Elevated bilirubin, alkaline phosphatase at onset, and drug metabolism are associated with prolonged recovery of drug-induced liver injury

Kristin Ashby[1], Wei Zhuang[1], Andres González-Jimenez[2], Ismael Alvarez-Alvarez[3], M. Isabel Lucena[3], Raúl J. Andrade[3], Guruprasad P. Aithal[4], Ayako Suzuki[5,6], Minjun Chen[1*]

1. Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, Arkansas
2. Bioinformatic platform. Instituto de Investigación Biomédica de Málaga (IBIMA), Málaga, Spain
3. Unidad de Gestión Clínica de Aparato Digestivo, Servicio de Farmacología Clínica, Instituto de Investigación Biomédica de Málaga, Hospital Universitario Virgen de la Victoria, Universidad de Málaga, CIBERehd, Málaga, Spain
4. National Institute for Health Research (NIHR) Nottingham Biomedical Research Center at the Nottingham University Hospital NHS Trust and University of Nottingham, Nottingham, UK
5. Division of Gastroenterology, Duke University, Durham, North Carolina
6. Durham VA Medical Center, Durham, North Carolina

Correspondence:

Minjun Chen, Ph.D., Email: Minjun.chen@fda.hhs.gov; Fax: 870-543-7865

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA.

Conflict of Interest:
The authors declare that they have no competing interests.

Disclaimer: The views presented in this article do not necessarily reflect those of the U.S. Food and Drug Administration. Any mention of commercial products is for clarification and is not intended as an endorsement.

Disclosures: None

Author contributions:
MC conceived and designed the study; KM, MC and AG analyzed the data and draft the manuscript; AS, GA, MC, and KM interpreted the data. WZ contributed to data analysis, model building and interpretation. AG, IAA, MIL, and RA collected and processed the cohort used for validation. MC, WZ, AS, MIL, and GA participated in critical revision of the manuscript. All authors contributed to the critical review and final approval of the manuscript.

Background & Aims: Although most drug-induced liver injury (DILI) cases resolve after the culprit medication is discontinued, time to recovery varies among patients with 6-12% developing a chronic disease. Here, we investigated clinical factors and drug properties as potential risk factors that influence DILI recovery time course and developed a model to predict cases at high risk for prolonged recovery.
Methods: We applied an accelerated failure time model to 294 cases collected by the International Drug-Induced Liver Network Consortium. Factors included in the multivariate recovery score model were selected through univariate analysis. The model was externally validated by 385 cases from the Spanish DILI registry and 191 cases from the LiverTox database.
Results: Higher serum bilirubin and ALP at DILI onset, a longer time to onset, and non-significant drug metabolism were associated with a longer recovery and were included in the multivariate recovery score model. We divided cases into risk groups based on the score assigned by the recovery score model. The estimated probability of recovery at six months was 0.46 (95% CI:0.26-0.61) for the high-risk group and 0.93 (95% CI:0.58-0.99) for the low-risk group. In both validation sets, the high-, and low-risk cases identified by the model showed a significantly different time-course for recovery, with a majority of low-risk cases recovering sooner.
Conclusion: Biochemical recovery from DILI is influenced by the extent of culprit drug metabolism and serum bilirubin and alkaline phosphatase at DILI onset. Understanding the factors that delay DILI recovery may give important insight into the mechanisms and development of chronic DILI and guide patient management after DILI events.

Lay summary: In this study, we investigated whether drug properties and clinical factors affect the time it takes to recover from drug-induced liver injury. We found that total bilirubin, ALP level at DILI onset, time to onset, and extent of drug metabolism were consistently associated with recovery time. Using these factors, we built a model to identify patients at a higher risk of delayed recovery and tested this model in two independent cohorts. Our findings will give important insight into factors contributing to delayed recovery from drug-induced liver injury.

Keywords: *Drug-induced liver injury, prolonged recovery, scoring model, risk factor, accelerated failure time*

Drug-induced liver injury (DILI) is a clinically significant adverse reaction. Although most DILI resolves after discontinuation of the culprit medication, the time to recovery varies among patients with 6-12% cases eventually developing chronic liver injury.[1, 2]

Chronic liver injury may result in fibrosis, bile duct loss, and cirrhosis and negatively impact quality of life.[1-4] Currently, the understanding of underlying mechanisms driving the development of chronic or persistent DILI is limited. Certain clinical factors have been reported as risk factors for chronic DILI, but the results are not entirely consistent. For example, Fontana et al.[1] reported that cholestatic pattern of DILI was more frequent in persistent DILI cases while Medina-Caliz et al.[3] did not find a significant association between chronicity and cholestatic pattern. Age is another clinical factor that was significant in some studies but insignificant in others.[1, 3, 5] A prospective study conducted by the US Drug-Induced Liver Injury Network described 17% of the DILI cases as chronic based on abnormal serum biochemistries six months after enrollment[5, 6] and Medina-Caliz et al. of the Spanish DILI registry found that 8% of DILI cases persisted for more than one year.[3]

One reason for the inconsistency between studies may stem from the different definitions of chronic DILI. Chronic cases are defined by abnormal serum chemistry values for an extended period of time, either six months [5, 6] or one year.[3, 7] Although these studies each categorized cases as chronic based on biological or medical rationales, the number of days of abnormal serum chemistry values is a continuous variable and these different cut-offs could lead to inconsistent observations.

In this study, we focused on the time to recovery after DILI injury and used an accelerated failure time model to explore host factors and drug properties and identify the potential risk factors that could influence the time-course of DILI recovery after discontinuation of the culprit medications. We further defined a model and score based on these factors (i.e. bilirubin and alkaline phosphatase (ALP) at DILI onset, time to onset, extent of drug metabolism) and validated the model using two independent ccohorts with 385 cases from the Spanish DILI registry[8] and 191 DILI cases collected from the National Institutes of Health (NIH) LiverTox database.[9]

Materials and Methods

DILI Cohort

Cases in this study were part of the International DILI Consortium (iDILIC), a large collaborative study with recruitment centers across Europe, Asian, and Australia. The 720 cases included in this study are from batches 1 and 2 of the fourth release of iDILIC. Batch 3 from the fourth release was excluded because follow-up serum biochemistries were not reported.

Inclusion criteria for cases in the iDILIC cohort were based on clinical chemistry criteria for DILI as defined by Aithal et al.,[7] which states that a qualified case must have either alanine aminotransferase (ALT) elevated at least five-fold above the upper level of normal (ULN), or at least a two-fold elevation of ALP above ULN or elevated levels of ALT at least three-fold above ULN while bilirubin concentrations are also over two-fold ULN. Cases were also assessed using the Roussel Uclaf Causality Assessment Method (RUCAM) scoring system and expert review consisting of a panel of three hepatologists.

In this study, only cases with a RUCAM causality scale of probable (i.e. score of greater than or equal to 6) were included. In addition, only patients with initial and follow-up serum biochemistries and without long intervals (more than six months) between the final elevated and the normalized serum biochemistry dates were included. We excluded cases where pre-existing liver disease were present and

cases that resulted in liver transplant or death. Only causal drug combinations that occurred more than five times in the iDILIC cohort were included. For example, amoxicillin-clavulanate and sulfamethoxazole-trimethoprim are frequently given in combination. However, causal drug combinations such as diclofenac and flucloxacillin, which are not frequently given in combination and occurred less than five times in the cohort, were excluded. To avoid conflict with the validation cohort, we excluded cases that were collected by the Spanish DILI group. The final analysis includes 294 cases. Figure 1 describes criteria for case inclusion and exclusions.

Patient Recovery

Patient recovery was defined as whether the patient's serum biochemistries returned to normal. Time to recovery, or time followed, was calculated in days from the day of withdrawal of the culprit medication[10, 11] to the date when liver serum biochemistries normalized (1xULN) or the last day of follow-up. Patients with serum ALT, AST, ALP, or bilirubin that did not return to 1xULN were censored at the date of their last recorded follow-up. Both censored cases and cases that returned to 1xULN are included in the regression modeling of the accelerated failure time analysis, which is a parametric time-to-event analysis.

Host Factors

Host factors were collected from the clinical data provided by iDILIC. Clinical information includes medical history information, concomitant medications, liver enzymes, and other clinical features (Table 1). The type of liver injury was categorized as hepatocellular, mixed, or cholestatic using the R value at DILI diagnosis as described by Benichou et al.[12] We defined DILI onset as the date of DILI diagnosis and the time to DILI onset as the days from the initial drug intake to the DILI diagnosis. Liver biochemistry tests that were taken at DILI diagnosis were included as well. Delayed drug discontinuation was defined as the number of days from DILI diagnosis to drug discontinuation. DILI severity was categorized as defined by Aithal et al.[7] where mild cases meet aforementioned clinical biochemistry criteria for DILI, moderate cases meet criteria for DILI and bilirubin values are greater than two times ULN, and severe cases meet moderate criteria and have one of the following: ascites, encephalopathy, international normalization ratio >1.5 and/or other organ failure due to DILI.

Drug Properties

In this study, we included the following drug properties: daily dose, lipophilicity, and extent of metabolism. Information on drug property was retrieved from the Liver Toxicity Knowledge Base[13] and literature sources. Extent of drug metabolism is defined as high when ≥ 50%; otherwise, it is defined as low following the definition in Lammert et al.[14] Drug combinations that occurred more than five times were included and their drug properties combined by taking the maximum value. Drug combinations occurring in five or fewer cases were removed from the analysis. Causal drugs and their frequencies are shown in Supplemental Table 1.

LiverTox Case Reports

We downloaded 389 case reports from the NIH LiverTox web site (www.livertox.nih.gov)[9] for an independent validation analysis. The case reports include liver biochemistries, DILI severity, pattern of injury, time to onset, age, sex, causal drug, and recovery time. After removing cases in which dietary supplements were culprit and cases that were missing initial serum biochemistries, 191 cases remained with a median follow-up time of 60 days (range: 4-300 days). Supplemental Table 2 includes LiverTox drug frequencies.

Spanish DILI Registry Cases

To further validate the model, we applied the recovery score prediction to 385 cases collected by the Spanish DILI registry.[8] These cases have a median follow-up time of 111 days (range: 5-3020 days). The inclusion criteria used in the main iDILIC cohort were applied to this cohort as well.

Statistical Methods

Univariate Analysis

Accelerated failure time (AFT) models were used for time-toevent analysis, where the event was DILI recovery as defined by the return of ALT, ALP, and bilirubin to normal values or 1xULN. In order to identify clinical factors and drug properties affecting patient recovery, we first screened clinically-relevant variables in a univariate analysis. These potential clinical risk factors included sex, age, time to onset, and liver biochemistries (bilirubin, ALP, and ALT) at DILI onset. We also included drug properties: daily dose, extent of metabolism, and lipophilicity. We transformed time to onset and the liver serum biochemistry values (ALT, ALP, and bilirubin) to their natural logarithms.

Two drugs, namely amoxicillin-clavulanate and flucloxacillin, account for 28% (N=82) and 27% (N=79) of the total cases in this cohort (N=294), respectively. Therefore, we conducted the univariate analysis in the entire cohort as well as in subsets including only amoxicillin-clavulanate or flucloxacillin.

Development of DILI Recovery Time Model

We then built a multivariate AFT model using factors approaching significance (p-value < 0.1) in the univariate analysis of the entire cohort. A factor will not be selected if its correlation with other factors is 0.3 or greater or is significantly associated with other factors (p-value <0.01). The model was used to calculate the recovery score and divide cases into high-, indeterminate-, and low-risk groups. Cases with a score of one standard deviation above the mean or greater were classified as high-risk of delayed recovery, those within one standard deviation of the mean were classified as indeterminate-risk, and those with a score one standard deviation below the mean or lower were classified as low-risk.

To determine which AFT distribution best fit the data, we compared the models estimated under three alternative distributional assumptions: the log-normal, Weibull, and log-logistic. All models retain the same other assumptions, e.g. independence across patients. The models were compared and evaluated by minimization of the Akaike information criterion (AIC) (Supplemental Table 3), range in size of standardized residuals (Supplemental Figure 1), graphical comparison of nonparametric modified Kaplan-Meier estimates that adjusts for covariates estimates against fitted survival estimates (Supplemental Figure 2), and Kaplan-Meier estimates of residuals (Supplemental Figure 3). The log-normal model had the smallest AIC, the smallest range of standardized residuals, and also fit the Kaplan-Meier estimates based on the comparison graphs. In addition, the overall fit of the log-normal model was also confirmed by the quantile-quantile plot of the log of time to recovery against the theoretical normal quantiles (Supplemental Figure 4). Therefore, we selected the log-normal distribution for the AFT multivariate model with covariates adjusted. The overall

goodness-of-fit of the AFT model was evaluated with an overlay of Cox–Snell residuals by follow-up time, in which a straight line suggests a good model fit (Supplemental Figure 5).

The defined multivariate score model was validated by predicting prolonged recovery cases in two independent validation cohorts consisting of 385 cases from the Spanish DILI registry and 191 cases from the NIH LiverTox database. The risk score cut-offs defined in the original study population were used to categorize these cases into high-, indeterminate-, and low-risk groups. Recovery rates were determined using the Kaplan-Meier method and the log-rank test was used to compare the recovery time between the high- and low-risk groups. We also considered performance in specific subgroups based on injury type, case severity, and RUCAM scores.

Other Statistical Analysis

Descriptive statistics, mean and standard deviation, were used to describe continuous variables, and frequency and percent were used to describe categorical variables. All analysis were performed using R (version 3.6.1)[15] and the survival[16] package for the accelerated failure time model, htmlTable[17] for clinical and drug tables, car[18] for Q-Q plots, and survminer[19] for Kaplan-Meier plots. Code for Kaplan-Meier estimator of residuals (Supplemental Figure 3) was adapted from Rizopoulos.[20]

Results

Clinical Characteristics of the Study Population

After applying inclusion and exclusion criteria (Figure 1), 294 cases remained. Of these, 140 cases recovered within the follow-up period and 154 either did not recover or were lost to follow-up. The mean/median follow-up time was 82/68 days (range: 1-587 days). The average age was 59 years (range: 14-91 years) and 60% of the cases were female. A majority of patients (94.2%) were Caucasian, 3.4% were Asian, and the remaining 2.4% were other or unknown race. The pattern of liver injury was cholestatic in 26% of cases, hepatocellular in 35%, and mixed in 39%. Amongst cases that recovered, the mean/median time to biochemical recovery was 79/73 days (range: 10-259 days). In censored cases that were lost to follow up before they fully recovered, the mean/median follow-up time was 84/59 days (range: 1-587 days). Clinical characteristics and comorbidities of the 294 cases are shown in Table 1.

Antibiotics were responsible for 68.4% of cases, followed by NSAIDs (6.5%), and antihyperlipidemics (3.4%). Amoxicillin-clavulanate and flucloxacillin were two of the most frequent drugs and were causal in 82 (27.9%) and 79 (26.9%) cases, respectively. Additional drug frequencies are shown in Supplemental Table 1.

Univariate Analysis

The results for the AFT univariate are shown in Table 2. In the entire cohort, a one-unit increase in $\log_e$ of bilirubin times the upper limit of normal at DILI onset was associated with a 46% ($p<0.001$) and ALP a 50% longer recovery time ($p<0.001$). Moderate to severe clinical severity (vs. mild) was associated with a 109% longer recovery time and was statistically significant ($p<0.001$). Hepatocellular injury was associated with 54% shorter recovery time compared with cholestatic injury ($p<0.001$). A higher bilirubin was also associated with a prolonged recovery in the amoxicillin-clavulanate and the flucloxacillin subgroups. A longer time to DILI onset was associated with prolonged recovery time in the amoxicillin-clavulanate subgroup. ALT at DILI onset, age, and sex were not significant in any of the subgroups. We also considered injury types classified by nR,[21] however; a large portion of cases (60%) lack AST values, and ALT alone was used for classification of these cases.The results were similar with only three mixed injury reclassified into hepatocellular injury by using nR.

Drug properties were tested in the entire cohort. Culprit drugs that are eliminated primarily through hepatic metabolism were significantly associated with 52% shorter recovery time (Table 2), compared with culprit drugs without significant hepatic metabolism.

DILI Recovery Time Model

Total bilirubin, ALP at DILI onset, time to onset, and extent of drug metabolism were selected for the multivariate analysis (see Table 3), which we used to calculate recovery score. Severity and injury type were not selected since they were derived from and correlated with other selected variables. A score model for DILI recovery derived from the AFT log-normal approach was defined as below:

$$\text{Recovery Score} = 0.227 * \log_e(\text{ALP xULN at onset}) + 0.277 * \log_e(\text{Bilirubin xULN at onset}) + 0.161 * \log_e(\text{time to onset}) - 0.440 * (\text{significant hepatic metabolism of culprit drug}).$$

The range of possible recovery scores was -0.60 to 2.03, where the higher score indicates a greater likelihood of prolonged recovery.

The cases were categorized into high-risk for prolonged recovery (recovery score > 1.30), and low-risk (recovery score ≤ 0.44) to evaluate association between the calculated scores and the likelihood of delayed recovery. Cases between the threshold were indeterminate. As shown in Figure 2, the risk groups had a significantly different time-course for recovery ($p<0.0001$). Specifically, the probability of recovery at 6 months for the high-risk group was estimated as 46% (95% CI:0.26-0.61) and the low group was 93% (95% CI:0.58-0.99). The estimated probability of recovery at three, six, and nine months was consistently higher in the low-risk group, as shown in Table 4.

As seen in Supplemental Figure 6, cases with a RUCAM score of eight or more were categorized by the recovery risk score model into significantly different recovery groups according to the log-rank test ($p<0.0001$) but those with a score of six or seven had a higher p-value($p=0.025$). This suggests that the model performs better when applied to higher quality data.

Model validation

We then validated the model by predicting delayed recovery cases in 385 cases from the Spanish DILI registry. The high-and low-risk cases identified by the model showed a significantly different time-course for recovery (Figure 3, $p=0.0028$).

We also validated the model in 191 cases downloaded from LiverTox and applied the recovery score model to the LiverTox cases, categorizing them into risk groups. The difference in recovery between the high- and low-risk groups was statistically significant using a log rank test (Figure 4, $p=0.0004$). The population characteristics of LiverTox cases are shown in Supplemental Table 5.

Discussion

Herein, we modeled recovery after DILI injury using clinical factors and drug properties by using an accelerated failure time model in a large cohort of well characterized acute DILI patients. We found that not only total bilirubin, ALP at DILI onset, time to onset but, importantly, extent of drug metabolism were consistently associated with DILI recovery time, and a scoring model based on these factors was developed from 294 DILI cases and validated in an independent cohort of 385 samples from the Spanish DILI registry[8] and 191 DILI cases collected from the LiverTox database.[9] To our knowledge, this is the first report to use regression survival analysis to investigate DILI resolution, including not just host factors but also drug properties.

Our study has a number of strengths, including the substantial size of the cohort, multinational source, strict inclusion criteria, and external validations of the model. The 294 well-defined DILI cases verified by expert review and causality assessment were retrieved from the iDILIC, which is part of the International Serious Adverse Event Consortium[22] and recruits patients primarily from DILI centers across Europe with clinical and culprit medication information. In our study, only cases with a RUCAM score of six or higher were included and cases with pre-existing conditions were also removed. The recovery score model performed better when applied to cases with a RUCAM of eight or more, which suggests that the model performs well in the higher quality cases (Supplemental Figure 6). We further considered performance in specific subgroups including injury types and severity. We found that the model performed best for hepatocellular cases, separating them into distinct groups (Supplemental Figure 7A, log-rank p=0.0084). Similarly results were found in the Spanish and LiverTox validation sets. In addition, the model performed better for moderate-severe DILI cases than mild DILI cases (Supplemental Figure 8).

Compared with the LiverTox and Spanish DILI cohorts,[3, 5] the cohort in this study has similar demographic distribution in sex and slightly older in age, i.e. the average age at DILI diagnosis is 59 years in this cohort, 52 years in the Spainish cohort and 49 years in the LiverTox cohort.[3, 5] Note, the cohort here contains 68.4% cases caused by antimicrobials, which is higher than those in Spanish cohort (34%) and LiverTox cohort (45%).[3, 5] Specifically, our cohort includes 28% Amoxicillin-clavulanate and 27% flucloxacillin cases, which leads to a higher portion of mixed (39%) and cholestatic (26%) cases. Compared to iDILIC, the LiverTox cohort includes a more even distribution of drugs with 113 drugs that only occur one time and 148 unique drugs (Supplemental Table 2). Even so, our model was successfully validated by two independent DILI cohorts, including 385 cases from the Spanish DILI Registry and 191 cases from the LiverTox database.

In addition, rather than reducing statistical power by using a cut point to define chronic cases, we considered patient time to recovery as a continuous variable. Using a continuous variable has statistical advantages over those classified by hard threshold (e.g. 6 months). It has been well documented that there are statistical disadvantages to categorizing continuous variables,[23, 24] which can result in a considerable loss of statistical power and increase the risk of false positives. Even cut-points based on medical rationales can be problematic for borderline cases. The accelerated failure time analysis also allowed us to include cases without complete follow-up information. By contrast, a logistic regression that compares chronic and acute cases using a six-month cutoff must exclude cases that do not resolve and have less than six-months of follow-up. In this cohort, only bilirubin was identified as a significant variable using logistic regression with a six-month cutoff.

We employed the AFT model rather than the Cox proportional hazards in this study. The Cox model is a semi-parametric model that does not assume that the survival times or outcome must follow a certain

statistical distribution. However, it does rely on the assumption of constant proportional hazard ratios, and a violation of such will result in an improper fitting of the model and incorrect inferences.[25] We found that the recovery time data violated the assumptions of the proportional hazards. Alternatively, the accelerated failure time model does not assume proportional hazards and is easier to interpret but does require a parametric distribution for the survival times. The AFT multivariate model fit the log-normal distribution well, except for departures at the tail of the distribution where very few cases remain.

Total bilirubin at DILI onset was found consistently associated with the length of time to DILI recovery across drugs in the entire cohort and the drug specific subsets. Previous studies have identified bilirubin as significantly associated with chronic[3] or persistent[1] DILI but not at DILI onset time; however, Medina-Caliz et al. did find that bilirubin was significantly elevated in chronic cases in the second month after onset and that jaundice at onset was a risk factor for DILI chronicity. Other studies define chronic DILI as six months or one year from onset,[3, 5, 26, 27] whereas our study relied on an accelerated failure time model rather than a binary cutoff, giving us more statistical power to identify influential factors (i.e. total bilirubin at onset).

ALP at onset was significant in the entire cohort and in the subgroup that excluded both amoxicillin-clavulanate and flucloxacillin cases (p=0.026). A similar finding was previously reported in other studies of DILI chronicity.[1, 3, 27] As in other studies, ALT was not a significant predictor of recovery time. This is in keeping with the clinical observations that degree of ALT elevation is not predictive of severe clinical outcomes such as acute liver failure. See reviewer comment and soften ALT.

Injury type was significant in the entire cohort (p=0.001) but not significant in the amoxicillin-clavulanate and flucloxacillin subgroups. Notably, Both of these drugs have a high prevalence of cholestatic and mixed cases with 25% and 58% of flucloxacillin cases attributable to cholestatic and mixed injury, and 33% and 41% of amoxicillin-clavulanate cases attributable to cholestatic and mixed injury, respectively. Previous studies differ on the significance of injury type with some finding injury type significant[1, 5] and others finding it not significant.[3] Significance of age was also inconsistent, with some studies reporting a higher frequency of chronic DILI cases in older patients,[1, 3] while another reported a higher frequency of younger patients with chronic DILI,[5] and still others reported no significant association with age.[4, 27] Interestingly, it has been suggested that the prevalence of certain medications in specific age categories may influence the significance of age in chronic DILI.[1] Since certain drugs have unique clinical signatures, it is possible the inconsistencies between studies is due in part to the different drug frequencies and drug properties. In addition, some of these studies considered elevated liver biochemistries at sixth months as chronic or persistent DILI and others at twelve months.

Besides clinical factors, drug factors such as lipophilicity, dose, and metabolism can also influence DILI phenotypes[28] or DILI risk.[14, 29-31] In this study, we found that the extent of hepatic metabolism of culprit drugs was significantly associated with recovery time. Significant hepatic metabolism has been linked to hepatocellular injury and severe DILI outcomes.[14] It is not surprising then that our model performed best when applied to hepatocellular and severe DILI cases (Supplemental Figure 7A and Supplemental Figure 8, respectively).

We also wanted to determine whether the model could separate cases with the same culprit drug into different risk categories. Supplemental Figure 9 demonstrates the Kaplan-Meier curve and risk

categories assigned to 79 flucloxacillin cases by the recovery score model. The high and indeterminant group have a clearly different recovery course, and the log-rank test was slightly significant with a p-value = 0.032.

A possible limitation of our study is that it is retrospective, and case follow up time is influenced in part by the clinician decision; cases which are selected as case reports, such as the Spainish DILI cases, may be followed longer than other cases. Because only five patients were followed for at least twelve months, extending the model past twelve months is beyond the scope of the model and
See reviewer 3, comment one. Add discussion here.

DILI recovery is determined by liver enzyme data, and very few cases have histological data for supporting evaluations. Overall statistically, the model works for the entire cohort; however, when applied to subgroups based on type of liver injury, the model performs best when applied to hepatocellular cases. This is likely due to mechanistic differences in the development of different kinds of liver injury. Further study is required to model cholestatic DILI and DILI extending beyond twelve months. We also noted that two drugs, amoxicillin-clavulanate and flucloxacillin, were the culprit medication in a large portion (55%) of the cases. In our study, we specifically investigated the subpopulation taking these two drugs. In addition, this cohort includes very few targeted therapies and only one tyrosine kinase inhibitor case; thus, the model is not well-trained for the targeted therapies. Further data including additional targeted therapies would improve performance of the model in this subpopulation.

In conclusion, we have identified drug related factors and clinical manifestation in the form of degree of bilirubin and ALP elevation at the onset of DILI that are statistically significantly associated with prolonged recovery. The model we developed was robust, maintaining significance in drug specific subgroups, as well as a separate cohort which included a number of different drugs. Considering that patients with prolonged recovery are associated with persistent symptoms such as itching reducing quality of life, our recovery score model stratified patients into high/low risk groups for prolonged DILI recovery, which could inform the plan of follow up in these patients following the initial diagnosis of DILI. Frequent investigations and clinic visits associated with prolonged recovery may also add to the cost of care of these patients adding to the burden on health services.

Abbreviations

ALP, alkaline phosphatase; ALT, alanine aminotransferase; DILI, drug-induced liver injury; RUCAM, Roussel Uclaf causality assessment ULN, upper limit of normal; method

Acknowledgements

We thank Dr. Jurgen Borlak's comments and suggestions and acknowledge iSAEC for providing the access to the iDILIC data used in this study. MIL, RJA, IAA, AG, AS, and GPA are members of European Cooperation in Science & Technology (COST) Action CA17112 Prospective European DILI Network.

References

[1] Fontana RJ, Hayashi PH, Barnhart H, Kleiner DE, Reddy KR, Chalasani N, et al. Persistent liver biochemistry abnormalities are more common in older patients and those with cholestatic drug induced liver injury. The American journal of gastroenterology 2015;110:1450.

[2] Hayashi PH, Bjornsson ES. Long-term outcomes after drug-induced liver injury. Current hepatology reports 2018;17:292-299.

[3] Medina-Caliz I, Robles-Diaz M, Garcia-Muñoz B, Stephens C, Ortega-Alonso A, Garcia-Cortes M, et al. Definition and risk factors for chronicity following acute idiosyncratic drug-induced liver injury. Journal of hepatology 2016;65:532-542.

[4] Björnsson E, Davidsdottir LJJoh. The long-term follow-up after idiosyncratic drug-induced liver injury with jaundice. 2009;50:511-517.

[5] Chalasani N, Bonkovsky HL, Fontana R, Lee W, Stolz A, Talwalkar J, et al. Features and outcomes of 899 patients with drug-induced liver injury: the DILIN prospective study. Gastroenterology 2015;148:1340-1352. e1347.

[6] Fontana RJ, Watkins PB, Bonkovsky HL, Chalasani N, Davern T, Serrano J, et al. Drug-induced liver injury network (DILIN) prospective study. Drug safety 2009;32:55-68.

[7] Aithal G, Watkins P, Andrade R, Larrey D, Molokhia M, Takikawa H, et al. Case definition and phenotype standardization in drug-induced liver injury. Clinical Pharmacology & Therapeutics 2011;89:806-815.

[8] Andrade RJ, Lucena MI, Fernández MC, Pelaez G, Pachkoria K, García-Ruiz E, et al. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 2005;129:512-521.

[9] Hoofnagle JH. LiverTox: a website on drug-induced liver injury. Drug-Induced Liver Disease: Elsevier; 2013. p. 725-732.

[10] Andrade RJ, Lucena MI, Kaplowitz N, García-Muṇoz B, Borraz Y, Pachkoria K, et al. Outcome of acute idiosyncratic drug-induced liver injury: long-term follow-up in a hepatotoxicity registry. 2006;44:1581-1588.

[11] Björnsson E, Kalaitzakis E, Av Klinteberg V, Alem N, Olsson RJAp, therapeutics. Long-term follow-up of patients with mild to moderate drug-induced liver injury. 2007;26:79-85.

[12] Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs—II. An original model for validation of drug causality assessment methods: case reports with positive rechallenge. Journal of clinical epidemiology 1993;46:1331-1336.

[13] Chen M, Zhang J, Wang Y, Liu Z, Kelly R, Zhou G, et al. The liver toxicity knowledge base: a systems approach to a complex end point. Clinical Pharmacology & Therapeutics 2013;93:409-412.

[14] Lammert C, Bjornsson E, Niklasson A, Chalasani N. Oral medications with significant hepatic metabolism at higher risk for hepatic adverse events. Hepatology 2010;51:615-620.

[15] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. . 2019.

[16] Therneau TM, Lumley T. Package 'survival'. Survival analysis Published on CRAN 2014.

[17] Gordon M, Gragg S, Konings P. htmlTable: Advanced Tables for Markdown/HTML. R package version 2018;1.

[18] Fox J, Weisberg S. An R Companion to Applied Regression (Third). Sage; 2019.

[19] Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using'ggplot2'. R package version 03 2017;1.

[20] Rizopoulos D. EP03: Survival Analysis in R Companion. Section 3.4. 2020.

[21] Robles–Diaz M, Lucena MI, Kaplowitz N, Stephens C, Medina–Cáliz I, González–Jimenez A, et al. Use of Hy's law and a new composite algorithm to predict acute liver failure in patients with drug-induced liver injury. Gastroenterology 2014;147:109-118. e105.

[22] Holden AL. The innovative use of a large-scale industry biomedical consortium to research the genetic basis of drug induced serious adverse events. Drug Discovery Today: Technologies 2007;4:75-87.
[23] Altman DG, Royston PJB. The cost of dichotomising continuous variables. 2006;332:1080.
[24] Royston P, Altman DG, Sauerbrei WJSim. Dichotomizing continuous predictors in multiple regression: a bad idea. 2006;25:127-141.
[25] Chapman J, O'callaghan C, Hu N, Ding K, Yothers G, Catalano P, et al. Innovative estimation of survival using log-normal survival modelling on ACCENT database. British journal of cancer 2013;108:784-790.
[26] Aithal P, Day CJG. The natural history of histologically proved drug induced liver disease. 1999;44:731-735.
[27] Fontana RJ, Hayashi PH, Gu J, Reddy KR, Barnhart H, Watkins PB, et al. Idiosyncratic drug-induced liver injury is associated with substantial morbidity and mortality within 6 months from onset. 2014;147:96-108. e104.
[28] Gonzalez-Jimenez A, McEuen K, Chen M, Suzuki A, Robles-Diaz M, Medina-Caliz I, et al. The influence of drug properties and host factors on delayed onset of symptoms in drug-induced liver injury. 2019;39:401-410.
[29] Chen M, Borlak J, Tong WJH. High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. 2013;58:388-396.
[30] Chen M, Suzuki A, Borlak J, Andrade RJ, Lucena MI. Drug-induced liver injury: Interactions between drug properties and host factors. Journal of hepatology 2015;63:503-514.
[31] McEuen K, Borlak J, Tong W, Chen MJIjoms. Associations of drug lipophilicity and extent of metabolism with drug-induced liver injury. 2017;18:1335.

Tables

Table 1. Clinical characteristics in 290 DILI cases.

		Injury Type		
	Entire cohort (N=290)	**Cholestatic (N=75)**	**Hepatocellular (N=101)**	**Mixed (N=114)**
Sex (n,%)				
Female	173 (60%)	40 (23%)	63 (36%)	70 (40%)
Male	117 (40%)	35 (30%)	38 (32%)	44 (38%)
Age (years)				
Mean (range)	61 (17 - 91)	65 (32 - 91)	56 (17 - 83)	62 (19 - 85)
Missing (n,%)	1 (0%)	0 (0%)	1 (100%)	0 (0%)
Age-Sex (n,%)				
Under 55 Female	71 (24%)	9 (13%)	35 (49%)	27 (38%)
Under 55 Male	30 (10%)	6 (20%)	14 (47%)	10 (33%)
Over 55 Female	101 (35%)	31 (31%)	27 (27%)	43 (43%)
Over 55 Male	88 (30%)	29 (33%)	25 (28%)	34 (39%)
Body Mass Index				
Mean (SD)	26.1 (±4.5)	25.4 (±4.2)	26.2 (±4.3)	26.3 (±4.9)
Missing (n,%)	17 (5.9%)	5 (29%)	7 (41%)	5 (29%)
Clinical Presentation (n,%)				
Jaundice	69 (24%)	20 (29%)	21 (30%)	28 (41%)
Hospital Admission	200 (69%)	51 (26%)	70 (35%)	79 (40%)
Missing	4 (1%)	1 (25%)	3 (75%)	0 (0.0%)
Hypersensitivity	8 (3%)	1 (12%)	4 (50%)	3 (38%)
Time to Onset (days)				
Mean (SD)	42.5 (±80.4)	38.5 (±49.8)	52.4 (±84.8)	36.4 (±91.7)
Median (range)	25 (1 – 955)	29 (4 - 365)	26 (1-519)	22 (1-955)
Follow up (days)				
Mean (SD)	82.3 (±73.9)	98.0 (±80.0)	69.2 (±65.2)	83.7 (±75.5)
Median (range)	68 (1-587)	82 (1-543)	53 (3-391)	68 (6-587)
Censored (n, %)	150 (51.7%)	44 (29%)	45 (30%)	61 (41%)
Mean (SD)	85.3 (±91.8)	106.2 (±98.3)	65.9 (±78.3)	84.5 (94.3)
Median (range)	58.5 (1-587)	83 (1 - 543)	39 (3 - 391)	55 (6 -587)
Time to Recovery*				
Recovered (n, %)	140 (48.3%)	31 (22%)	56 (40%)	53 (38%)
Mean (SD)	79.2 (±47.9)	86.4 (±40.9)	71.8 (±53.0)	82.7 (±45.9)
Median (range)	73 (10 - 259)	82 (20 - 194)	66 (10 - 231)	76 (11 - 259)
Laboratory Parameters at DILI onset				
ALT (/ULN), mean (SD)	12.3 (±11.7)	4.9 (±3.5)	21.0 (±15.5)	9.6 (±4.5)
AST (/ULN), mean (SD)	10.0 (±12.1)	3.6 (±2.7)	16.3 (±15.9)	7.3 (±6.9)
Missing, n (%)	174 (60%)	48 (28%)	55 (32%)	71 (41%)
ALP (/ULN), mean (SD)	2.9 (±2.3)	4.6 (±3.5)	1.6 (±0.9)	3.0 (±1.3)
Bilirubin (/ULN), mean (SD)	5.1 (±4.5)	6.2 (±6.1)	4.4 (±4.2)	5.1 (±3.3)
Severity (n,%)				
Mild	42 (14%)	12 (29%)	17 (40%)	13 (31%)
Moderate	244 (84%)	61 (25%)	84 (34%)	99 (41%)
Severe	3 (1%)	1 (33%)	0 (0%)	2 (67%)

Table 1. Clinical characteristics in 290 DILI cases.

	Entire cohort (N=290)	Injury Type		
		Cholestatic (N=75)	Hepatocellular (N=101)	Mixed (N=114)
Fatal or Transplant removed				
Missing	1 (0%)	1 (100%)	0 (0%)	0 (0%)
RUCAM Causality Score (n,%)				
Definite or highly probable	111 (38%)	26 (23%)	38 (34%)	47 (42%)
Probable	179 (62%)	49 (27%)	63 (35%)	67 (37%)
Comorbidities (n,%)				
Diabetes mellitus	9 (3%)	4 (44%)	3 (33%)	2 (22%)
Hypertension	55 (19%)	19 (35%)	18 (33%)	18 (33%)
Tuberculosis	6 (2%)	0 (0%)	6 (100%)	0 (0%)
Lipid metabolism disorders	5 (2%)	2 (40%)	0 (0%)	3 (60%)
Psoriasis	11 (4%)	4 (36%)	4 (36%)	3 (27%)
Dermatitis	7 (2%)	2 (14%)	1 (14%)	5 (71%)
Missing	24 (8%)	8 (33%)	5 (21%)	11 (46%)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; Hypersensitivity: fever, rash, and/or eosinophilia; Severity, Mild: elevated ALT/ALP meeting DILI criteria; Moderate: elevated ALT/ALP meeting DILI criteria and bilirubin ≥ 2xULN; Severe: elevated ALT/ALP, bilirubin ≥ 2xULN, and one of the following: ascites, encephalopathy, international normalization ratio >1.5 and/or other organ failure due to DILI; Fatal: death or transplantation due to DILI; RUCAM, Roussel Uclaf Causality Assessment Method. *Time to recovered includes only cases that resolved within follow-up time

Table 2. Univariate accelerated failure time estimates of the impact of host factors and drug properties on the time to biochemical recovery.

	Entire Cohort Time ratio (95% CI) N=290	**Amoxicillin-clavulanate Time ratio (95% CI) N=82**	**Flucloxacillin Time ratio (95% CI) N=79**
Age	1.00 (1.00-1.01)	1.01 (0.99-1.02)	1.00 (0.99-1.02)
Age below median (61 years)	0.98 (0.77-1.25)	0.99 (0.73-1.36)	1.00 (0.62-1.62)
Sex (Male)	0.85 (0.67-1.09)	0.82 (0.60-1.10)	0.66 (0.41-1.06)†
Injury Type			
Hepatocellular	**0.54 (0.40-0.73)*** **	**0.59 (0.40-0.85)** **	0.83 (0.39-1.80)
Mixed	0.79 (0.58-1.06)	**0.67 (0.47-0.96)***	0.76 (0.43-1.36)
Severity: Moderate-severe versus mild	**2.09 (1.54-2.83)*** **	-	-
Log_e of ALT at onset (xULN)	0.96 (0.83-1.11)	0.87 (0.71-1.06)	1.02 (0.64-1.62)
Log_e of ALP at onset (xULN)	**1.50 (1.25-1.81)*** **	1.23 (0.94-1.62)	1.44 (0.80-2.63)
Log_e of bilirubin at onset (xULN)	**1.46 (1.31-1.62)*** **	**1.41 (1.15-1.72)*** **	**1.44 (1.03-1.99)***
Log_e of time to onset (days)	1.12 (1.00-1.26)†	**1.33 (1.13-1.56)*** **	0.95 (0.66-1.36)
Drug exposure (days)	1.00 (1.00-1.00)	1.01 (0.97-1.04)	1.03 (0.98-1.09)
Delayed drug discontinuation	1.02 (0.99-1.05)	-	-
Extent of metabolism ≥ 50%	**0.52 (0.40-0.66)*** **		
Daily Dose	1.00 (1.00-1.00)		
Lipophilicity (LogP)	1.00 (0.92-1.08)		

Univariate accelerated failure time estimates of the percentage differences in time to biochemical recovery. Covariates with a time ratio greater than 1 are associated with a prolonged time to recovery. For example, a time ratio of 0.54 indicates that hepatocellular injury is associated with a 54% decrease in time to recovery, as compared to cholestatic injury. A time ratio of 1.46 implies that an increase in bilirubin will increase time to recovery by 46%. Results are shown in the entire cohort and two subgroups, one comprised of only amoxicillin-clavulanate cases and the other of only flucloxacillin cases. There was not enough data to estimate severity and delayed drug discontinuation in the amoxicillin-clavulanate and flucloxacillin cases. Bolded numbers indicate statistical significance.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, † $p < 0.1$.

Table 3. Multivariate log-normal accelerated failure time model for DILI recovery score. β is used as the coefficients of the recovery score model.

Covariates	β	SE	TR	95% CI	P value
Log_e ALP(xULN) at DILI onset	0.227	0.086	1.25	1.06-1.48	0.008
Log_e Bilirubin(xULN) at DILI onset	0.277	0.054	1.32	1.19-1.47	<0.001
Log_e of Time to Onset (days)	0.161	0.049	1.17	1.07-1.29	0.001
Extent of Metabolism (>50%)	-0.440	0.127	0.64	0.50-0.83	<0.001

Number of observations = 290, Number of events = 140, R-squared = 0.219
Likelihood ratio X^2 test = 72.24 (df = 4, $p < 0.0001$). Scale= 0.729
TR, Time ratio.

Table 4. Clinical characteristics of the high and low risk group for prolonged DILI recovery

	Total (N=290)	Risk Category	
		Low, recovery score ≤ 0.44 (N=53)	High, recovery score > 1.30 (N=59)
Probability of Recovery (Pr, 95% CI)			
3 months	45 (0.37-0.51)	58 (0.38-0.72)	22 (0.08-0.35)
6 months	74 (0.65-0.80)	93 (0.58-0.99)	46 (0.26-0.61)
9 months	85 (0.75-0.91)	93 (0.58-0.99)	54 (0.28-0.70)
Time to Onset			
Mean (SD)	43.2 (±80.7)	28.6 (±27.5)	65.2 (±130.9)
Median (range)	25 (1 - 955)	22 (1 - 150)	33 (5 - 955)
Drug-cessation (n, %)			
at DILI onset	52 (17.7%)	15 (28.3%)	4 (6.8%)
prior to DILI onset	185 (62.9%)	18 (34.0%)	48 (81.4%)
after DILI onset	57 (19.4%)	20 (37.7%)	7 (11.9%)
Follow up (days)			
Mean (SD)	81.8 (±73.6)	50.7 (±43.9)	109.6 (±91.5)
Median (range)	68 (1 - 587)	31 (6 - 181)	86 (6 - 543)
Censored (n, %)	154 (52.4%)	24 (45.3%)	42 (71.2%)
Time to Recovery*			
Mean (SD)	79.2 (±47.9)	51.6 (±44.0)	96.6 (±34.0)
Median (range)	73 (10 - 259)	29 (10 - 147)	89 (55 - 190)
Injury Type (n,%)			
Cholestatic	77 (26.2%)	10 (18.9%)	28 (47.5%)
Hepatocellular	103 (35.0%)	30 (56.6%)	8 (13.6%)
Mixed	114 (38.8%)	13 (24.5%)	23 (39.0%)
Laboratory Parameters at Onset			
ALT (/ULN), mean (SD)	12.8 (±13.6)	12.0 (±18.0)	9.8 (±8.0)
AST (/ULN), mean (SD)	10.0 (±12.1)	7.8 (±6.1)	10.9 (±16.5)
Missing	178 (60.5%)	26 (49.1%)	38 (64.4%)
ALP (/ULN), mean (SD)	2.9 (±2.4)	2.1 (±2.5)	4.3 (±3.3)
Bilirubin (/ULN), mean (SD)	5.2 (±4.7)	1.1 (±1.1)	10.0 (±6.3)
Extent of Metabolism (n,%)			
≥ 50 %	85 (29%)	42 (79%)	2 (3%)
< 50 %	209 (71%)	11 (21%)	57 (97%)

ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase. *Time to recoverey includes only cases that resolved within follow-up time.

Figure legends

Fig. 1. Inclusion and exclusion criteria. Of the 720 patients in the iDILIC cohort, 386 met inclusion criteria. These cases included initial and follow-up serum biochemistries and a RUCAM score of six or more. Cases collected by the Spanish DILI registry were excluded here due to potential conflict with their inclusion in the Spain validation cohort. After cases with a pre-existing liver disease or multiple causal drugs were excluded, 294 cases remained.

Fig 2. Kaplan-Meier cumulative event rates for time to recovery in 294 cases. Cases were divided into high-, indeterminate-, and low-risk groups using the recovery score model. The difference between high- and low-risk groups was significant according to the log-rank test ($p<0.0001$).

Fig 3. Kaplan-Meier cumulative event rates for time to recovery in the Spanish DILI registry cohort. Cases were divided into risk groups using the recovery score model with scores greater than 1.30 assigned to high-risk, scores less than 0.44 assigned to low-risk, and cases between described as indeterminate. Recovery in the high- and low-risk groups was significantly different according to the log-rank test ($p=0.0025$).

Fig 4. Kaplan-Meier cumulative event rates for time to recovery in 191 LiverTox cases. Cases were divided into risk groups using the recovery score model with scores greater than 1.30 assigned to high-risk and scores less than 0.44 assigned to low-risk. Indeterminate cases are shown in gray. The difference between the low- and high-risk groups was significant according to the log-rank test ($p=0.0004$).

Figure 1

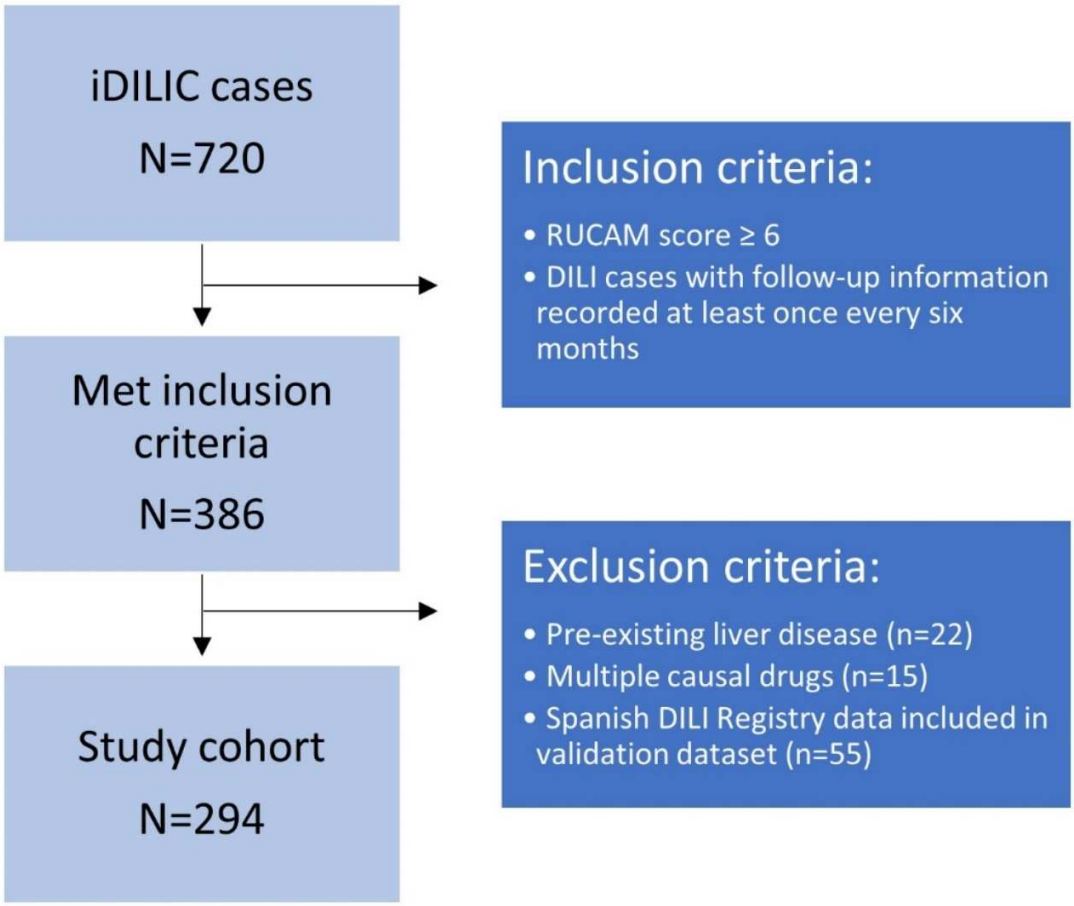
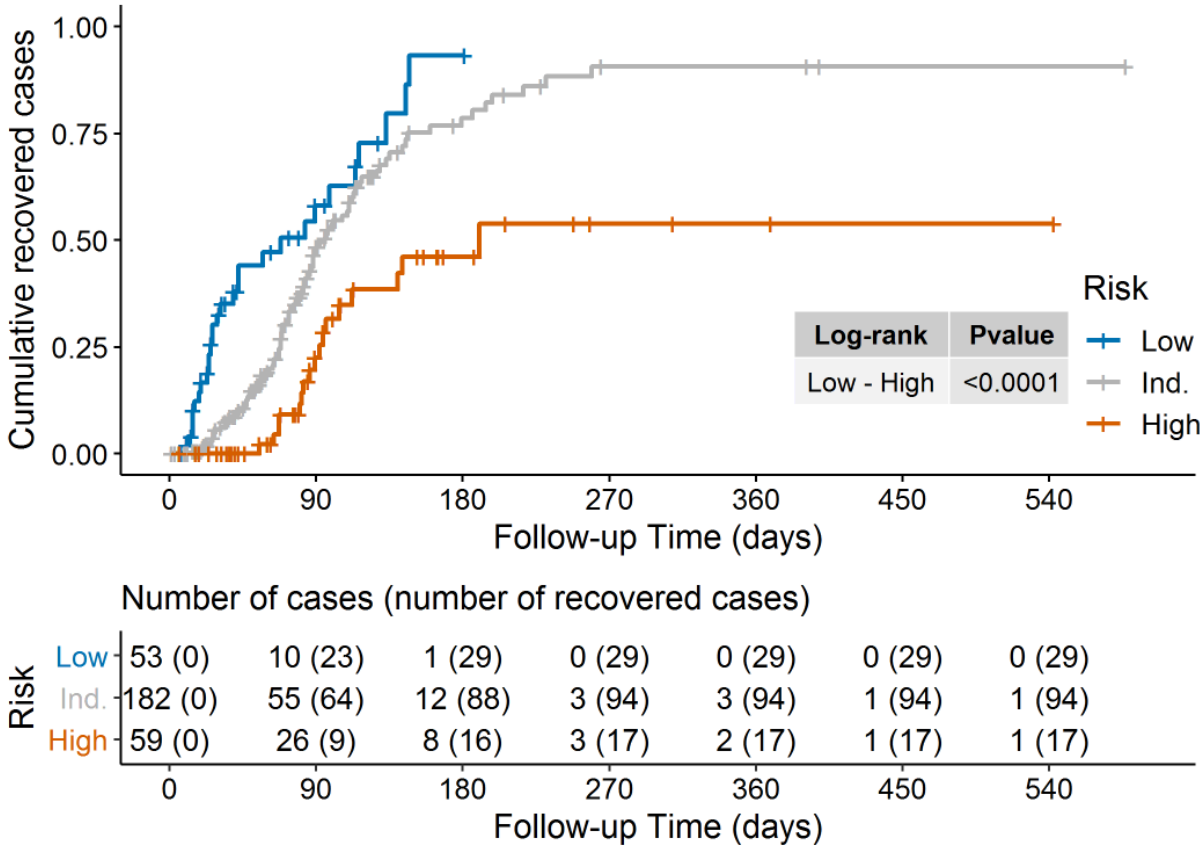

iDILIC cases
N=720
Met inclusion criteria
N=386
Study cohort
N=294
Inclusion criteria:
• RUCAM score ≥ 6
• DILI cases with follow-up information recorded at least once every six months
Exclusion criteria:
• Pre-existing liver disease (n=22)
• Multiple causal drugs (n=15)
• Spanish DILI Registry data included in validation dataset (n=55)

Figure 2



Number of cases (number of recovered cases)

Risk							
Low	53 (0)	10 (23)	1 (29)	0 (29)	0 (29)	0 (29)	0 (29)
Ind.	182 (0)	55 (64)	12 (88)	3 (94)	3 (94)	1 (94)	1 (94)
High	59 (0)	26 (9)	8 (16)	3 (17)	2 (17)	1 (17)	1 (17)

0 90 180 270 360 450 540

Follow-up Time (days)

Figure 3.

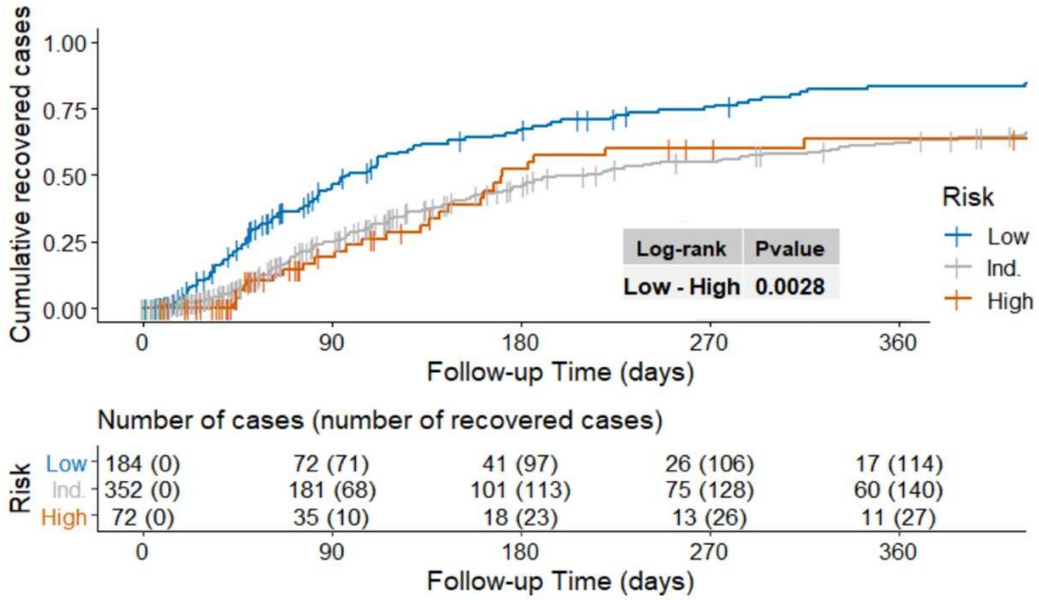


Figure 4