DR. PAUL H HAYASHI (Orcid ID : 0000-0002-5371-1020)

Article type : Original Article

A Revised Electronic Version of RUCAM for the Diagnosis of Drug Induced Liver Injury

Paul H. Hayashi MD, MPH[1*]

M. Isabel Lucena, MD[2*]

Robert J. Fontana, MD[3]

Einar S Bjornsson, MD[4, 5]

Guruprasad P. Aithal, MD, PhD[6]

Huiman Barnhart, PhD[7]

Andres Gonzalez Jimenez, PhD[2]

Qinghong Yang, PhD[7]

Jiezhun Gu, PhD[7]

Raul J Andrade, MD[2]

Jay H. Hoofnagle, MD[8]

[1]Food and Drug Administration, Silver Spring, MD, USA
[2] Servicio de Aparato Digestivo y de Farmacología Clínica, Hospital Universitario Virgen de la Victoria, Instituto de Investigación Biomédica de Málaga-IBIMA, Facultad de Medicina, Universidad de Málaga, Málaga, Spain. Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), UICEC-IBIMA, Plataforma Investigación Clínica, Madrid, Spain
[3]University of Michigan-Ann Arbor, MI, USA
[4]Landspitali University Hospital, Reykjavik, Iceland
[5]Faculty of Medicine, University of Iceland, Reykjavik Iceland
[6]NIHR Nottingham BRC, Nottingham University Hospitals NHS Trust and the University of Nottingham, United Kingdom
[7]Duke University, Durham, NC, USA
[8]National Institute of Diabetes and Digestive and Kidney Diseases-Bethesda, MD, USA
*These two authors contributed equally to this research.
Corresponding author:
Paul H. Hayashi, MD, MPH, FAASLD
10903 New Hampshire Avenue
CDER/OND, Division of Hepatology and Nutrition
WO-22/5387
Silver Spring, MD 20993
Paul.Hayashi@fda.hhs.gov

Abbreviations:

AASLD: American Association for the Study of Liver Diseases; ALT: alanine aminotransferase; ANA: anti-nuclear antibody; ALP: alkaline phosphatase; ASMA: anti-smooth muscle antibody; AST: aspartate aminotransferase; AUC: area under the receive operator curve; CIOMS: Council of

International Organizations of Medical Sciences; CMV: cytomegalovirus; CT: computerized tomography; DILI: drug-induced liver injury; DILIN: Drug-Induced Liver Injury Network; DRESS: drug reaction with eosinophilia and systemic symptoms; EBV: Epstein-Barr Virus; HAV: hepatitis A virus; HBc: hepatitis B core; HBsAg: hepatitis B surface antigen; HCV: hepatitis C; HDS: herbal and dietary supplements; HEV: hepatitis E virus; HSV: herpes simplex virus; IgG: immunoglobulin G; MRI: magnetic resonance imaging; NIDDK: National Institute of Diabetes and Digestive and Kidney Diseases; RECAM: Revised Electronic Causality Assessment Method; RUCAM: Roussel Uclaf Causality Assessment Method; SIRS: systemic inflammatory response syndrome; SJS: Stevens Johnson syndrome; ULN: upper limit of normal; US: ultrasound

Funding: The Drug-Induced Liver Injury (DILN) Network is structured as a U01 cooperative agreement with funds provided by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under grants: U24-DK065176 (Duke), U01-DK065201 (UNC), U01-DK065184 (Michigan), U01- DK065211 (Indiana), U01DK065193 (UConn), U01-DK065238 (UCSF/CPMC), U01-DK083023 (UTSW), U01-DK083027 (Einstein), U01-DK082992 (Mayo), U01-DK083020 (USC), U01-DK100928 (Mount Sinai). Additional support is provided by CTSA grants: UL1 RR025761 (Indiana), UL1TR000083 (UNC), UL1 RR024134 (UPenn), UL1 RR024986 (Michigan), UL1 RR024982 (UTSW), UL1 RR024150 (Mayo) and by the Intramural Research Program of the NIH, National Cancer Institute. Clinical Trials.gov number: NCT00345930. The Spanish DILI Registry is funded by grants of Instituto de Salud Carlos III cofounded by Fondo Europeo de Desarrollo Regional - FEDER (contract numbers PI 18/01804; PT 20/00127) and Agencia Española del Medicamento. Plataforma ISCiii de Investigación Clínica and CIBERehd are funded by ISCIII. 2016 American Association for the Study of Liver Disease (AASLD) Innovations Fund.

Acknowledgements:
"This publication is partially based upon work from COSTAction "CA17112 - Prospective European Drug-Induced Liver Injury Network" supported by COST (European Cooperation in Science and Technology) www.cost.eu. MIL, ESB, GPA, AGJ and RJA are members of COST Action CA17112.

We appreciate Dr Lauryna Aukštikalnė for her help in refining the online implementation of the RECAM.

We thank the American Association for the Study of Liver Disease (AASLD) for supporting this research through the AASLD Innovations Fund (2016).

Conflicts and disclaimers regarding this manuscript:

Guruprasad P. Aithal has advised Pfizer, GSK and NuCANA through Nottingham University Consultants.

Raul J Andrade has no conflicts of interests.

Huiman Barnhart has no conflict of interests.

Einar S. Bjornsson has no conflicts of interests.

Robert J. Fontana has received research support from Gilead and AbbVie and consulted for Sanofi.

Andres Gonzalez-Jimenez has no conflicts of interests.

Jiezhun Gu has no conflicts of interests.

Paul H. Hayashi has no conflicts of interests. He is employed by the US Food and Drug Administration (FDA), but conclusions in this paper do not reflect any opinion of the FDA. The FDA has not evaluated the findings, assessments or conclusions in this paper.

Jay H. Hoofnagle has no conflicts of interests.

M Isabel Lucena has no conflicts of interests.

Qinghong Yang has no conflicts of interests.

Abstract

Background and Aims: Roussel Uclaf Causality Assessment Method (RUCAM) for drug-induced liver injury (DILI) has been hindered by subjectivity and poor reliability. We sought to improve the RUCAM using data from the Drug-Induced Liver Injury Network (DILIN) and the Spanish DILI Registry, published literature and iterative computer modelling.

Approach and Results: RUCAM criteria were updated, clarified and computerized. We removed criteria 3 (risk factors) for lack of added value and criteria 4 because we felt it more useful to assess each drug separately. Criteria 6 (drug specific risk) was anchored to LiverTox® likelihood scores. Iterative testing in subsets of 50-100 single agent, non-herbal cases from both registries was done to optimize performance. We used classification tree analysis to establish diagnostic cut-offs for this revised electronic version (RECAM) and compared RECAM with RUCAM for correlation with expert opinion diagnostic categories in 194 DILI cases (98 DILIN, 96 Spanish DILI). Area under receiver operator curves (AUC) for identifying at least probable DILI were the same at 0.89 for RECAM and RUCAM. However, RECAM diagnostic categories have better observed overall agreement with expert opinion (0.62 vs. 0.56 weighted kappa, $p = 0.14$), and had better sensitivity to detect extreme diagnostic categories (73 vs. 54 for highly likely or high probable, $p=0.02$; 65 vs. 48 for unlikely/excluded, $p = 0.08$) than RUCAM diagnostic categories.

Conclusions: RECAM is an evidenced based update that is at least as capable as RUCAM in diagnosing DILI compared to expert opinion but is better than RUCAM at the diagnostic extremes. RECAM's increased objectivity and clarity will improve precision, reliability and standardization of DILI diagnosis but further refinement and validation in other cohorts are needed.

Introduction

The diagnosis of drug-induced liver injury (DILI) is primarily based on clinical judgment and the elimination of alternate diagnoses. Lack of an evidence-based and reliable diagnostic tool is a significant hindrance to clinical care and research. In 1993, Danan and Benichou published the Rousell Uclaf Causality Assessment Method (RUCAM, also credited to CIOMS, Council of International Organizations of Medical Sciences), which is a diagnostic scorecard based on 7 clinical criteria.[1] It is the most widely used and accepted DILI diagnostic tool.[2, 3] However, clinical and research usefulness is still debated.[4, 5]

Since 1993, there have been three major problems with RUCAM: (1) unclear operating instructions and subjectivity leading to poor reliability and usability, (2) unclear validity due to lack of an accepted gold standard and (3) domain criteria that are not evidence-based.[6] Even the updated RUCAM, which is quite similar to the original, retains a significant degree of subjectivity in terms of ruling out competing diagnoses.[7] Nevertheless, RUCAM's criteria include most of the critical elements needed to make a diagnosis of DILI, thus providing a framework for evaluation. Despite its limitations, this framework has led to RUCAM's durability in publications. However, establishing a causal relationship between exposure to an agent and the appearance of liver injury remains the Achilles heel in DILI research, and improved standardization, automation and reproducibility in causality assessment are needed.

Using an evidence-based approach, we sought to revise the RUCAM, with an aim of having an instrument that not only had criterion and construct validity against the current RUCAM but improved precision and reproducibility. We used data from two large prospective DILI registries of well-vetted cases, the US Drug-Induced Liver Injury Network (DILIN) and the Spanish DILI Registry, to refine and develop instrument domains and scoring. We then piloted the instrument in randomly selected cases to determine the instruments performance properties in comparison to RUCAM.

Methods

Process overview: Since 2015, the authors met regularly to modify the RUCAM criteria using data from the DILIN and Spanish Registry cases. In addition, a review of the published literature and expert opinion were used when robust data were lacking. The development was restricted to provide assessment of single medication cases because full separate assessment for each competing agent would have been needed to achieve reliable scoring. Herbal and dietary supplements (HDS) product cases were also excluded due to the uncertainty of product contents and less well established causality assessment methods.

The new instrument was developed through 5 sequential stages: (1) Each of the 7 RUCAM criteria were separately analyzed and revised to optimize diagnostic scoring. Registry data for latency and dechallenge were robust and well-suited to optimize cut-off values and scores. Contrary to expectations, distinction between hepatocellular and cholestatic/mixed injuries was not necessary for latency and dechallenge scoring. Other criteria changes were based on a combination of registry data, expert opinion and available literature; (2) the revised criteria, renamed domains, were tested for ability to detect at least probable drug-induced liver injury cases in the DILIN. During this stage, revisions were made including elements added or discarded based on performance contribution; (3) computer programming was applied to extract data directly from the DILIN database and Spanish Registry with single agent DILI cases of varying levels of prior causality scoring. We assessed concordance of computer scoring with human scoring to ensure proper computer programming; (4) the revised electronic causality assessment method (RECAM) scored groups of 50-100 single medication cases from the DILIN stratified equally on DILIN's 5 expert diagnostic categories (see next paragraph). Scoring outputs were used to revise the RECAM and programming to optimize performance. (5) RECAM was then applied to groups of 50-100 single agent, non-HDS cases randomly selected from the Spanish DILI Registry to assess instrument performance including domain validity and comparison of scoring obtained with RUCAM. Through this final phase, the RECAM went through modifications by an iterative process of testing both DILIN and Spanish-DILI cases. RECAM was applied across the range of DILI likelihood categories used by the DILIN and Spanish DILI Registry. Throughout the process, an emphasis was placed on clarity, performance and

precise language that would be adaptable to a clinically useful website application with minimal subjective opinion from the user.

Likelihood Categories and Causality Assessment in the DILIN: DILIN uses a consensus expert opinion method of causality assessment previously described.[8] Each case was evaluated by 3 DILIN hepatologists who independently assigned an ordinal causality score or category representing percent likelihood of attribution (1 = definite or > 95% likelihood, 2 = highly likely or 75-95%, 3 = probable or 50-74%, 4 = possible or 25-49%, and 5 = unlikely or < 25%). Consensus was reached by e-mail and monthly conference calls. The enrolling DILIN investigator also provides a RUCAM score for each case.

Likelihood Categories and Causality Assessment in the Spanish DILI Registry: Each case referred to the Spanish Registry was independently assessed and adjudicated by at least 3 expert investigators. Expert opinion is used to assess whether DILI consideration was reasonable and further data requested from the referring providers as needed. Case likelihood categorization is based on traditional RUCAM categories, but expert opinion can over-ride the RUCAM assigned category as necessary (e.g. drugs with long half-lives and known long latencies after drug stop, mandatory testing of hepatitis E).[9,10]

RECAM and RUCAM Performance in Diagnosing DILI in DILIN and Spanish DILI Registry

A total of 100 and 96 single agents, non-HDS cases from the DILIN and Spanish-DILI, respectively were randomly chosen for testing the 12th and final version of RECAM. We used the R-value ([ALT/ULN] ÷[ALP/ULN]) to categorize cases as hepatocellular (R >5), cholestatic (R <2) or mixed (2< R <5). The DILIN cases were stratified equally across its 5 likelihood categories. One DILIN case was excluded due to data entry error in DILIN adjudication requiring re-assessment and another DILIN case was excluded due to an indirect, atypical liver injury of drug induced sphincter of Oddi dysfunction. Therefore, 98 DILIN cases were used for RECAM scoring.

RECAM scoring was undertaken via semi-automated computer data extraction and scoring from both registries. Computer programming used software version 9.4 and R language version 4.02 for the DILIN cases where R language version 3.5.0 was used for Spanish Registry cases. However, both registry databases contain free text fields (e.g., imaging, histology findings) that required some human interpretation and input for the computer to score the RECAM correctly.

Area under the curves (AUCs) and Diagnostic Cut-offs for RECAM: For the purposes of comparing performance between registries and combining data, the DILIN definite and highly likely cases were combined and considered equivalent to the highly probable Spanish cases. Similarly, the unlikely and excluded cases in the Spanish Registry were combined and considered equivalent to the DILIN unlikely cases. The other category labels of probable and possible are the same in both registries. AUC values were generated for both RECAM and RUCAM scores. RECAM and RUCAM AUC values for identification of at least high probable (or at least highly likely), at least probable and at least possible DILI were determined for both registries. Overall, correlation of RECAM and RUCAM to DILIN and Spanish Registry expert opinion diagnostic categories was assessed by using Spearman's Rho coefficient.

Using the combined DILIN and Spanish Registry data, we built a classification tree[11] based on RECAM scores to obtain three cut-offs for classifying each case into four categories: highly likely/high probable, probable, possible and unlikely/excluded. Performance of RECAM classification based on these cut-offs was compared to the performance of RUCAM classification based on its published cut-off scores of highly probable (≥9), probable (8 to 6), possible (5 to 3), unlikely/excluded (≤2). We tested the overall percent agreement, and Cohen's weighted kappa coefficient between the RECAM and RUCAM scales with expert's opinion. Diagnostic performance, sensitivity and specificity values were calculated for the diagnostic categories. P-values are reported for testing the equality of agreement metrics (overall agreement, sensitivity and specificity) of RECAM and RUCAM diagnostic categories with expert's opinion via the generalized estimating equations and for testing equality of weighted Kappa statistics of RECAM and RUCAM diagnostic

categories with expert's opinion via bootstrap approach to account for correlation of RECAM and RUCAM diagnostic categories within the same subject.

Role of Funding Source and Institutional Board Review (IRB): An American Association for the Study of Liver Diseases (AASLD) Innovations Fund awarded in 2016 supported this project. The DILIN Network is structured as a U01 cooperative agreement with funds provided by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Separate IRB approvals were maintained at each center throughout study participation. The Spanish DILI Registry is funded by competitive grants from National Health Institute-FEDER and the Spanish Medicines Agency. All patients enrolled in both registries provided written informed consent.

Results

RUCAM Modifications for RECAM Development:

The 7 original RUCAM Criteria (Supplement Table 1) were modified, reordered and renamed as Domains. The resulting 5 domain RECAM is shown in Table 1. Below, we describe how each RUCAM criteria was modified and resulted in the 5 domain RECAM.

Criteria 1 (Time to onset): We retained latency from both drug start and stop to form Domains 1a and 1b, but time intervals for scoring revised and the need to stratify by type of liver injury (determined by R-value) was eliminated (Table 1a). The original RUCAM was unclear as to whether both or only one latency is to be scored. Unlike the updated RUCAM which scores one or the other,[7] the RECAM requires both latencies from drug start (Domain 1a) and stop (Domain 1b) to be scored. However, latency after stopping drug can only hurt the case for DILI by subtracting up to 6 points. Some drugs (e.g. monoclonal antibodies) clearly have long half-lives and long latency to DILI. For these drugs, Domain1b is passed over with no points taken. Cut-offs for point allocations were based on the DILIN and Spanish registry latency data across likelihood categories. Time intervals were expanded creating a wider range of scores compared to RUCAM.

Criteria 2 (Course): In RUCAM, dechallenge time cut-offs and scoring are different for hepatocellular and cholestatic/mixed cases. Based on analysis of DILIN and Spanish Registry data, dechallenge timing is similar for hepatocellular and cholestatic/mixed cases in terms of causality. Therefore, dechallenge time cut-offs are the same regardless of R-value and were based on the observed distribution of dechallenge times across definite to unlikely DILIN cases. R-value still defines which liver biochemistry to use for dechallenge scoring. Hepatocellular injury cases follow the course of ALT, while cholestatic and mixed injury cases follow alkaline phosphatase or total bilirubin whichever yields a higher score (Table 1a). This modified dechallenge criteria became Domain 2.

Criteria 3 (Risk factors): For the standard RUCAM and new RECAM, these 3 variables did not contribute significantly to logistic regression modeling to diagnose at least probable DILIN cases (age, odds ratio [OR] 1.12 (95% CI 0.71-1.76), p = 0.62; alcohol and pregnancy, OR 0.90 (0.47-1.73), p = 0.75). This lack of Domain 3 contribution coincided with expert opinion and clinical experience of the group.[12, 13] Therefore, Criteria 3 (Risk Factors) was eliminated.

Criteria 4 (Concomitant drugs): We reasoned that concomitant medications of clinical significance should be scored separately for simplicity and reliability of scoring. The assessment of competing drugs in RUCAM is prone to subjectivity (e.g. "suggestive" timing, "known as hepatotoxin") and does not provide detailed assessment for these agents (Supplemental Table 1).[1] Therefore, we limited this revised RUCAM to assess drugs individually, and these concomitant drug criteria are not included in the RECAM.

Criteria 5 (Search for Non-Drug Causes): This RUCAM Criteria became RECAM Domain 4 (Table 1b). All competing diagnoses in the RUCAM were retained, but HEV, congestive hepatopathy, infiltrating cancer and cholestasis of sepsis based on what is considered necessary evaluation testing in the literature was added.[14, 15] We chose to only penalize for competing diagnoses because DILI is a diagnosis of exclusion where competing causes should only hurt the case. All diagnoses in this Domain should be addressed. At this point, the RECAM will suggest obtaining these data before proceeding. Otherwise, points are taken away for missing such information. Specific tests and

scoring instructions are provided to minimize subjectivity. Viral tests are specified, including HEV antibodies. Evaluation for acute hepatitis C include HCV RNA, history of prior hepatitis C and risk factors. The RECAM provides scores based on pre-specified test results. Consideration for alcoholic hepatitis diagnosis is prompted by the AST:ALT ratio and AST less than 500 U/L. Only if prompted, will the user need to enter information about the amount of alcohol use. Imaging data are clarified with 3 binary questions based on evidence of pancreaticobiliary disease, and cancer infiltration. Autoimmune marker interpretation was aligned more closely with the simplified autoimmune hepatitis score[16] but also scored differently for certain medications known to cause DILI with autoimmune marker positivity.

Criteria 6 (Previous Information on Hepatotoxicity of the Drug): We moved these criteria to Domain 3 reasoning that most clinicians seek this information early in their consideration of DILI. To increase objectivity and reliability, scoring was anchored to LiverTox® likelihood scores[17] which are loaded into the RECAM. Based on iterative performance testing, the likelihood scores were grouped into 3 categories of LiverTox® likelihood scores (Table 1a), and the RECAM will automatically input the corresponding score upon entering the implicated medication. If an agent is not listed in LiverTox® (e.g., flucloxacillin), then the user will be given the opportunity to assign a score of 0, 1 or 3 (Table 1a).

Criteria 7 (Response to Readministration): Because rechallenge was so infrequent in both registries and clinical practice, these criteria were incorporated as part of a new Domain 5 of additional (optional) data (Table 1c). We distinguish between a rechallenge prospectively documented with laboratory testing and a retrospective rechallenge which is elicited in a patient history only and laboratory data may be lacking. We provide specifics on scoring each. Rechallenge is infrequent, but a positive prospective rechallenge is highly indicative of DILI and awarded more points than any other component in the RECAM (+6).

RECAM Domain 5 (Additional Data):

Besides rechallenge, liver histology, atypical viral testing and presence of severe skin reactions were newly included in Domain 5. Liver histology is uncommonly diagnostic of DILI, so points awarded were limited. However, the case is penalized heavily if the biopsy findings yielded an obvious competing diagnosis (Table 1c). The presence of severe cutaneous drug reactions adds a point. The presence of non-hepatotropic viral infection, for which testing should be done according to clinical context (e.g., fever, lymphadenopathy, immunocompromise), leads to loss of points.

RECAM warnings and stops: When a firm alternate diagnosis or inconsistent timing for DILI is evident, the user is warned to stop with a -6 final score automatically rendered. The user may over-ride this warning, but -6 points will be deducted from the overall score, and the user should recognize that DILI as sole cause of liver injury is questionable due to a competing explanation or inconsistent timing, regardless of total score obtained.

RECAM and RUCAM Performance:

RECAM went through 12 versions based on iterative testing of cases and meetings. The RUCAM and final version of RECAM scoring was done on 98 DILIN and 96 Spanish DILI cases. Characteristics of each cohort are shown in Table 3. Spanish cases were older and had a greater proportion of probable cases. The DILIN had more definite and highly likely cases compared to the Spanish Registry. Supplemental Table 2 shows the most common medications implicated.

Both RECAM and RUCAM had similarly high statistical correlation between the resulting scores and the four ordinal diagnostic categories provided by experts (Spearman Rho 0.85, $p<0.001$ and 0.87, $p<0.001$, respectively). By using classification tree approach, we estimated RECAM diagnostic cut-offs of ≥ 8 for highly likely/high probable, 7 to 4 for probable, 3 to -3 for possible, and ≤ -4 for unlikely/excluded DILI, respectively. Classification of combined DILIN and Spanish Registry cases along diagnostic categories using the RECAM and traditional RUCAM cut-off are shown by boxplots

in Figure 1. In a stratified analysis by separate cohorts, the 96 Spanish DILI cases were better classified when using the RECAM compared to the 98 DILIN cases (Supplemental Figure 1). The AUCs for cumulative cut-offs in likelihood category for both cohorts combined are shown in Table 4. RECAM and RUCAM performed similarly well across all three cut-offs (AUC > 0.8 in all likelihood categories). In a stratified analysis by cohort the RECAM and the RUCAM scale AUCs showed better performance in Spanish DILI cases compared to DILIN cases. For the Spanish cases, the RECAM AUCs ranged from 0.95 (at least probable cases) to 0.99 (at least possible cases), while in DILIN cases AUCs ranged from 0.80 (at least possible cases) to 0.86 (at least probable cases) (Supplemental Table 3).

The overall percent agreements between the RECAM and RUCAM scales with expert's opinion were 62% and 59%, respectively (p=0.44). By Cohen's weighted Kappa coefficient, RECAM had better observed overall agreement compared to RUCAM (0.62 vs 0.56), although statistical significance was not reached (p=0.16) (Table 4). The RECAM had a markedly greater sensitivity for classifying extreme likelihood categories of high likely/high probable and unlikely/excluded. Both scales showed great and similar specificity along likelihood categories, except for probable cases, where the RECAM scale showed better performance (Table 4).

Discussion

This revised electronic causality assessment method, RECAM, provides an evidence-based update of RUCAM. Both RECAM and RUCAM had good diagnostic performance in classifying cases across varying cut-offs in likelihood of DILI based on expert opinion in two large DILI registries. However, RECAM tended to have better observed overall agreement with expert opinion and to better discriminate diagnostic categories especially at the extremes, i.e., highly likely/probable and unlikely. It also had greater specificity to correctly classify probable cases. These differences were likely due to a wider scoring range for latency and dechallenge that was developed from case data and the heavier penalization for lack of data or data indicating a non-DILI diagnosis. RECAM also offers an automated scoring with less subjective input which should lead to better inter-rater reliability.

Computerization of RECAM (http://gihep.com/dili-recam/) is important because RUCAM's poor inter-rater reliability has limited its adaptation in clinical practice and research. The RECAM categorically scores test results, latency, dechallenge, medication specific DILI risk and most competing diagnoses without the need for subjective user opinion or knowledge. The user merely enters the objective data of dates, lab values and test results. The only subjective information needed for Domains 1 through 4 are the presence of biliary obstruction, >50% malignant liver infiltration on imaging, sepsis, shock or congestive hepatopathy as these defied consistent objective parameters for computer entry. Similarly, subjective opinion in Domain 5 is limited to histology and presence of drug reaction with eosinophilia and systemic symptoms (DRESS) or Steven Johnson Syndrome.

The heterogeneity of DILI phenotypes makes it difficult to develop a single, easy-to-use diagnostic tool for all medications. Thus, the RECAM did not completely mirror expert opinion for a variety of reasons. Firstly, experts rely on knowledge of recent DILI research and emerging phenotypes that can be difficult to translate into algorithmic scoring. Second, some patients had symptoms but delayed seeking medical care artificially lengthening the latency or had DILI due to agents with prolonged latencies of months to years (e.g., nitrofurantoin, minocycline). Experts correctly adjusted their opinion of what RECAM considered a latency too long for DILI. Finally, death, transplant, and chronic DILI also prevented receipt of dechallenge points, while experts accounted for typical cases of fatal or chronic DILI. Inability to capture such clinical factors into the RECAM led to the overall, complete agreement rate of just 62% (59% for RUCAM), but the AUCs of 0.87 to 0.89 across diagnostic category cut-offs which are quite good and competitive with other clinical diagnostic tools. For the clinician, the cut-off of at least probable may be most useful when weighing the risks of rechallenge with a highly needed medication or need for further diagnostic evaluation. RECAM's AUC of 0.89 and better ability to separate diagnostic categories (Figure 1) provide a useful framework for such decision making. The improved stratification may better classify cases for genetic (e.g., HLA) and other DILI biomarker development, and increased consistency will make it a better teaching tool.

RECAM's remarkably high AUCs in the Spanish DILI Registry (Supplemental Table 3) provide some criterion validity as the Spanish experts rely more on RUCAM for their diagnostic categories.

The high performance suggests enough retained similarity to support RECAM's application to that Registry and others currently based on RUCAM. The comparable AUCs for RUCAM and RECAM also confirms that the risk factors of age ≥55, alcohol intake, and pregnancy do not add value to the diagnosis of DILI (Supplemental Case 1) and suggests that the 5-domain RECAM without differentiation between hepatocellular and cholestatic/mixed injury is adequate. RECAM's separation of diagnostic categories, especially unlikely and excluded cases, was also better in Spanish cases (Supplemental Figure 1c) possibly due to the fact that the DILIN often excludes cases that have definitive competing diagnoses prior to enrollment, while the Spanish group retains such cases in their data analyses.

The RECAM has several other notable improvements. The elimination of alternate diagnoses only prevents a loss of points because ruling out competing etiologies does not directly support a DILI diagnosis in the same way as latency and dechallenge do. The RECAM has automatic warnings for data inconsistent with DILI, which is not a part of RUCAM. In the RUCAM, an alternate diagnosis or other data could rule out DILI, but the case would still gain points in other criteria (Supplemental Cases 2 and 3). Even when data clearly diagnose acute viral hepatitis or autoimmune hepatitis by simplified autoimmune hepatitis score [16] points are still given for latency, dechallenge or underlying hepatotoxicity risk of the drug. In these situations of highly implausible DILI, RECAM gives warnings to stop with an imputed total score of -6. One can over-ride these warnings, if one believes DILI may be concurrent with the non-DILI diagnosis. However, -6 points are still assessed. Similarly, warnings to consider stopping or proceeding with a -3 penalty occur when critical data are missing. Such prompts firmly remind the user of tests needed during DILI evaluation. These stops and penalizations led to downward distribution of scores in both registries, particularly unlikely or excluded cases.

The RUCAM assigns a single point for any latency from 5 to 90 days after drug start, while the RECAM has 3 different scores within the span of 2 to 90 days regardless the type of liver injury. Gradation of cut-offs was increased for latency times based on latencies in DILIN cases, expert opinion and iterative testing of cases. This may have led to better identification of highly likely or high probable cases (Supplemental Case 3). A pre-assessment DILI risk score (Domain 3) for specific

medications is automatically assigned based on LiverTox® likelihood score, thus clarifying one of the more ambiguous domains in RUCAM.[18] These changes also may have helped RECAM better identify more of the highly probable cases.

Incorporating liver histology into a categorical scoring system was challenging. Certain findings may be quite consistent with a specific DILI episode (e.g., ring granulomas with allopurinol liver injury), but we felt even these readings are open to interpretation and need clinical context. Thus, only 1 point is awarded for histologic findings, but histology can hurt the case for DILI when a clear alternate diagnosis is found like infiltrating cancer or ischemic injury. In these cases, a heavy penalty of -6 and warning are given. In both registries, liver biopsy was often not obtained, and pathognomonic signs of DILI or alternate diagnosis were even less common. Therefore, the impact of histology on RECAM performance was minimal. Nevertheless, the computer program used to develop the RECAM will allow us to adjust this variable as more data on how histology influences the diagnosis of DILI become available.[19]

The RECAM also has several important limitations. It was developed in US and Spanish cohorts, so we do not know how it may perform in other regions, particularly Asia. Also, both registries have minimum enrollment criteria for liver enzyme and bilirubin elevation, so it is unclear how the RECAM may perform in less severe cases.[8, 10] The RECAM also needs testing by a broader group of clinicians including non-hepatologists. It is currently limited to single agent medication cases leaving the user to score each medication individually in multi-drug cases. However, any competing medication causing loss of points in the RUCAM, probably deserves its own RECAM score. The RECAM is also not designed nor tested for HDS liver injury which is increasingly reported.[20-22] While simplified with fewer Domains and clearer operating instructions, the web application increases the amount of data entry compared to the RUCAM. Yet, we believe the increased data entry will be offset by automated latency and dechallenge calculations by the computer. Also, users no longer need to render a subjective opinion on competing diagnoses. They simply choose test results regarding competing diagnoses from short dropdown menus. The RECAM retains a few parameters that need clinical judgement. Whether a biliary stricture is clinically insignificant is still left up to the user. Drugs not included in LiverTox® must still be scored by opinion of labeling and available

literature. Finally, the RECAM will need updating as DILI epidemiology and research evolve. For example, the cutoff of >90-day latency garnering 0 points was based on a broad range of cases with most having shorter latencies, but as longer acting medications (e.g., monoclonal check point inhibitors) grow in use and latencies increase, this cut off may need adjustment. Pharmacogenomic data and new biomarkers may also need to be incorporated with the computerization of RECAM lending itself well to such modifications.

RUCAM has been a valuable clinical framework for DILI diagnosis since 1993. However, user subjectivity made it unreliable, and it was overdue for an evidence-based update. RECAM has better sensitivities at the extreme diagnostic categories and tends to have better overall agreement with expert opinion. It will likely have better inter- and intra-rater reliability due to computerized categorical, data entry and minimized subjective opinion. The RECAM also eliminates unnecessary variables that were not diagnostically helpful. Domains are based on data from well-vetted cases that were often followed for a minimum of 6 months. Accuracy of 80-90% for identifying at least probable DILI compared to expert opinion is high, but not high enough to make the RECAM a standalone diagnostic tool. For now, nothing can replace good history taking, chart review, and thorough evaluation for competing causes. There will always be cases that defy proper scoring by any single algorithm that seeks to account for the extensive heterogeneity in DILI phenotypes and presentation (e.g., very long latency DILI, chronic DILI). Therefore, further refinement and validation are anticipated. Indeed, the RECAM provides an opportunity to conduct causality assessment using standardized, quantitative and categorical data fields which should lead to improved case identification, earlier diagnosis, and medical management. The electronic, automated platform of the RECAM that is available for all to use on the Internet should also help with efforts at harmonization and standardization in DILI research.

References

1. Danan G, Benichou C. Causality assessment of adverse reactions to drugs--I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. J Clin Epidemiol 1993;46:1323-30.
2. Chalasani NP, Maddur H, Russo MW, et al. ACG Clinical Guideline: Diagnosis and Management of Idiosyncratic Drug-Induced Liver Injury. Am J Gastroenterol 2021;116:878-898.
3. European Association for the Study of the Liver. Electronic address eee, Clinical Practice Guideline Panel C, Panel m, et al. EASL Clinical Practice Guidelines: Drug-induced liver injury. J Hepatol 2019;70:1222-1261.
4. Kaplowitz N. Causality assessment versus guilt-by-association in drug hepatotoxicity. Hepatology 2001;33:308-10.
5. Hayashi PH. Causality assessment in drug-induced liver injury. Semin Liver Dis 2009;29:348-56.
6. Garcia-Cortes M, Stephens C, Lucena MI, et al. Causality assessment methods in drug induced liver injury: strengths and weaknesses. J Hepatol 2011;55:683-691.
7. Danan G, Teschke R. RUCAM in Drug and Herb Induced Liver Injury: The Update. Int J Mol Sci 2015;17.
8. Fontana RJ, Watkins PB, Bonkovsky HL, et al. Drug-Induced Liver Injury Network (DILIN) prospective study: rationale, design and conduct. Drug Saf 2009;32:55-68.
9. Andrade RJ, Lucena MI, Fernandez MC, et al. Drug-induced liver injury: an analysis of 461 incidences submitted to the Spanish registry over a 10-year period. Gastroenterology 2005;129:512-21.
10. Stephens C, Robles-Diaz M, Medina-Caliz I, et al. Comprehensive analysis and insights gained from long-term experience of the Spanish DILI registry. J Hepatol 2021.
11. Breiman LF, J. H.; Olshen, R. A.; Stone, C. J. Classification and Regression Trees. Wadsworth, Belmont, CA 1984.
12. Lucena MI, Andrade RJ, Kaplowitz N, et al. Phenotypic characterization of idiosyncratic drug-induced liver injury: the influence of age and sex. Hepatology 2009;49:2001-9.
13. Dakhoul L, Ghabril M, Gu J, et al. Heavy Consumption of Alcohol is Not Associated With Worse Outcomes in Patients With Idiosyncratic Drug-induced Liver Injury Compared to Non-Drinkers. Clin Gastroenterol Hepatol 2018;16:722-729 e2.
14. Aithal GP, Watkins PB, Andrade RJ, et al. Case definition and phenotype standardization in drug-induced liver injury. Clin Pharmacol Ther 2011;89:806-15.

15. Agarwal VK, McHutchison JG, Hoofnagle JH, et al. Important elements for the diagnosis of drug-induced liver injury. Clin Gastroenterol Hepatol 2010;8:463-70.
16. Hennes EM, Zeniya M, Czaja AJ, et al. Simplified criteria for the diagnosis of autoimmune hepatitis. Hepatology 2008;48:169-76.
17. LiverTox. Clinical and Research Information on Drug-Induced Liver Injury [Internet]. Bethesda (MD): National Institute of Diabetes and Digestive and Kidney Diseases. (Accessed Oct 30, 2020);https://www.ncbi.nlm.nih.gov/books/NBK547852/.
18. Bjornsson ES, Hoofnagle JH. Categorization of drugs implicated in causing liver injury: Critical assessment based on published case reports. Hepatology 2016;63:590-603.
19. Ahmad J, Bonacini, M., Ghabril, M.S., Hayashi, P.H., Odin, J., Rockey, D.C., Rossi, S., Serrano, J., Tillmann, H.L., Kleiner, D.E., Drug-Induced Liver Injury Network Investigators. The Value of Liver Biopsy in the Diagnosis of Drug-Induced Liver Injury. Hepatology 2020;72:730A.
20. Navarro VJ, Barnhart H, Bonkovsky HL, et al. Liver injury from herbals and dietary supplements in the U.S. Drug-Induced Liver Injury Network. Hepatology 2014;60:1399-408.
21. Medina-Caliz I, Garcia-Cortes M, Gonzalez-Jimenez A, et al. Herbal and Dietary Supplement-Induced Liver Injuries in the Spanish DILI Registry. Clin Gastroenterol Hepatol 2018;16:1495-1502.
22. Shen T, Liu Y, Shang J, et al. Incidence and Etiology of Drug-Induced Liver Injury in Mainland China. Gastroenterology 2019;156:2230-2241 e11.

Figure Legends

Figure 1: Box and whisker plots showing median, inter-quartile, outliers and Spearman's rho values for (a) RECAM and (b) RUCAM scores by expert opinion diagnostic categories. 98 DILIN and 96 Spanish Registry cases combined (n = 194). Horizontal lines represent diagnostic score cut-offs for RECAM and RUCAM. Downward pointing arrowheads indicate that the cut-off integer value is included in the category below the line. DILIN categories of definite and highly likely were combined and considered equivalent to Spanish Registry high probable category (labeled High Probable/Highly Likely). Spanish Registry unlikely and excluded categories were combined and considered equivalent to DILIN unlikely category (labeled Unlikely/Excluded).

Supplemental Figure 1:

Figure 1: Box and whisker plots showing median, inter-quartile, outliers and Spearman's rho values for (a) RECAM and (b) RUCAM scores by expert opinion diagnostic categories for 98 DILIN cases. Similar box and whisker plots for (c) RECAM and (d) RUCAM scores by expert opinion diagnostic categories for 96 Spanish Registry cases. Horizontal lines represent diagnostic score cut-offs for RECAM and RUCAM. Downward pointing arrowheads indicate that the cut-off integer value is included in the category below the line. DILIN categories of definite and highly likely were combined, and Spanish Registry unlikely and excluded categories were combined.

Table 1a: RECAM algorithm (Domains 1-3)

Domain 1a & 1b:	Points
Score *both* sections 1a (Onset after drug start) and 1b (Onset after drug stop)	
1a: Onset after drug start (points given)	
Days after drug start where *day 1 is first day drug taken*	
≤ 1 day	-6
2 through 9 days (inclusive)	3
10 through 60 days (inclusive)	4
61 through 90 days (inclusive)	2
>90 days	0
1b: Onset after drug stop (points taken) **[For long 1/2 life agents*, enter zero points for Domain 1b]**	
Days after drug stop where *day 1 is the first day the drug is not taken*	
≤ 30 days	0
31 through 60 days (inclusive)	-1
61 through 90 days (inclusive)	-2
91 through 120 days (inclusive)	-4
>120 days	-6
Domain 2: Dechallange or Washout	Points
Initial R value > 5: apply washout criteria below to serum ALT	
Initial R value ≤ 5: apply washout criteria below to either AP or Bilirubin, whichever gives a higher score	
ALT, AP or Bilirubin (whichever used by R-value criteria above) declines to less than 50% of peak	
If drug still taken when greater than 50% of peak decline occurs	-6
Days from peak value to less than 50% of peak (assumes drug was discontinued)	
1 through 30 days	4
31 through 90 days	3
91 through 182	2
183 through 365	1
> 365	0
All other instances where ALT, AP or Bilirubin does not decline, has not yet declined to less than 50% of peak	0
ALT, AP or Bilirubin (whichever used by R-value criteria above) is > 90% of peak value at anytime >182 days and prior to any transplant without other explanation recurrent or persistent elevation.	-6
Domain 3: Literature supporting liver injury	Points
LiverTox Category (reference: https://livertox.nlm.nih.gov/index.html)	
A, B	3
C or D or E*	1
E or X	0

Before using RECAM, the user should rule out non-liver related sources for enzyme elevations (e.g., muscle, hemolysis and bone) and acetaminophen liver injury, for which this tool is not designed.

-6: Data entered suggests a DILI is not explanatory of liver injury. User should consider this case as excluded or unlikely DILI with a total score of -6. If user chooses to proceed, 6 points will be deducted from the running score, and user should recognize that DILI as the cause of liver injury is questionable due to inconsistent latency or dechallenge, regardless of total score obtained.

*Agents with estimated half-life or pharmacodynamic effect greater than or equal to 15 days.

LiverTox® categories of DILI risk: A: Well-known, well described and characteristic signature. More than 50 well reported cases in the literature; B: Known or highly likely to cause DILI with characteristic

signature. 12-49 cases in the literature; C: Probably causes DILI. No characteristic signature. Less than 12 cases in the literature; D: Possible cause of DILI. Less than 3 cases in the literature. E: Unlikely to causes DILI due to extensive use. Cases in the literature may exist but are unconvincing. E*: Unproven but suspected to cause DILI. Suggestion of liver injury exists outside of published literature (e.g. trial data reported to regulatory agencies) X: Unknown. Agents recently approved or rarely used. For complete information go to LiverTox® online.[17]

Table 1b: RECAM (Domain 4)

Domain 4: Exclusion of competing diagnoses*	Points
Hepatitis A	
Missing HAV IgM anti-HAV data	-3
IgM anti-HAV negative (if total anti-HAV is negative, consider IgM negative as well)	0
IgM anti-HAV positive	-6 *
Hepatitis B	
Missing IgM anti-HBc [note: (-) anti-HBc *total* means IgM is negative, but (+) anti-HBc total does *not inform IgM result*]	-3
HBsAg *and* IgM anti-HBc negative (if total anti-HBc is negative, consider IgM negative; anti-HBc IgG may be + or -)	0
HBsAg positive and IgM anti-HBc negative (if total anti-HBc is negative, consider IgM negative; anti-HBc IgG may be + or -)	-1
IgM anti-HBc positive regardless of HBsAg result or missing	-6 *
Hepatitis C	
Missing anti-HCV *or* HCV RNA	-3
Anti-HCV *and* HCV RNA both negative	0
Anti-HCV *and/or* HCV RNA (+) then score according to initial R-value:	
R ≤ 5 HCV RNA (-) & anti-HCV (+)	**0**
R ≤ 5 HCV RNA (+) & anti-HCV (+) or HCV RNA (+) & anti-HCV (-)	**-1**
R > 5 with known chronic infection	-1
R > 5, no known chronic infection and no exposure risk in ≤ 100 days prior to onset	-1
R > 5, no known chronic infection and exposure risk in ≤ 100 days prior to onset	-6 *
HEV (IgM serologies)	
Missing IgM anti-HEV data	-3
IgM anti-HEV negative	0
IgM anti-HEV positive	-6 *
Alcohol (AST and ALT values at onset)	
AST:ALT ≥ 2 with AST ≤ 500 *and* missing alcohol history	-3
AST:ALT < 2 and/or AST >500	0
AST:ALT ≥ 2 with AST≤ 500 then score according to alcohol history below:	
Average of ≤ 2 standard drinks/d for women, ≤ 3 standard drinks/d for men within 6 weeks of injury onset	0
Average of >2 and ≤4 standard drinks/d for women, > 3 and ≤ 6 standard drinks/d for men within 6 weeks of injury onset	-3
Average of >4 standard drinks/d for women, >6 standard drinks/d for men within 6 weeks of injury onset	-6 *
Biliary or parenchymal disease assesed by imaging (US, CT, MRI, MRCP or cholangiogram)	
Missing imaging data	-3
Imaging shows no biliary stenosis(es) or obstruction, no or <50% malignant infiltration	0
Imaging shows biliary stenosis(es) or obstruction or infiltrating malignancy occupying ≥ 50% of the liver.	-6 *
Autoimmune Hepatitis: Use either (a) or (b) below	
(a) Autoimmune Hepatitis assessment for non-minocycline and non-nitrofurantion cases	
Missing ANA and ASMA and IgG	-3
ANA <1:80, ASMA <1:80, IgG < 1.1 ULN. Can be missing 1-2 of these, but those obtained must be below these levels.	0
ANA ≥ 1:80 *or* ASMA ≥ 1:80 *or* IgG ≥ 1.1 ULN	-1
(ANA ≥ 1:80 *or* ASMA ≥ 1:80) ***and*** IgG ≥ 1.1 ULN, and liver biopsy with typical features of AIH	-6 *
(b) Autoimmune Hepatitis assessment for minocycline and nitrofurantion cases	
Missing ANA and ASMA and IgG	-3
ANA <1:80, ASMA <1:80, IgG < 1.1 ULN. Can be missing 1-2 of these, but those obtained must be below these levels.	0
ANA ≥ 1:80 *or* ASMA ≥ 1:80 *or* IgG ≥ 1.1 ULN	1
Liver injury due to ischemic liver injury (shock liver) and/or acute congestive hepatopathy^	
No information on possible hypoxia, hypotension, shock or acute congestive hepatopathy (history incomplete or inadequate)	-1
No known or suspected prolonged hypoxia, hypotension, shock or acute congestive hepatopathy within 1 wk prior	0
Known or suspected episodes of prolonged hypoxia, hypotension, shock or acute congestive hepatopathy within 1 wk prior	-2
Sepsis causing cholestasis	
No information on sepsis or systemic inflammatory resonse (SIRS), and R-value <5	-1
R-value ≤ 5 but no sepsis or systemic inflammatory resonse (SIRS), or R-value >5	0
Sepsis or systemic inflammatory resonse (SIRS) present *and* R-value < 5	-2

When critical data are missing in Domain 4, -3 points are assessed, but user should consider obtaining these data before proceeding. -6*: Data entered suggests a non-DILI explanation for liver injury. User should consider the case as excluded DILI with a total score of -6. If user chooses to continue, 6 points will be deducted, and user should

recognize that DILI as sole cause of liver injury is questionable, regardless of total score obtained. ^Consider ischemia or shock when transaminases are extremely high (e.g., >7,500 U/L) with elevated LDH and AST>ALT.

Table 1c: RECAM (Domain 5)

Domain 5: Additional data	Points
The following information may be available in the evaluation, but are not required.	
Retrospective Rechallenge: h/o DILI w/ jaundice to same drug	
No history of prior exposure or no DILI with jaundice after exposure to this drug or agent in the past	0
Positive history of DILI with jaundice after exposure to drug or agent; no documentation by lab results necessary	1
Prospective Rechallenge (documented with labs)	
No rechallenge or no data regarding rechallenge	0
Re-expsoure results in rise in liver enzymes 2-3 x ULN (or baseline)	0
Re-exposure: same R-value category, latency <60 da., ALT, AST > 3x ULN(or baseline) or AP >2x ULN(or baseline)	6
Re-expsoure to same drug results in no injury with <2 x ULN (or baseline) rise in liver enzymes	-3
Liver biopsy	
No biopsy done	0
Non-diagnositic (can be suggestive of DILI, but not diagnostic)	0
Biopsy carries features consistent with a specific DILI	1
Diagnostic of non-DILI diagnosis (e.g. infiltrating cancer, ischemic injury, alcoholic hepatitis)	-6 *
CMV (IgM =IgM anti-CMV)	
Missing both IgM and PCR	0
Negative (both IgM and PCR negative or at least one negative and other not done)	0
Positive IgM or PCR	-2
Positive IgM and PCR	-6
EBV (IgM can be any IgM anti-EBV antibody, heterophile test, monosopot or EBV early antigen)	
Missing IgM and PCR	0
Negative (both IgM and PCR negative or at least one negative and other not done)	0
Positive IgM or PCR	-2
Positive IgM and PCR	-6
HSV (IgM = IgM anti-HSV)	
Missing IgM and PCR	0
Negative (both IgM and PCR negative or at least one negative and other not done)	0
Positive IgM or PCR	-2
Positive IgM and PCR	-6
Drug reaction with eosinophila and systemic symptoms (DRESS) or Steven Johnsons Syndrome (SJS)	
Absent or no information	0
Present	1

-6*: Data entered suggests a non-DILI explanation for liver injury. User should consider the case as excluded DILI with a total score of -6. If user chooses to continue, 6 points will be deducted from the running score, and user should recognize that DILI as sole cause of liver injury is questionable due to a competing explanation, regardless of total sum score obtained.

Table 2. Critical clinical elements for the diagnosis of DILI

Element	Comments
Minimum liver test elevations[14]	
ALT ≥5x ULN* **ALP ≥2x ULN** **ALT > 3x ULN + total Bilirubin > 2x ULN**	ULN may be replaced by the mean baseline values obtained prior to exposure to drug if baseline values are abnormal.
Temporal sequence for latency & dechallenge (RECAM Domains 1 & 2)	Consider temporal relationship between drug exposure, injury onset and improvement.
Competing Medications	Obtain thorough pharmacologic history of other drugs that have appropriate temporal relationship between drug exposure, injury onset and improvement. Consider obtaining a separate RECAM score for these drugs.
Alternative diagnoses (RECAM Domains 4)	
Viral hepatitis A, B, C, and E	For chronic hepatitis B or C try to establish a baseline and course for liver enzymes, bilirubin and viral load to help exclude disease exacerbation.
Alcoholic hepatitis	Obtained detailed alcohol intake history
Biliary obstruction	Imaging studies needed
Autoimmune hepatitis	Testing for ANA, ASMA, total IgG
Hypotension due to shock and/or heart failure	Clinical diagnosis
Cholestasis of sepsis	Clinical diagnosis
Malignant infiltration of the liver	Imaging studies needed. Biopsy may be needed.

*ULN = upper limit of normal

Table 3: Clinical characteristics of 98 DILIN and 96 Spanish DILI Registry cases

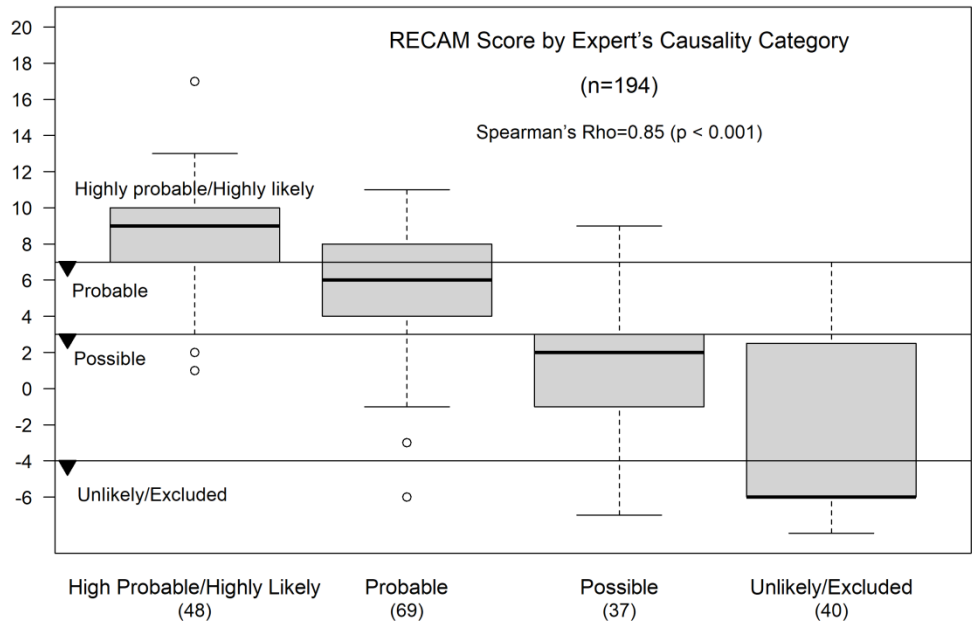
Patient Characteristics from DILIN and Spanish DILI Registries		
Characteristic	DILIN N= 98	Spanish Registry N=96
Age in years, mean (SD)	48 (18.4)	58 (17.3)
Women	56 (57%)	48 (50%)
Race		
Caucasian	80 (82%)	95 (99%)
Black	9 (9%)	0 (0%)
Asian	4 (4%)	0 (0%)
Other	5 (5%)	1 (1%)
Injury Pattern*		
Cholestatic	22 (23%)	17 (18%)
Mixed	22 (23%)	21 (22%)
Hepatocellular	51 (54%)	58 (60%)
Likelihood category:		
Definite/Highly likely or High probable	38 (39%)	10 (10%)
Probable	20 (20%)	49 (51%)
Possible	20 (20%)	17 (18%)
Unlikely or Excluded	20 (20%)	20 (21%)

*Based on R-value (ALT/ULN ÷ ALP/ULN). R-value > 5 hepatocellular, 2< R-value <5 mixed, R-value < 2 cholestatic.[1]

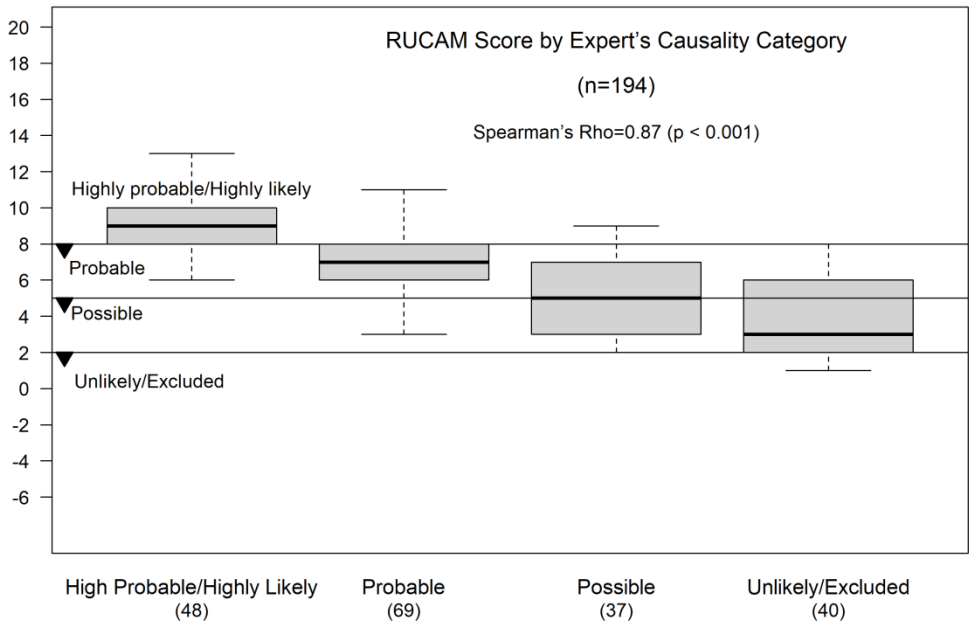
Table 4: Diagnostic performance of RECAM and RUCAM compared to expert opinion for DILIN and Spanish Registry cases combined (n = 194)

Performance category	RECAM	RUCAM	p-value
Area under the receiver operator curve (95% CI)			
At least Highly likely or Highly probable	0.87 (0.81, 0.92)	0.85 (0.80, 0.91)	0.73
At least Probable	0.89 (0.84, 0.93)	0.89 (0.84, 0.93)	0.92
At least Possible	0.88 (0.81, 0.94)	0.87 (0.81, 0.93)	0.90
Overall Agreement (95% CI)			
Percent agreement	62.4 (55.6 - 69.2)	58.8 (51.8 - 65.7)	0.44
Weighted Kappa	0.62 (0.53, 0.70)	0.56 (0.48, 0.65)	0.16
Sensitivity (95% CI)			
Highly probable, Definite or Highly likely	72.9 (60.4 - 85.5)	54.2 (40.1 - 68.3)	0.02
Probable	49.3 (37.5 - 61.1)	68.1 (57.1 - 79.1)	0.03
Possible	70.3 (55.5 - 85.0)	59.5 (43.6 - 75.3)	0.20
Unlikely or Excluded	65.0 (50.2 - 79.8)	47.5 (32.0 - 63.0)	0.08
Specificity (95% CI)			
Definite, Highly likely, or Highly probable	86.3 (80.7, 91.9)	89.0 (84.0, 94.1)	0.41
Probable	82.4 (75.7, 89.1)	63.2 (54.8, 71.7)	< 0.01
Possible	82.8 (76.9, 88.7)	89.2 (84.3, 94.0)	0.08
Unlikely or Excluded	97.4 (94.9, 99.9)	99.4 (98.1, 1.00)	0.18

CI = confidence interval



hep_32327_f1a.tif



hep_32327_f1b.tif