Causality Assessment in Drug-Induced Liver Injury Using a Structured Expert Opinion Process: Comparison to the Roussel-Uclaf Causality Assessment Method

Don C. Rockey[1], **Leonard B. Seeff**[2], **James Rochon**[3], **James Freston**[4], **Naga Chalasani**[5], **Maurizio Bonacini**[6], **Robert J. Fontana**[7], and **Paul H. Hayashi**[8] **for the US Drug-Induced Liver Injury Network**

[1]Division of Digestive and Liver Diseases, University of Texas Southwestern Medical Center, Dallas, TX

[2]National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD

[3]Duke Clinical Research Institute, Durham, NC

[4]Department of Medicine, University of Connecticut Health Center, Farmington, CT

[5]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN

[6]Department of Transplantation, California Pacific Medical Center, San Francisco, CA

[7]Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, MI

[8]Liver Program, Division of Gastroenterology and Hepatology, University of North Carolina School of Medicine, Chapel Hill, NC

Abstract

Drug-induced liver injury (DILI) is largely a diagnosis of exclusion and is therefore challenging. The US Drug-Induced Liver Injury Network (DILIN) prospective study used two methods to assess DILI causality: a structured expert opinion process and the Roussel-Uclaf Causality Assessment Method (RUCAM). Causality assessment focused on detailed clinical and laboratory data from patients with suspected DILI. The adjudication process used standardized numerical and descriptive definitions and scored cases as definite, highly likely, probable, possible, or unlikely. Results of the structured expert opinion procedure were compared with those derived by the RUCAM approach. Among 250 patients with suspected DILI, the expert opinion adjudication process scored 78 patients (31%) as definite, 102 (41%) as highly likely, 37 (15%) as probable, 25 (10%) as possible, and 8 (3%) as unlikely. Among 187 enrollees who had received a single implicated drug, initial complete agreement was reached for 50 (27%) with the expert opinion process and for 34 (19%) with a five-category RUCAM scale (*P* = 0.08), and the two methods demonstrated a modest correlation with each other (Spearman's *r* = 0.42, *P* = 0.0001). Importantly, the RUCAM approach substantially shifted the causality likelihood toward lower probabilities in comparison with the DILIN expert opinion process.

Address reprint requests to: Don C. Rockey, M.D., Division of Digestive and Liver Diseases, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-8887. don.rockey@utsouthwestern.edu; fax: 214-648-8446.

Conclusion—The structured DILIN expert opinion process produced higher agreement rates and likelihood scores than RUCAM in assessing causality, but there was still considerable interobserver variability in both. Accordingly, a more objective, reliable, and reproducible means of assessing DILI causality is still needed.

A diagnosis of hepatotoxicity must be considered when liver injury is identified in a person taking a prescription drug, herbal, or over-the-counter product, even if there is already preexisting liver disease.[1-5] Because there is currently no specific marker of drug-induced liver injury (DILI), the diagnosis rests on excluding other conditions that can mimic such injury. The diagnosis is especially difficult when affected persons are taking multiple products, any one of which might be responsible, and because of possible synergism between drugs.[1,6-8]

In the traditional diagnostic approach to suspected DILI, which involves clinical, biochemical, and histological evaluation, attempts are made to establish the latency between the start of the drug and the onset of injury, its clinical signature, the exclusion of alternate etiologies, evidence of improvement of the liver injury upon drug withdrawal (dechallenge), and the effect of deliberate or inadvertent rechallenge. When performed by an experienced clinician, the assessment is considered by expert opinion. However, even for experts, the diagnosis of DILI can be problematic because of the inherently subjective nature of this approach. Efforts have therefore turned toward developing more objective diagnostic strategies through the creation of specific instruments such as the Roussel-Uclaf Causality Assessment Method (RUCAM), the Maria and Victorino method, and the Naranjo scale, the last designed to assess all forms of adverse drug reactions.[9-13] In a head-to-head comparison of these instruments, RUCAM has been found to perform best for diagnosing hepatotoxicity, but it is cumbersome and therefore is rarely used in clinical practice.

The Drug-Induced Liver Injury Network (DILIN) is a multicenter study whose primary aims are to identify and collect information on bona fide cases of drug-induced liver disease and to obtain serum, DNA, and liver tissue to allow for mechanistic investigation. When the study was being planned, the decision was made to assess causality with both expert opinion and RUCAM. A highly structured expert opinion method was developed that was specifically designed to include standardized terminology and specific methodology, and it is hereafter called structured expert opinion. It was hypothesized that this approach may have certain advantages in comparison with RUCAM. This report describes how the expert opinion approach was developed and refined and compares its effectiveness to that of RUCAM.[14]

Patients and Methods

Established in 2003, DILIN originally comprised five clinical sites (and their affiliates), a data coordinating center (DCC), a serum, DNA, and tissue repository, and a central histopathology core.[15] The study has since expanded to include eight clinical sites. Network investigators were charged with identifying and enrolling persons who developed DILI, carefully phenotyping the clinical condition, and collecting appropriate biological samples. Details concerning the planning, initial study design, study outcomes, eligibility criteria, and conduct of the study have been reported,[14] and the clinical features of the first 300 patients enrolled have been summarized.[16] The study protocols were approved by the institutional review boards at all participating institutions and were registered at ClinicalTrials.gov.

Eligibility Requirements for the Prospective Study

In brief, enrollees in the prospective study were persons receiving single or multiple drugs, herbals, or other over-the-counter products identified to have biochemically defined liver

dysfunction, provided that they could be evaluated within 6 months of onset of the liver disease.[14] Biochemical criteria for enrollment included (1) two consecutive serum alanine aminotransferase (ALT) or aspartate aminotransferase (AST) values > 5 times the upper limit of normal (ULN) or > 5 times the baseline abnormal value, (2) two consecutive serum alkaline phosphatase (AP) values greater than twice the ULN or twice the baseline abnormal value, or (3) an otherwise unexplained total serum bilirubin value > 2.5 mg/dL or an international normalized ratio (INR) > 1.5 on two consecutive occasions. Symptoms or signs of liver injury were not required. Exclusion criteria were liver injury due to acetaminophen, preexisting autoimmune hepatitis or sclerosing cholangitis, and previous receipt of a bone marrow or liver transplant. Persons were not excluded for preexisting chronic hepatitis B or C or human immunodeficiency virus infection, provided that baseline laboratory test results were available.

Structured DILIN Expert Opinion Process

Sequential steps in causality assessment are outlined in Fig. 1. Complete clinical data, including serial laboratory test results together with the local ULN values, were extracted from the clinical records and entered into a 65-page case report form (CRF). To exclude conditions that can mimic drug-induced liver disease, the patients and their medical records were screened for previous liver disease, alcohol use, serological and virological evidence of hepatitis A, B, or C infection, autoantibodies, ceruloplasmin, alpha-1-antitrypsin, ferritin, and iron; additionally, results of imaging studies were reviewed. Patients who had not been fully evaluated when they were first identified underwent testing for any missing laboratory data at enrollment. Liver biopsy was not required for adjudication purposes, but if it was performed as part of routine clinical care, the results were collected and made available to reviewers. To facilitate adjudication, the extensive database was summarized in an abbreviated CRF that included such key elements as the date of onset of liver injury, complete information about all medications taken within 6 months of onset of the event, the presence of symptoms and signs of liver disease, pertinent past medical history, complete laboratory tests, imaging and liver biopsy results, and serial results for ALT, AST, AP, serum bilirubin, and the prothrombin time or INR.

In addition to the short CRF summary, a succinct case summary called the clinical narrative was completed by the study investigator who enrolled the subject. The narrative provided detailed information on the history and chronology of the illness with dates of drug initiation and liver disease onset, pertinent features of the liver disease, and the time to improvement or recovery. The narrative also included information on past use of the implicated agent and significant concomitant drugs, the past medical history, the extent of alcohol use, whether there had been an episode of hypotension, and information on the course of the illness, including hospitalization, a history of hepatic decompensation or organ failure, and death or liver transplantation. Finally, the investigator provided a rationale for ascribing the event to a specific medication or medications without offering a personal view on the estimated strength of the association.

Origin and Status of the Reviewers

The CRF summary and clinical narrative were first assessed by the DCC for consistency and omissions and, after approval, were forwarded to three reviewers, including the submitting investigator and two members of the DILIN causality committee from other sites. The three reviewers each worked independently, without knowledge of who the other two were or what scores they awarded. The two nonsubmitting reviewers were selected in rotation from the full causality committee, which consisted of principal investigators and coprincipal investigators from the five clinical sites and the DCC and project officers and scientific advisors from the National Institute of Diabetes and Digestive and Kidney Diseases (see

Appendix 1 in the supporting information). All the reviewers were hepatologists with experience in evaluating DILI. All contributed to the design of the study and, from the outset, participated in an in-depth discussion of the issues related to hepatotoxicity and in fashioning the DILIN causality process through frequent conference calls, e-mail communications, and face-to-face meetings. This allowed for the thorough evaluation of the scoring systems and ended in the development of standard operating procedures for both the DILIN system and RUCAM. The RUCAM standard operating procedure was generated after one of its originators was contacted for clarification purposes and with a broad examination of relevant literature. Thereafter, experience was gained by frequent discussion of representative examples of DILI and by re-review of specific cases.

To avoid a conflict of interest, reviewers who had a declared relationship with the manufacturer of any implicated drug were excluded. Detailed conflict-of-interest forms were completed annually by all DILIN participants.

DILIN Adjudication Process

After independently evaluating the cases, each reviewer, using a five-point or category scale, provided an assessment of the likelihood that the medication caused the liver injury. They also completed a RUCAM form (see Appendix 2 in the supporting information) and used the instructions provided with the form. Information necessary to complete the RUCAM assessment was included in the short CRF and the clinical narrative.

The five-point (category) DILIN likelihood causality scale used both a percentage figure and descriptive legal terminology to grade cases as definite, highly likely, probable, possible, or unlikely (Table 1). Causality was considered to be definite if attribution of the drug to the liver injury was believed to exceed 95% likelihood with an association beyond a reasonable doubt. Cases were awarded this grade if the medication was well recognized to cause liver injury, it had a characteristic or typical signature, and there was no evidence of a competing diagnosis. The designation *highly likely* was applied when there was an estimated 75% to 95% likelihood of an association and by the legal phrase indicating clear and convincing evidence for the association. These cases were regarded as convincingly due to the medication, with minor reservations because of a somewhat atypical course or presentation or the remote possibility of another diagnosis. Cases were called probable when the likelihood of an association was considered to be between 50% and 75%, with legal terminology indicating that the association was supported by the predominance of the evidence. Although appearing to show an association, such cases would not be graded higher because of an atypical course, the absence of essential clinical information, or the presence of another possible explanation or diagnosis. Cases were considered to be possible if they were believed to have a 25% to 50% likelihood of an association because, although it was still possibly related, the involvement by the drug was equivocal and was not supported by the preponderance of the evidence. Cases were ranked as unlikely if they were regarded to have less than a 25% likelihood of resulting from the medication, and another etiology was considered to be responsible. These definitions attached semiquantitative values to these inherently subjective terms and brought increased uniformity to the adjudication process. For a more complete summary of the definitions of each category, please see Supporting Table 1.

If more than one drug, herbal, or nutritional supplement was considered potentially responsible, a separate assessment by expert opinion and RUCAM was completed for each drug. The case was assessed first for the overall likelihood that a drug caused liver disease with the five-category DILIN scale, and then each drug (up to three were allowed) was assessed separately. Combination drugs, such as amoxicillin/clavulanate, trimethoprim/sulfamethoxazole, and the majority of herbal preparations, were assessed as if they were a

single agent without an attempt to appraise the role of each component separately. Implicated drugs graded for likelihood by the three reviewers were assessed also for the severity of the liver injury by a single reviewer, and the results were submitted to the DCC for addition to the database.

A causality conference call was arranged monthly to review cases adjudicated for that month by the three reviewers using the structured expert opinion method and RUCAM. If all three had independently reached the same causality scores before the call, this was accepted as a final result and not discussed further. If, however, there was discrepancy among the three reviewers, the chair of the causality committee attempted to reconcile the differences among them before the conference call through open and transparent dialogue. If accord was still not reached at the time of the conference call, the three reviewers were given one last opportunity on the call to reach agreement. Failing to find consensus, the full causality committee then voted on the case, and the majority result was accepted as the final score.

Liver biopsy was performed inconsistently and often at different stages in the course of the liver injury. For these reasons, liver biopsy was not used as a formal feature of the adjudication process. However, local biopsy readings were available to reviewers.

Statistics

Standard descriptive statistics were used to summarize the features of enrolled patients, which included demographic characteristics, signs and symptoms, laboratory data, and type of injury. To assess discrepancies, pairwise differences among the three primary reviewers were first compiled, and the maximum of the absolute values of these differences [the maximum absolute difference (MAD) among them] was recorded. Spearman's correlation was used to assess the association between the RUCAM and DILIN structured expert opinion scores. Between-group comparisons of the DILIN causality score were made with Fisher's exact test. McNemar's test[17] was used to compare the rate of complete agreement among the three reviewers in the two causality approaches.

Results

Characteristics of the Study Cohort

The analysis focused on the first 250 adjudicated cases, 187 (75%) of whom had received a single drug or herbal product. Their demographic, clinical, and biochemical features (Table 2) closely resembled those of the 300 patients in the prospective study previously described.[16] The average age of the patients was 49 years, and 58% were women. Approximately two-thirds were jaundiced (bilirubin > 2.5 mg/dL), 58% were hospitalized, and 5% died within 6 months of onset of liver injury or required liver transplantation. Although almost 20% of the subjects reported during the initial interview that they had had prior liver disease, objective evidence for this was present in only 6%, the most common cause being underlying hepatitis C virus infection.

The liver injury was attributed to a wide variety of drugs and herbal products, which included antimicrobials (46%), central nervous system agents (15%), immunomodulatory agents (7%), herbals (5%), anti-neoplastic agents (4%), lipid-lowering agents (4%), analgesics (3%), and others (16%). The most frequent presenting pattern of injury was that of hepatocellular liver disease (i.e., R value ≥ 5).[18]

Causality Assessment Using the DILIN Structured Expert Opinion Adjudication Process

Causality assessment by the structured expert opinion method was conducted in two phases, the first consisting of the frequency with which the three independent reviewers reached

initial common agreement and the second consisting of the frequency with which they were willing to alter their initial causality grade after group discussion. The frequency of initial agreement among the three reviewers was relatively high, as indicated by the MAD in causality assessment scores for the 250 assessed cases (Table 3). All three agreed in the assessment in 27% of the cases (MAD of 0), and there was agreement by two of the three in another 43% of patients, the third reviewer differing by only one category or point (MAD of 1).

Results of the final assessment using the DILIN structured expert opinion approach and its comparison with the initial assessment are shown in Table 4. The two most frequent scores assigned initially by the three reviewers were definite and highly likely, and these evaluations changed little at the final assessment. Thus, the final conclusion was that 31% of cases were considered definite, 41% were highly likely, 15% were probable, 10% were possible, and 3% were unlikely. In general, when the full causality committee voted on adjudication, they tended to adopt the majority opinion reached among the three reviewers, unless one reviewer established compelling evidence to the contrary.

Comparison of the DILIN Expert Opinion Adjudication Process and the RUCAM Method

All cases were assessed by each reviewer separately with both the DILIN structured expert opinion approach and RUCAM; the results of the two are compared for the 187 patients who had received a single drug (Table 5). Because each case had been evaluated by three reviewers, the total number of reviews should have totaled 561; all 561 reviews were completed with the expert opinion approach, but 4 were missing for RUCAM, so completion of 557 scores (99.3%) was permitted. RUCAM assigns scores that range from +15 to −3, with *highly probable* requiring a score of >8, *probable* requiring a score of 6 to 8, *possible* requiring a score of 4 to 6, and *unlikely* requiring a score of 1 to 3; DILI is excluded for a score of <1.

Reviewers, using structured expert opinion, scored 409 cases (196 + 213) as definite or highly likely (total of 72%), but only 132 (24%) were assigned the equivalent RUCAM score of highly probable (Table 5). Furthermore, although reviewers scored 22 cases (4%) as unlikely with the DILIN structured expert opinion process, 38 of the cases (8%) were assessed correspondingly by RUCAM as either unlikely (22) or excluded (16). Finally, although the structured expert opinion method scored 23.1% of the cases as either probable or possible, RUCAM attributed 69.5% of the cases to the equivalent probable or possible categories. These comparative results are displayed in a box and whisker plot (Fig. 2). There was considerable variability in the comparison of the RUCAM score to each level of the DILIN structured expert opinion scores, with RUCAM displaying lower levels of causality (Spearman's correlation, $r = 0.42$ in absolute value; $P = 0.0001$).

A comparison of the agreement among the reviewers in causality assessment between the structured expert opinion and RUCAM methods, restricted to the 187 patients who had received only a single drug, is shown in Table 6. Complete agreement (MAD = 0) was reached in 27% with expert opinion versus 19% with RUCAM ($P = 0.08$). The average MAD was 1.12 with the DILIN strategy and 1.18 with the RUCAM strategy.

In order to adequately assess the relationship between the conclusions of the DILIN process and RUCAM, it should be possible to directly compare the results of the two different assessment methods. Such a comparison is, however, compromised by the fact that, even though both systems use five levels of likelihood, the terminological differences hinder a direct comparison. In an effort to circumvent this problem, two different types of comparisons were undertaken.

The first consisted of directly comparing the results of the two approaches in a 5 × 5 table with the established terms for each of them, even though an individual term, such as *possible*, might not have the identical weight. Nevertheless, the comparison is based on the relative ranking on the two ordinal scales. As shown in cross-tabulation (diagonal box) in Table 7, there was agreement in the relative ranking in 230 of the 557 reviews (41.3%). Moreover, scores fell within one category of each other in 479 reviews (86.0%). The majority of cases scored at DILIN's highest causality category (definite) were scored at lower levels by RUCAM. Similarly, disagreements at DILIN's second causality level (very likely) were scored more often at lower causality levels by RUCAM. In contrast, disagreements at DILIN's third and fourth causality levels (probable and possible) were scored more often at higher causality levels by RUCAM. Thus, RUCAM graded more cases in the middle ranges, whereas the DILIN process scored a greater number of cases in higher and lower likelihood categories.

A second analysis took into account the fact that a score of probable or higher in both systems would probably signify a valid case of DILI. Thus, the comparison was collapsed into a 2 × 2 table, and the outcomes for both were separated into "yes = DILI" and "no = not DILI" (Table 8). Even at this most basic level, there was agreement in only 384 of the reviews (68.9%), as displayed in the cross-tabulation. In this analysis, the DILIN expert opinion process was more likely than RUCAM to ascribe the case to DILI [DILIN, 495/557 (88.9%) versus RUCAM, 352/557 (63.2%)].

Discussion

Because there is no specific objective marker of drug-induced liver disease, an accurate diagnosis of hepatotoxicity has constantly been challenging. DILI must always be considered, however, when there is a temporal association between observed liver injury and the receipt of a drug. The warning signal has been either an acute onset of clinical symptoms (e.g., rash, fever, abdominal pain, or jaundice) or, more commonly, biochemical dysfunction, which includes raised levels of ALT, AST, AP, gamma-glutamyl transpeptidase, and/or serum bilirubin.[1,19,20] Although these abnormalities strongly suggest liver disease, they are in fact nonspecific indicators and, moreover, do not provide an etiological diagnosis.

The use of expert opinion to identify DILI has long been regarded as the gold standard for diagnosing hepatotoxicity, especially when reports come from well-recognized authorities rather than from an inexperienced occasional observer.[2-5,21,22] However, because this approach is subjective and lacks defined criteria, a group of international experts convened a meeting under the auspices of the Council for International Organizations of Medical Scientists with the goal of introducing structure and uniformity to the causality process through the development of highly defined diagnostic criteria for drug-induced liver disease. The meeting was supported by Roussel-Uclaf Pharmaceuticals, and hence the instrument is called RUCAM. The strategy awards points for seven different domains.[10] Other attempts have been made to develop causality instruments with the hope of simplifying the adjudication process,[11,23] but the value of some has been questioned[24]; this has left RUCAM as the preferred causality instrument. Although used by some experts in the field and often referred to in discussing DILI causality, RUCAM has not been adopted in general clinical practice or, in fact, by most practicing hepatologists and gastroenterologists. The chief reason is that it is a time-consuming process with insufficient and sometimes confusing information on how to score some of the elements of its domains. Nevertheless, during the planning for DILIN, the decision was made to use and compare two approaches for establishing causality: a refined and highly structured expert opinion method and the RUCAM instrument.

A direct comparison of the DILIN structured expert opinion and RUCAM revealed that the DILIN process was more likely than RUCAM to generate a score supportive of drug-induced liver disease. Using the DILIN system, reviewers scored 73% of cases (405/557) as definite or highly likely, whereas only 24% of the cases (132/557) were scored as highly probable in the corresponding RUCAM category (Table 6). Indeed, the DILIN process grouped subjects more toward the definite/highly likely category, whereas the same cases evaluated by RUCAM were scored in the middle ranges (probable/possible), and this suggested that RUCAM skewed the distribution toward a lesser overall likelihood. The RUCAM approach was thus more conservative in assigning a high level of causality than the DILIN strategy. A drawback to this comparison, however, is that the two grading categories are not strictly parallel, and collapsing of categories was required to bring them to a reasonable accord. Furthermore, such grouping of categories was not part of the actual design of either causality method.

Also of note is the fact that the DILIN approach afforded substantially greater agreement in the initial blinded evaluation than the RUCAM approach. With the DILIN system, all three reviewers agreed completely in 50 of the cases (27%), and they disagreed by only one point in an additional 83 (44%); they thus achieved generally similar conclusions in 70% of the adjudicated cases. In contrast, when RUCAM, restricted to persons who had received only a single agent, was used, complete agreement was even lower at 19% of subjects (34/187). This is somewhat surprising because RUCAM was designed to be an objective causality score. The variability is likely due to the ambiguities of some of the RUCAM score parameters. Nevertheless, even though there was greater reviewer agreement with the DILIN structured expert opinion method than with the RUCAM approach, there was still disagreement in almost one-third of cases with the DILIN adjudication method. This is not unexpected, however, because the structured expert opinion process persists in being a subjective form of assessment until a definitive diagnostic marker is established, and thus assessments will continue to vary according to individual reviewer perspectives. Indeed, difficulties in reaching consensus among multiple reviewers working independently have been described previously,[25] although disagreements appear less likely when reviewers are experts trained in the use of a standardized causality assessment method.[26,27]

The RUCAM scoring system appears to be problematic even for experienced persons, let alone for nonexpert health professionals in clinical practice. Indeed, in a previous report from the DILIN study group, RUCAM was found to have poor reproducibility, even when repeated by the same reviewers.[28] However, as already noted, the refined expert opinion process developed for this study also has its limitations. One of these is that there was unquestionably selection bias in recruiting subjects into this study because the site investigators, all experts, tended to choose cases with a high probability of a diagnosis of DILI, especially those with severe injury. Nevertheless, identical cases were reviewed by the two modalities, so any bias would apply to both systems. Another limitation is that the DILIN approach used three and sometimes more expert reviewers, a luxury not available in routine practice, and this limits its general clinical applicability. Most important is that, without a specific diagnostic biomarker, adjudication by expert opinion remains subjective in the hands of even highly experienced clinician investigators.

In conclusion, the goal of this network study was to develop a carefully standardized approach for assessing DILI that would yield high-quality and consistent results because it involved experienced hepatologists. Additionally, it was believed that the use of RUCAM would complement the expert opinion approach. Instead, the correlation between the two adjudication methods was weak. Indeed, neither approach, as currently designed, can be considered fully effective for assessing causality of DILI outside a research setting. There is clearly a need for a more objective, quantitative, and effective method of adjudication for

drug-induced liver disease. Undoubtedly, such an instrument would contain many of the fields currently included in the RUCAM instrument, but it would require modifications and improved, more user-friendly definitions and may ultimately include genomic and/or proteomic assessment. The components of such an instrument would require careful and precise definitions without ambiguity. Moreover, this new instrument would ideally be developed in a web-based, computerized form that could be programmed to rapidly produce a meaningful score. The DILIN study continues to work with this goal in mind.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Zimmerman HJ. Drug-induced liver disease. Clin Liver Dis. 2000; 4:73–96. [PubMed: 11232192]

2. Sgro C, Clinard F, Ouazir K, Chanay H, Allard C, Guilleminet C, et al. Incidence of drug-induced hepatic injuries: a French population-based study. Hepatology. 2002; 36:451–455. [PubMed: 12143055]

3. Galan MV, Potts JA, Silverman AL, Gordon SC. The burden of acute nonfulminant drug-induced hepatitis in a United States tertiary referral center [corrected]. J Clin Gastroenterol. 2005; 39:64–67. [PubMed: 15599214]

4. Vuppalanchi R, Liangpunsakul S, Chalasani N. Etiology of new-onset jaundice: how often is it caused by idiosyncratic drug-induced liver injury in the United States? Am J Gastroenterol. 2007; 102:558–562. [PubMed: 17156142]

5. Carey EJ, Vargas HE, Douglas DD, Balan V, Byrne TJ, Harrison ME, et al. Inpatient admissions for drug-induced liver injury: results from a single center. Dig Dis Sci. 2008; 53:1977–1982. [PubMed: 18392678]

6. Bissell DM, Gores GJ, Laskin DL, Hoofnagle JH. Drug-induced liver injury: mechanisms and test systems. Hepatology. 2001; 33:1009–1013. [PubMed: 11283870]

7. Liu ZX, Kaplowitz N. Immune-mediated drug-induced liver disease. Clin Liver Dis. 2002; 6:755–774. [PubMed: 12362579]

8. Watkins PB, Seeff LB. Drug-induced liver injury: summary of a single topic clinical research conference. Hepatology. 2006; 43:618–631. [PubMed: 16496329]

9. Benichou C. Criteria of drug-induced liver disorders. Report of an international consensus meeting. J Hepatol. 1990; 11:272–276. [PubMed: 2254635]

10. 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. J Clin Epidemiol. 1993; 46:1331–1336. [PubMed: 8229111]

11. Maria VA, Victorino RM. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. Hepatology. 1997; 26:664–669. [PubMed: 9303497]

12. Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther. 1981; 30:239–245. [PubMed: 7249508]

13. Garcia-Cortes M, Lucena MI, Pachkoria K, Borraz Y, Hidalgo R, Andrade RJ. Evaluation of Naranjo adverse drug reactions probability scale in causality assessment of drug-induced liver injury. Aliment Pharmacol Ther. 2008; 27:780–789. [PubMed: 18284654]

14. Fontana RJ, Watkins PB, Bonkovsky HL, Chalasani N, Davern T, Serrano J, et al. Rationale, design and conduct of the Drug-Induced Liver Injury Network (DILIN) prospective study. Drug Saf. 2009; 32:55–68. [PubMed: 19132805]

15. Hoofnagle JH. Drug-induced liver injury network (DILIN). Hepatology. 2004; 40:773. [PubMed: 15382161]

16. Chalasani N, Fontana RJ, Bonkovsky HL, Watkins PB, Davern T, Serrano J, et al. Causes, clinical features, and outcomes from a prospective study of drug-induced liver injury in the United States. Gastroenterology. 2008; 135:1924–1934. [PubMed: 18955056]

17. Fleiss, JL.; Levin, B.; Paik, MC. Statistical Methods for Rates and Proportions. 3rd. New York, NY: John Wiley & Sons; 2003.

18. 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; 46:1323–1330. [PubMed: 8229110]

19. Andrade RJ, Lucena MI, Fernandez MC, Pelaez G, Pachkoria K, Garcia-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. [PubMed: 16083708]

20. Bjornsson E, Kalaitzakis E, Olsson R. The impact of eosinophilia and hepatic necrosis on prognosis in patients with drug-induced liver injury. Aliment Pharmacol Ther. 2007; 25:1411–1421. [PubMed: 17539980]

21. Andrade RJ, Lucena MI, Kaplowitz N, Garcia-Munoz B, Borraz Y, Pachkoria K, et al. Outcome of acute idiosyncratic drug-induced liver injury: long-term follow-up in a hepatotoxicity registry. Hepatology. 2006; 44:1581–1588. [PubMed: 17133470]

22. De Valle MB, Av Klinteberg V, Alem N, Olsson R, Bjornsson E. Drug-induced liver injury in a Swedish University hospital outpatient hepatology clinic. Aliment Pharmacol Ther. 2006; 24:1187–1195. [PubMed: 17014577]

23. Aithal GP, Rawlins MD, Day CP. Clinical diagnostic scale: a useful tool in the evaluation of suspected hepatotoxic adverse drug reactions. J Hepatol. 2000; 33:949–952. [PubMed: 11131457]

24. Lucena MI, Camargo R, Andrade RJ, Perez-Sanchez CJ, Sanchez De La Cuesta F. Comparison of two clinical scales for causality assessment in hepatotoxicity. Hepatology. 2001; 33:123–130. [PubMed: 11124828]

25. Arimone Y, Begaud B, Miremont-Salame G, Fourrier-Reglat A, Moore N, Molimard M, et al. Agreement of expert judgment in causality assessment of adverse drug reactions. Eur J Clin Pharmacol. 2005; 61:169–173. [PubMed: 15827761]

26. Arimone Y, Miremont-Salame G, Haramburu F, Molimard M, Moore N, Fourrier-Reglat A, et al. Inter-expert agreement of seven criteria in causality assessment of adverse drug reactions. Br J Clin Pharmacol. 2007; 64:482–488. [PubMed: 17711539]

27. Arimone Y, Begaud B, Miremont-Salame G, Fourrier-Reglat A, Molimard M, Moore N, et al. A new method for assessing drug causation provided agreement with experts' judgment. J Clin Epidemiol. 2006; 59:308–314. [PubMed: 16488362]

28. Rochon J, Protiva P, Seeff LB, Fontana RJ, Liangpunsakul S, Watkins PB, et al. Reliability of the Roussel Uclaf Causality Assessment Method for assessing causality in drug-induced liver injury. Hepatology. 2008; 48:1175–1183. [PubMed: 18798340]

Abbreviations

ALT	alanine aminotransferase
AP	alkaline phosphatase
AST	aspartate aminotransferase
CRF	case report form

DCC	data coordinating center
DILI	drug-induced liver injury
DILIN	Drug-Induced Liver Injury Network
INR	international normalized ratio
MAD	maximum absolute difference
RUCAM	Roussel-Uclaf Causality Assessment Method
ULN	upper limit of normal

DILIN Causality Adjudication Flowchart

1. Site collects the data from charts and records; completes the CRF.

2. Site sends CRF to the DCC; data entry is performed. Queries are generated.

3. Queries are resolved by the sites and returned to the DCC.

4. Data are declared "Causality Clean"; CRF subset is generated.

5. Clinical narrative is sent by the site investigator to the DCC.

6. CRF subset and clinical narrative are sent to 3 independent reviewers.

7. Case is reviewed by each reviewer; causality scores are returned to the DCC.

8. If complete agreement among reviewers, scores are taken as the final scores.

9. If any disagreement, case is placed on agenda for next Causality Committee

10. If three reviewers reach agreement, consensus scores are taken as final scores.

11. If continued disagreement, final scores are determined by majority vote.

12. Causality adjudication is complete

Fig. 1.
Overview of the DILIN causality adjudication process. A flow diagram of the different steps in the entire causality process is shown.
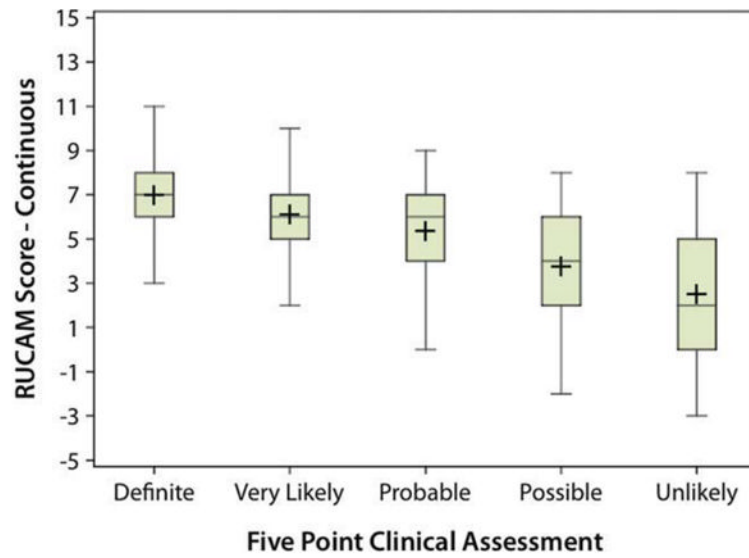
Fig. 2.
Correlation of the RUCAM and DILIN causality scores. A box and whisker plot of the RUCAM score at each level of the DILIN expert opinion score is shown. There is a general relationship between the two scales (r = 0.42 in absolute value), although there is considerable variability in the RUCAM score at each DILIN score. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 1
Clinical Assessment of Causality Scale: Definitions

| Label (Score)	Likelihood	Description
Definite (1)	>95%	The evidence for the drug causing the injury is beyond a reasonable doubt.
Highly likely (2)	75%-95%	The evidence for the drug causing the injury is clear and convincing but not definite.
Probable (3)	50%-74%	The preponderance of the evidence supports the link between the drug and the liver injury
Possible (4)	25%-49%	The evidence for the drug causing the injury is equivocal but present.
Unlikely (5)	<25%	There is evidence that an etiological factor other than a drug caused the injury.

Table 2
Demographic and Clinical Characteristics

Characteristic	Mean ± Standard Deviation/%[*]
Demographics	
Age at DILI onset (n = 249)	48.5 ± 18.6
Age ≥ 55 years (n = 249)	37.3%
Gender: female	57.6%
Ethnicity: not Hispanic or Latino	92.8%
Race (n = 249)	
White	78.3%
Black or African American	13.7%
Prior history	
Prior history of a liver problem	19.2%
Prior allergy to a medication (n = 247)	53.4%
Body mass index at DILI onset (n = 231)	26.5 ± 6.5 kg/m^2
At least one alcoholic drink prior to drug use (n = 248)	52.8%
Days from drug exposure to DILI recognition [median (25th percentile, 75th percentile)]	42 (20, 106.5)
Hospitalized (n = 249)	58.2%
Time from onset of illness to presentation	
<1 week	8.4%
1 week	3.2%
2–4 weeks	27.2%
>4 weeks	61.2%
Prescribed prednisone	18.8%
Death or orthotopic liver transplantation	5%
Selected symptoms and signs	
Jaundice[†]	66.0%
Nausea	63.2%
Anorexia	56.0%
Itching	52.8%
Abdominal pain	49.6%
Vomiting	40.4%
Fever	32.8%
Rash	29.6%
Ascites	10.0%
Hepatomegaly	7.2%
Splenomegaly	2.4%
Lymphadenopathy	1.6%
Peak liver tests (n = 249)	
Peak AST (×ULN)	21.1 ± 34.3
Peak ALT (×ULN)	18.4 ± 19.5
Peak AP (×ULN)	2.9 ± 3.0

Characteristic	Mean ± Standard Deviation/%[*]
Peak serum total bilirubin (×ULN)	9.0 ± 8.9
INR (n = 213)	1.6 ± 1.6
Eosinophil count (n = 57)	200.6 ± 196.8
Pattern of liver injury (n = 249)[‡]	
Hepatocellular (R ≥ 5)	55.4%
Mixed (2 < R < 5)	22.1%
Cholestatic (R ≤ 2)	22.5%

[*] Based on 250 cases unless otherwise indicated.

[†] Jaundice was based on clinical criteria (yellow sclera, skin, or both).

[‡] R is defined as (ALT/ULN)/(AP/ULN) at the time of onset.

Table 3

MAD at the Initial Evaluation of All 250 Cases Using the DILIN Structured Expert Opinion Assessment Method

MAD	Frequency	%
0	67	27
1	108	43
2	51	20
3	24	10

The average was 1.13.

Table 4

Distribution of Initial and Final Scores Using the DILIN Structured Expert Opinion Strategy

DILIN Causality Score	Initial Reviewer Scores (n = 750)		Final Committee Scores (n = 250)	
	Frequency	%	Frequency	%
Definite	241	32.1	78	31.2
Highly likely	305	40.7	102	40.8
Probable	119	15.9	37	14.8
Possible	58	7.7	25	10.0
Unlikely	27	3.6	8	3.2

Initial reviewer scores are based on three reviewers per case; committee scores are based on the final score per case.

Table 5
Distribution of Causality Scores at the Initial Evaluation

A. DILIN Structured Expert Opinion Score (n = 187)[†]

Likelihood	Frequency	%
Definite	196	34.9
Highly likely	213	37.9
Probable	86	15.3
Possible	44	7.8
Unlikely	22	3.9
Total	561	100.0

B. RUCAM (n = 187)[†]

Likelihood	Frequency	%
Highly probable	132	23.7
Probable	220	39.5
Possible	167	30.0
Unlikely	22	4.0
Excluded	16	2.9
Total	557	100.0

[†] Each case was evaluated by three reviewers, and this yielded 561 reviews.

[†] The RUCAM categories were as follows: highly probable, >8; probable, 6 to 8; possible, 3 to 5; unlikely, 1 to 2; and excluded, <1. RUCAM scores were missing for 4 reviews, so there were 557 reviews instead of 561.

Table 6

MAD Among the Reviewers in the Two Causality Scales During the Initial Review Process

MAD	DILIN Structured Expert Opinion Score		RUCAM	
	Frequency	%	Frequency	%
0	50	26.7	34	18.6
1	83	44.4	92	50.3
2	25	18.7	48	26.2
3	19	10.2	8	4.4
4	0	0	1	0.6
Average	1.12		1.18	

This table is restricted to cases in which a single agent was implicated and reviews were available from all three reviewers. There were 187 cases for the causality score and 183 cases for RUCAM.

Table 7

Cross-Tabulation of Initial DILIN Causality Scores Against Categorized RUCAM Scores[a]

DILIN Expert Opinion Process	RUCAM					
	Highly Probable	Probable	Possible	Unlikely	Excluded	Total
Definite	80	72	41	0	0	193
Very Likely	38	100	70	4	0	212
Probable	10	35	33	7	1	86
Possible	2	10	18	8	6	44
Unlikely	2	3	5	3	9	22
Total	132	220	167	22	16	557

[a]This table is restricted to cases in which a single agent was implicated (n = 187 cases). RUCAM scores were missing for 4 reviews, and this resulted in 557 reviews.

Table 8

Condensed Cross-Tabulation of Initial DILIN and RUCAM Scores[a]

	RUCAM		
DILIN Expert Opinion Process	**Yes**	**No**	**Total**
Yes	335	156	491
No	17	49	66
Total	352	205	557

[a]This table is restricted to cases in which a single agent was implicated (n = 187 cases). RUCAM scores were missing for 4 reviews, and this resulted in 557 reviews.

[b]For each DILIN and RUCAM, yes = probable or higher; no = possible or lower.