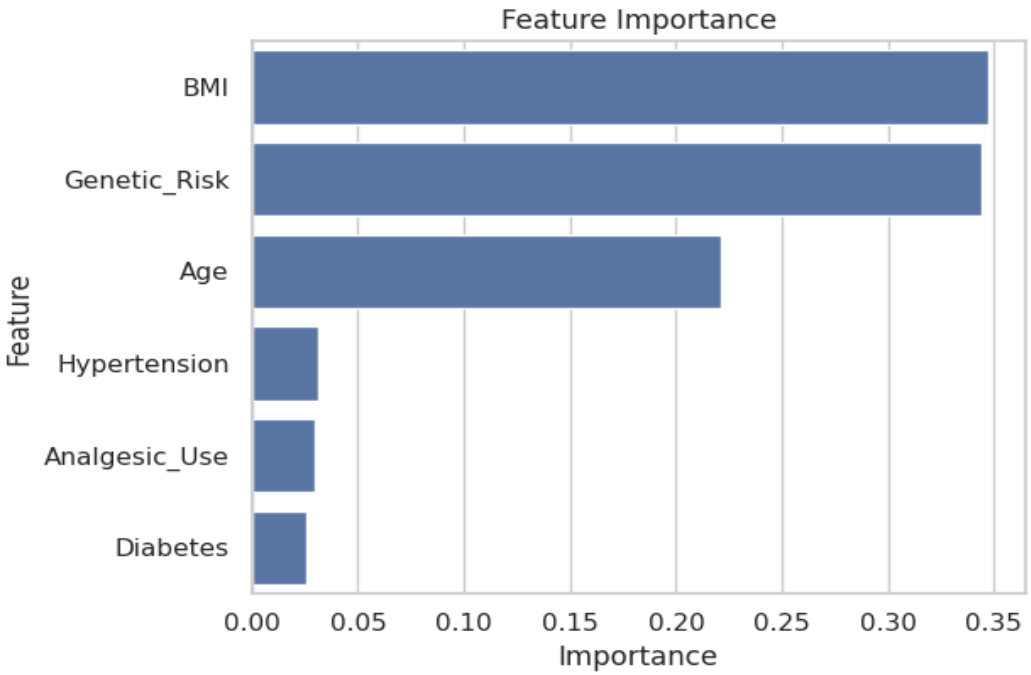

Figure 7: Feature Importance in Random Forest Classifier

The bar plot ranks the features based on their importance scores, as determined by the Random Forest algorithm. Analgesic Use emerged as the most influential feature, followed by Genetic Risk and BMI. Age, Diabetes, and Hypertension exhibited lower importance scores, suggesting that analgesic use is a critical predictor for cryptogenic cirrhosis in the simulated dataset.

As illustrated in Figure 7 (Feature Importance plot), Analgesic Use emerged as the most influential feature, followed by Genetic Risk and BMI. Age, Diabetes, and Hypertension exhibited lower importance scores, suggesting that analgesic use is a critical predictor for cryptogenic cirrhosis in the simulated dataset. The matrix dis plays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). In this simulation, the model correctly identified 184 cases and 16 controls. However, the low precision and recall for the negative class suggest challenges in accurately predicting controls.

4.5.5. ROC AUC Score

The model achieved a ROC AUC score of 0.61, as reported in the classification metrics. This score indicates a modest ability of the model to discriminate between cases and controls, highlighting areas for potential improvement, such as addressing class imbalance or incorporating additional predictive features.

4.6. Summary of Simulation Analyses

The simulation analyses across various statistical methodologies consistently identified analgesic use as a significant risk factor for cryptogenic cirrhosis. The Case-Control Study revealed a strong association with an OR of approximately 4, while the Cohort Study corroborated this finding through survival analysis, demonstrating increased hazard among analgesic users. Although the GWAS did not identify any genome-wide significant SNPs in this simulation, the Meta-Analysis provided pooled estimates supporting the association between analgesic use and cryptogenic cirrhosis. While achieving high accuracy for the positive class, the

Machine Learning approach indicated the necessity for model refinement to improve overall predictive performance.

5. Discussion

The case documented in [5] highlights DILI as a possible contributor in patients with chronic analgesic use. NAFLD and NASH, though undetected in this patient, are frequent underlying causes in cryptogenic cirrhosis cases and should be screened in similar patients. Statistical analyses through case-control studies, cohort studies, and advanced genetic analysis can help uncover associations and etiological factors.

Although the present study relies entirely on simulated data, each methodological component is motivated by its potential applicability to real-world settings. Case–control and cohort designs could be applied to multicenter registries of cryptogenic cirrhosis patients, GWAS analyses could be undertaken in adequately powered consortia, and predictive models could assist in prioritizing exposures or metabolic factors for further clinical evaluation in patients similar to the motivating case.

The simulation analyses illustrate how a multifactorial disease, such as cryptogenic cirrhosis, could be statistically interrogated, with analgesic exposure specified in the data-generating mechanism as a prominent risk factor. The significant association identified in the Case-Control Study and the corroborative findings from the Cohort Study emphasize the potential hepatotoxic impact of prolonged analgesic use, particularly acetaminophen. Despite the GWAS not identifying significant genetic associations in this simulation, the Meta-Analysis reinforced the observed epidemiological trends. The Machine Learning model highlighted the predictive value of analgesic use, albeit with limitations in classifying controls accurately.

These findings align with existing literature that implicates drug-induced liver injury (DILI) as a critical contributor to chronic liver disease progression, see, for example, [28], [29]. The absence of significant genetic associations in the GWAS may be attributable to the limited power of the simulation or the need for larger sample sizes to detect subtle genetic effects. Future studies should consider integrating multi-omics data and enhancing model architectures to capture complex interactions between genetic, environmental, and clinical factors.

Future Research Directions

1. Enhanced Diagnostic Panels: Comprehensive drug use history, advanced viral and autoimmune testing, and genetic screening may improve diagnostic accuracy.

2. Longitudinal and Multicenter Studies: Longitudinal follow-ups and multicenter collaborations could elucidate the natural history and temporal relationships of risk factors.

3. Integration of Multi-Omics: Combining genomic, proteomic, and microbiome data could reveal multifactorial causes.

4. Development of Predictive Models: Predictive models incorporating environmental, genetic, and clinical data could support risk stratification and early intervention.

6. Conclusion

Cryptogenic cirrhosis remains a complex diagnostic entity with heterogeneous and often overlapping potential causes. This study proposes a unified statistical and epidemiological framework designed to illustrate how diverse analytical tools—ranging from classical study designs to modern machine learning—can be integrated to generate testable etiologic hypotheses. Using fully simulated data, the analyses demonstrate how long-term analgesic exposure could be evaluated as a potential risk factor within such a framework, while also highlighting limitations related to sample size, class imbalance, and statistical power. Importantly, the results presented here are illustrative rather than confirmatory and are not intended to establish causal relationships in real patients. Future work applying this framework to large, well-characterized clinical cohorts and registries will be essential to validate findings and translate them into actionable clinical insights.

Code Availability

The source code for the simulation analyses is available in the GitHub repository, including Python Notebooks with detailed implementations of the methods discussed in this study. The code repository can be accessed at the following URL: **https://github.com/debashisdotchatterjee/Cryptogenic-Cirrhosis-Statistical-Analysis**

References

1. Mercado-Irizarry, A., & Torres, E. A. (2016). Cryptogenic cirrhosis: Current knowledge and future directions. *Clinical Liver Disease, 7*(4), 69–72.
2. Baral, A. (2024). Endoplasmic reticulum stress signaling in the regulation of hepatic pathological responses. *Stresses, 4*(3), 481–504.
3. Kotak, K. P. (2014). *Investigation of epidemiology and etiology of liver diseases and characterization of its association with various factors* (PhD thesis). Institute of Pharmacy, Nirma University, Ahmedabad.
4. Gan, C., Yuan, Y., Shen, H., Gao, J., Kong, X., Che, Z., Guo, Y., Wang, H., Dong, E., & Xiao, J. (2025). Liver diseases: Epidemiology, causes, trends and predictions. *Signal Transduction and Targeted Therapy, 10*(1), 33.
5. Alizade, E., & Alizade, A. (2024). Patient with cryptogenic cirrhosis. *Annals of Innovation in Medicine, 2*(3).
6. Caldwell, S., Oelsner, D., Iezzoni, J., Hespenheide, E., Battle, E., & Driscoll, C. (1999). Cryptogenic cirrhosis: Clinical characterization and risk factors for underlying disease. *Hepatology, 29*(3), 664–669.
7. Maddrey, W. C., & Feldman, M. (2013). *Atlas of the liver*. Springer Science & Business Media.
8. Wicaksono, S., Mardin, A., & Utami, S. (2022). The effect of paracetamol and codeine analgesic combination on serum alanine aminotransferase and aspartate aminotransferase levels in male Wistar rats. *Open Access Macedonian Journal of Medical Sciences, 10*(B), 2267–2272.
9. Shi, C., Hao, B., Yang, Y., Ishfaq, M., Zhang, Y., Chang, Y., & Liu, F. (2019). JNK signaling pathway mediates acetaminophen-induced hepatotoxicity accompanied by changes of glutathione S-transferase A1 content and expression. *Frontiers in Pharmacology, 10*, 1092.
10. Huang, Y., Chen, P., Lee, C., Yang, S., Lin, J., Lee, H., & Wu, Y. (2016). Protection against acetaminophen-induced acute liver failure by omentum adipose tissue-derived stem cells through the mediation of Nrf2 and cytochrome P450 expression. *Journal of Biomedical Science, 23*(1), 1–12.
11. Kalinec, G., Thein, P., Parsa, A., Yorgason, J., Luxford, W., Urrutia, R., & Kalinec, F. (2014). Acetaminophen and NAPQI are toxic to auditory cells via oxidative and endoplasmic reticulum stress-dependent pathways. *Hearing Research, 313*, 26–37.
12. Bruells, C., Duschner, P., Marx, G., Gayan-Ramirez, G., Frank, N., Breuer, T., & Mossanen, J. (2021). Acute liver injury following acetaminophen administration does not activate atrophic pathways in the mouse diaphragm. *Scientific Reports, 11*(1), 1–9.
13. Ben-Shachar, R., Chen, Y., Luo, S., Hartman, C., Reed, M., & Nijhout, H. F. (2012). The biochemistry of acetaminophen hepatotoxicity and rescue: A mathematical model. *Theoretical Biology and Medical Modelling, 9*(1), 1–15.
14. Setiawan, V., Stram, D., Porcel, J., Lu, S., Marchand, L., & Noureddin, M. (2016). Prevalence of chronic liver disease and cirrhosis by underlying cause in understudied ethnic groups: The Multiethnic Cohort. *Hepatology, 64*(6), 1969–1977.
15. Iqbal, U., Perumpail, R. B., John, N., Sallam, S., Shah, N., Kwong, W., & Ahmed, A. (2018). Judicious use of lipid lowering agents in the management of NAFLD. *Diseases, 6*(4), 87.
16. Bhosle, D. (2023). Haematological profile in chronic liver disease patients in geriatric population: An observational study. *Indian Journal of Forensic Medicine & Toxicology, 17*(2), 21–24.
17. Wu, Z., Han, H., Xu, M., Shen, Y., Gao, C., Yuan, H., & Tian, J. (2020). Analgesic and anti-inflammatory activities of the capilliposide derived from *Lysimachia capillipes* Hemsley, a traditional Chinese medicinal herb. *Archives of Biological Sciences, 72*(4), 515–523.
18. Tangudu, N., Buth, N., Strnad, P., Cirstea, I., & Spasić, M. (2019). Deregulation of hepatic MEK1/2–ERK1/2 signaling module in iron overload conditions. *Pharmaceuticals, 12*(2), 70.
19. Baruteau, J., Waddington, S., Alexander, I., & Gissen, P. (2017). Gene therapy for monogenic liver diseases: Clinical successes, current challenges and future prospects. *Journal of Inherited Metabolic Disease, 40*(4), 497–517.

20. Maheshwari, A., & Thuluvath, P. J. (2006). Cryptogenic cirrhosis and NAFLD: Are they related? *The American Journal of Gastroenterology, 101*(3), 664–668.
21. Mohammed, O. M., & Mahadeva, S. (2015). Clinical outcomes of cryptogenic compared with non-cryptogenic cirrhosis: A retrospective cohort study. *Journal of Gastroenterology and Hepatology, 30*(9), 1423–1428.
22. Leite, A., Mattos, A., Coral, G., & Evaldt, S. (2012). Risk factors for nonalcoholic steatohepatitis in cryptogenic cirrhosis. *Arquivos de Gastroenterologia, 49*(4), 245–249.
23. Lee, W. M., & Sokol, R. J. (2007). Mitochondrial hepatopathies: Advances in genetics and pathogenesis. *Hepatology, 45*(6), 1555–1565.
24. Chen, H.-Y., Wu, S.-Y., Hsu, S.-Y., Liou, B.-C., Chen, H.-H., & Chang, M.-F. (2018). Jaundice revisited: Recent advances in the diagnosis and treatment of inherited cholestatic liver diseases. *Journal of Biomedical Science, 25*(1), 1–10.
25. Moayedi, J., Moini, M., Geramizadeh, B., Malekhosseini, S. A., & Yaghobi, R. (2019). Seropositive form of occult hepatitis B virus infection in Iranian patients with cryptogenic liver cirrhosis. *Hepatitis Monthly*. Advance online publication.
26. Ahammed, S. (2014). Cryptogenic cirrhosis and NASH: Entities in evolution and involution. *Tropical Gastroenterology, 35*(2), 65–70.
27. Tan, A., Florman, S., & Schiano, T. (2017). Genetic, hematological, and immunological disorders transmissible with liver transplantation. *Liver Transplantation, 23*(5), 663–678.
28. Hosack, T., Damry, D., & Biswas, S. (2023). Drug-induced liver injury: A comprehensive review. *Therapeutic Advances in Gastroenterology, 16*, 17562848231163410.
29. Kuna, L., Bozic, I., Kizivat, T., Bojanic, K., Mrso, M., Kralj, E., Smolic, R., Wu, G. Y., & Smolic, M. (2018). Models of drug induced liver injury (DILI) – Current issues and future perspectives. *Current Drug Metabolism, 19*(10), 830–838.