

# Causality Assessment for Suspected DILI During Clinical Phases of Drug Development

Arie Regev · Leonard B. Seeff · Michael Merz ·  
Sif Ormarsdottir · Guruprasad P. Aithal ·  
Jim Gallivan · Paul B. Watkins

© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** Causality assessment is a critical step in establishing the diagnosis of drug induced liver injury (DILI) during drug development. DILI may resemble almost any type of liver disease, and often presents a serious challenge to clinical investigators and drug makers. The diagnosis of DILI is largely based upon a combination of a compatible clinical course, exclusion of all other reasonable causes, resemblance of clinical and pathological features to known features of liver injury due to the drug (i.e., “drug’s signature”), and incidence of liver injury among patients treated with the drug compared to placebo or comparator. Causality assessment for suspected DILI is currently performed using either evaluation by physicians with expertise in liver disorders (i.e., expert opinion) or standardized scoring instruments such as the Roussel Uclaf Causality Assessment Method (RUCAM). Both approaches are widely used in the post marketing setting. Causality

## Key Points

Causality assessment is a critical step in establishing the diagnosis of drug-induced liver injury (DILI) and in determining the hepatic safety profile of the drug during clinical phases of drug development

Although widely used in the post marketing setting, scoring instruments for causality assessment such as the Roussel Uclaf Causality Assessment Method (RUCAM) have not been systematically validated in clinical trial patients, and may have many limitations when used during drug development

Causality assessment using expert opinion remains the gold standard for causality assessment of suspected DILI, and is the preferred approach for causality assessment during drug development

A. Regev (✉)  
Global Patient Safety, Eli Lilly and Company, Lilly Corporate Center, Drop Code 2121, Indianapolis, IN 46285, USA  
e-mail: regev\_arie@lilly.com

L. B. Seeff  
Einstein Healthcare Network, Philadelphia, PA, USA

M. Merz  
Novartis Institutes for BioMedical Research, Basel, Switzerland

S. Ormarsdottir  
Global Patient Safety, AstraZeneca R&D, Mölndal, Sweden

G. P. Aithal  
National Institute for Health Research Nottingham Digestive Diseases Biomedical Research Unit, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK

J. Gallivan  
Marketed Health Products Directorate, Health Canada, Ottawa, ON, Canada

P. B. Watkins  
The Hamner-University of North Carolina Institute for Drug Safety Sciences, 6 Davis Drive, PO Box 12137, Research Triangle Park, NC 27709, USA  
e-mail: pwatkins@thehamner.org

P. B. Watkins  
Schools of Medicine, Pharmacy and Public Health, University of North Carolina, Chapel Hill, NC, USA

assessment based on expert opinion is considered superior to standardized instruments such as RUCAM, in the setting of drug development, and is currently the preferred approach during clinical trials. There is a need for a systematic revision of RUCAM that will render it more suitable for the setting of clinical trials and drug development. Careful monitoring and meticulous data collection during clinical trials are essential in all cases with established liver injury to allow for a proper causality assessment. A workshop was convened to discuss best practices for the assessment of drug-induced liver injury (DILI) in clinical trials. This publication is based on the conclusions of this workshop.

## 1 Introduction

A workshop was sponsored and organized jointly by the European Innovative Medicines Initiative (IMI) and the Hamner Institute for Drug Safety Sciences (IDSS), with the aim of addressing gaps in current guidance and initiating alignment of liver safety assessment on a global scale.

On November 9, 2012, regulatory experts from the FDA, EMA, Health Canada, and the Japanese National Institute of Health Sciences, with representatives from industry and academia, convened and discussed what could be considered best practices in clinical liver safety assessment, focusing on four key areas: 1) data elements and data standards, 2) methodologies to systematically analyze liver safety data, 3) tools and methods for causality assessment, and 4) liver safety assessment in special populations such as hepatitis and oncology patients.

This section summarizes current issues related to causality assessment during clinical trials as discussed at the workshop, and provides respective recommendations for use in clinical drug development. Causality assessment for suspected drug induced liver injury (DILI) remains a major challenge both in clinical practice and during drug development. In contrast to many other liver disorders, there is currently no specific biomarker or a combination of tests that will establish the diagnosis of DILI and differentiate it from other causes of liver injury. DILI may resemble almost any type of liver disease, and the clinicopathologic spectrum may range from nonspecific injury, to acute and chronic hepatitis, granulomatous liver disease, cholestasis, fatty infiltration, vascular lesions, and hepatic tumors [1]. The diagnosis of DILI is therefore virtually always presumptive, as it is based on clinical assessment and exclusion of other possible causes rather than on absolute criteria and specific diagnostic tests.

Abnormal liver tests may be caused by numerous liver disorders as well as extra-hepatic disorders (Table 1), most of which are considerably more common than is

typical DILI. It is therefore critical to exclude other liver diseases before attributing a liver injury to a drug. Exclusion of other causes requires detailed information pertaining to the patient's clinical course and laboratory data. Failure to test for other causes may result in assigning guilt by association, which may often be erroneous.

## 2 Clinical Threshold for Initiation of Causality Assessment

The threshold for initiating a full evaluation, that would allow causality assessment of suspected DILI during drug development has been a matter of debate. While the FDA [2] has suggested starting a full evaluation when ALT/AST crosses the 3X ULN level, ALP crosses 2X ULN or TBIL crosses 2X ULN, the Health Canada guidance recommends to repeat testing at ALT or AST of 3X ULN and to monitor if the level is unchanged; a full evaluation would only be required if the level is increasing [3]. Others have suggested that an ALT threshold of 3X ULN may be too low and should be increased to 5X ULN [4]. In a recently published consensus paper, an international DILI Expert Working Group [4] has argued that the frequency of detection of mild ALT elevations unrelated to DILI is growing due to the increasing prevalence of NAFLD, and the increased frequency with which liver tests are being performed. In addition, there is a high frequency of transient, mild drug-related ALT elevations which may return to baseline even if therapy is continued. These elevations, which occur with several drugs, may represent true mild liver injury with spontaneous resolution, or "adaptation," but they do not represent clinically important liver injury. The DILIN Expert Working Group has therefore suggested that an ALT threshold of 5X ULN is more likely to exclude clinically unimportant and self-limited events as well as ALT increases related to NAFLD.

In the setting of drug development (particularly early phases) waiting for a threshold of 5X ULN before initiating evaluation in patients that had normal or near normal liver tests at baseline ( $ALT < 1.5X$  ULN), may be problematic, especially since some regulatory agencies have recommended that an ALT value of 5X ULN should be the level for drug discontinuation in clinical trials [3]. However, since most regulators and experts in this field agree on the need to promptly (within 48–72 hours) repeat liver tests when the ALT threshold of 3X ULN is crossed, it may be prudent to assess the need for evaluation based on the trend of the ALT change rather than on the initial abnormal ALT level. If the subsequent ALT level decreases or remains in the 3X ULN range, it is probably appropriate to continue monitoring, and there is

**Table 1** Alternative causes of abnormal liver tests

---

<b>Hepatitis viruses A, B, C, D, E</b> (Can mimic acute hepatocellular DILI)
<b>Other infectious agents (CMV, EBV, HSV)</b> (Typically mild AT elevations, but may be severe in immune suppressed patients)
<b>Nonalcoholic fatty liver disease (NAFLD)</b> (Usually mild to moderate elevations in AT with mild elevations in ALP)
<b>Alcoholic liver disease</b> (Typically AT levels <300 U/L, AST > ALT, bilirubin level may be elevated)
<b>Autoimmune hepatitis</b> (Can mimic acute hepatocellular DILI)
<b>Congestive heart failure</b> (May lead to elevated bilirubin and prolonged INR in addition to elevated AT)
<b>Hypotension/cardiac arrhythmia</b> (Usually very high and rapidly reversed AT spike, often in the presence of heart failure and/or hypoxia)
<b>Systemic infection/sepsis</b> (Generally mild AT elevation; bilirubin may be elevated particularly with gram negative infection)
<b>Wilson's disease</b> (Can mimic acute hepatocellular DILI – hemolysis often present)
<b>Hemochromatosis</b> (Chronic low level AT elevation. Typically would not resemble a hepatocellular DILI episode)
<b>Primary or secondary hepatic tumors</b> (Typically predominantly ALP elevation)
<b>Gallstone disease</b> (Passing a stone can cause a very high ALT spike with a rapid resolution. ALP is generally elevated, and abdominal pain is expected)
<b>Vascular disorder (Budd Chiari, portal vein thrombosis)</b> (Liver enzymes vary. May be acute, subacute or chronic presentation)

---

*AT* aminotransferase, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *ALP* alkaline phosphatase, *DILI* drug induced liver injury, *CMV* cytomegalovirus, *EBV* Epstein-Barr virus, *HSV* herpes simplex virus

no need for a full evaluation. However, if the subsequent ALT level shows a clear trend upwards, it is advisable to initiate a full evaluation to allow for a comprehensive causality assessment. For ALP, it is appropriate to initiate evaluation when the repeated levels remain above 2X ULN. Of note is that the results of the causality assessment have important immediate implications, as they may influence the decision of whether or not the drug should be discontinued. In general, if the cause for liver injury has been established and is clearly not drug-related, there is usually no reason to stop the study drug and it can be continued with close monitoring of liver tests.

The threshold for initiating evaluation in patient with abnormal liver tests at baseline, is a more complicated issue and may need to be individualized based on specific levels and the underlying liver disease. It has been suggested that if the ALT value is abnormal before beginning drug treatment, close monitoring should be initiated if there is a 2-fold increase above baseline values during treatment [2]. In treatment trials of viral hepatitis, the baseline serum ALT may be elevated but may fall quickly in response to viral eradication establishing a new baseline (see Special Populations section.)

An alternative approach for initiating causality assessment for possible DILI is to use multiples of each individual's baseline ALT value rather than using a threshold based on multiples of the ULN. This approach may avoid the issue of somewhat arbitrary and imperfectly selected ULN values, which may vary between different laboratories and patient populations. There is ongoing debate on the potential advantages and utility of this approach, which is beyond the scope of this publication.

## 2.1 Recommendations

1. In patients with normal or near normal liver tests at baseline, an increase of serum ALT/AST to >3X ULN, ALP to >2X ULN or TBIL to >2X ULN should be followed by repeat testing within 48 to 72 hours of all four of the routine serum measures (ALT, AST, ALP, and TBL) to confirm the abnormalities and to determine the trend of the change in liver tests. If there is a clear trend upwards in ALT or AST, and/or if ALP or TBIL levels remain above 2X ULN, an evaluation should be started for alternative causes. Depending on the degree of worsening, the drug might need to be withheld awaiting determination of cause.
2. For subject with elevated values before drug exposure (e.g., baseline ALT levels >1.5X ULN) it is appropriate to start close monitoring when ALT shows a 2-fold increases above baseline, and to initiate evaluation for alternative causes, when subsequent testing shows a clear trend upwards.

## 3 Causality Assessment Methods and Instruments

In general, approaches to causality assessment for suspected DILI can be divided into two main groups: approaches, which rely on expert opinion and those, which rely on causality assessment instruments such as the Roussel Uclaf Causality Assessment Method (RUCAM) [5], and the Maria and Victorino method (M&V) [6], which attempt to standardize and objectivize the process of causality assessment.

### 3.1 Causality Assessment Based on Expert Opinion

The approach used for causality assessment in clinical practice has relied largely on detailed history taking, and comprehensive clinical, and laboratory assessment. The final decision regarding causality assessment is often based on four major components: (1) compatible clinical course, including typical changes in hepatic biochemical tests, compatible extrahepatic manifestations, and typical time to onset and time to resolution (2) exclusion of all other reasonable causes, including viral hepatitis, alcoholic liver disease, gallstone related disorders, autoimmune hepatitis, metabolic and inherited liver diseases and DILI related to a concomitant drug (3) resemblance of clinical and pathological features to known features of liver injury due to the drug in question (i.e., “drug’s signature”) and (4) the incidence of liver injury among patients treated with the drug. The last two components would often not be evident for a new drug candidate in clinical development, particularly in early clinical trials, although a signature may emerge during clinical trials from careful review of liver biochemistry abnormalities in milder cases of liver injury. That is, when severe DILI is associated with a drug either in clinical trials or post-marketing, it is typical for the event to have many of the same characteristics in terms of latency and biochemical profile of milder and often transient liver chemistry abnormalities. Recent genetic studies have also found common risk factors for the severe and the minor and often asymptomatic liver injuries observed in clinical trials, which may indicate common underlying mechanisms [7–9].

When a clinically important liver injury is encountered in a clinical trial, finding similar but milder events in a retrospective examination supports a causal link and, conversely, not finding similar episodes makes a causal link less likely [2].

This causality assessment approach usually requires great expertise in the field of hepatology and therefore, under optimal circumstances, should be performed by hepatology experts. At a minimum, it is advisable to utilize hepatology experts for causality assessment in more severe cases of liver injury such as cases belonging to categories 3–4 (severe injury, liver transplant, or death) according to the following DILI Severity Index recently published by a DILI Expert Working Group [4], although in some cases it may be prudent to seek the advice of such experts for moderate cases as well (category 2)

#### CATEGORY 1, MILD LIVER INJURY

- Elevated alanine aminotransferase (ALT) or alkaline phosphatase (ALP) levels reaching criteria for DILI<sup>1</sup> but total bilirubin (TBL) concentration  $<2 \times$  ULN.

#### CATEGORY 2, MODERATE LIVER INJURY

- Elevated ALT or ALP levels reaching criteria for DILI and TBL concentration  $\geq 2 \times$  ULN, or symptoms consistent with liver injury

#### CATEGORY 3, SEVERE LIVER INJURY

- Elevated ALT or ALP levels reaching criteria for DILI, TBL concentration  $\geq 2 \times$  ULN, and one of the following:
  - International normalized ratio (INR)  $\geq 1.5$
  - Ascites and/or encephalopathy, disease duration  $<26$  weeks, and absence of underlying cirrhosis
  - Other organ failure considered to be due to DILI

#### CATEGORY 4, FATAL OR LIVER TRANSPLANTATION

- Death or liver transplantation due to DILI

Despite its many advantages, causality assessment based on expert opinion is largely subjective, and suffers from interobserver variance [10]. Furthermore, the use of this approach requires the availability of hepatology experts, which is not always feasible. Nevertheless, recent evidence suggests that this approach, with all its limitations, is superior to the proposed standardized methodologies for causality assessment, most notably RUCAM [10]. Moreover, expert opinion was used by several authors as the “gold standard” for development, validation and comparison of existing scoring systems [6, 11]. Even authorities that recommend the use of RUCAM in clinical practice often agree that consensus opinion among hepatologists with expertise in DILI adjudication remains a “gold standard” [4].

When using expert opinion to assess hepatic cases, especially cases of serious liver injury, it is probably advantageous to use three experts to increase the chances of a majority opinion. In general, each expert should perform causality assessment independently, and each expert should be blinded to treatment assignment where feasible. Ideally, during drug development, causality assessment should be performed prior to unblinding the treatment assignment within the company. This avoids the possibility that greater attempts were made to gather relevant data,

<sup>1</sup> Clinical chemistry criteria for DILI, based on the definition of the DILI Expert Working Group [4] include any one of the following: (1) More than or equal to 5X elevation above the ULN for ALT. (2) More than or equal to 2X elevation above the ULN for ALP (particularly with accompanying elevations in concentrations of gamma-glutamyl transpeptidase) in the absence of known bone pathology driving the rise in ALP level. (3) More than or equal to 3X elevation in ALT concentration and simultaneous elevation of TBL concentration exceeding 2X ULN.

such as herbal or dietary supplement use, for subjects revealed to be receiving study drug. Discrepant scores between reviewers can be resolved through subsequent communications in order to achieve consensus, or at least a majority opinion.

### 3.2 Causality Assessment Based on Standardized Instruments

Two causality assessment instruments have been developed in an attempt to standardize the process of causality assessment for liver injury and achieve an objective assessment. Despite some advantages, these instruments suffer from several drawbacks, which limit their accuracy and restrict their routine use. Of the two instruments, the RUCAM is the more well recognized. It is widely used, especially in European countries, and is considered superior to the Maria and Victorino method (M&V) [11]. RUCAM was developed in 1989 by a group of experts at the request of CIOMS (Council for International Organizations of Medical Sciences). It is based on seven criteria that may receive scores ranging from  $-3$  to  $+3$ . The total score for acute liver injury can range between  $-7$  and  $+14$  and be classified in 5 degrees of relatedness: score equal to or less than 0, relationship “Excluded”; 1–2: “Unlikely”; 3–5: “Possible”; 6–8: “Probable”; above 8: “Highly Probable” [5]. A worksheet with assigned scores for the RUCAM is provided in the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) LiverTox website (<http://livertox.nih.gov/Causality.html>).

The RUCAM was used in numerous publications assessing liver injury related to drugs and herbal products, which was diagnosed in routine clinical practice [12–14]. However, RUCAM was mainly developed and validated in patients in the post marketing setting [15]. It was not systematically validated for the drug development setting or in clinical trial patients. Furthermore, the RUCAM score was validated initially using cases with a positive rechallenge [15], but it is now clear that rechallenge can be negative in cases of well-established DILI [17]. In addition, many of the factors included in the RUCAM score are not well defined and are open to variable interpretation, which would often need expertise in the field of hepatology and DILI. In a study by the U.S. Drug-Induced Liver Injury Network (DILIN) [10] that assessed the reliability of RUCAM in well-defined cases of hepatotoxicity, RUCAM was found to have low test-retest reliability (0.54), and low inter-rater reliability (0.45).

RUCAM has additional limitations when used in the setting of clinical trials for new drugs:

1. RUCAM requires previous information on the hepatic effects of the drug, which is often lacking in clinical trials, especially in early phases.
2. RUCAM includes pregnancy and excessive alcohol consumption as potential risk factors for DILI, but both pregnancy and excessive alcohol drinking are often exclusion criteria for most clinical trials. Furthermore, it should be noted that the role of these conditions as general risk factors for DILI has not been clearly established.
3. RUCAM relies heavily on response to readministration of the drug after resolution of liver injury (“re-challenge”). Although this may be justified in clinical practice where a clear benefit of the drug is established, rechallenge is currently not recommended under most circumstances pertaining to clinical trials [16]. Furthermore, a few reports suggest that rechallenge may be negative in a substantial proportion of cases, even when the drug was considered the cause of liver injury [17].

## 4 Causality Assessment Scales

The causality scales used in the medical literature have varied, but most authorities would agree that the key cutoff of concern is whether the study drug probably caused the observed liver injury, or whether a causality link is unlikely. This can be viewed as whether the likelihood of a causal link is greater than 50 % or less than 50 %, respectively. This binary classification may be sufficient for assessments of infrequent cases or at the early stages of clinical development. However, at later stages of clinical development, for example in preparation for a new drug application (NDA) submission, or in the event of serious liver injuries (such as Hy’s Law cases), it may be more appropriate to further subcategorize causation categories. DILIN has defined categories based on percent likelihood [18] (Table 2). These percentage values, although lacking mathematical accuracy, may be helpful in further defining the causal association, particularly by non-hepatologists.

The “Definite” category requires a typical temporal relationship with no competing diagnosis. The event should fit a pattern (“signature”) of liver chemistry elevations observed in other patients treated with the study drug but not the comparator. The “Highly likely” category requires a convincing temporal relationship and no competing diagnosis. Similarly, the “Probable” category requires a temporal relationship considered compatible with drug related injury, with no competing causes or when competing causes are considered less likely than is injury from the study drug.

**Table 2** Causality assessment scoring in the drug-induced liver injury network (DILIN) prospective study [18]<sup>a</sup>

Causality score	Likelihood (%)	Description
1 = Definite	>95	Clinical features of the liver injury are typical for the drug or herbal product ('signature' or pattern of injury, timing of onset, recovery). The evidence for causality is 'beyond a reasonable doubt'
2 = Highly likely	75–95	The evidence for causality is 'clear and convincing' but not definite
3 = Probable	50–74	Causal relationship is supported by 'the preponderance of evidence' as implicating the drug but the evidence cannot be considered definite or highly likely
4 = Possible	25–49	Causal relationship is not supported by 'the preponderance of evidence'; however, one cannot definitively exclude the possibility
5 = Unlikely	<25	The evidence for causality is 'highly unlikely' based upon the available information
Insufficient data	Not applicable	Key elements of the drug exposure history, initial presentation, alternative diagnoses and/or diagnostic evaluation prevent one from determining a causality score

<sup>a</sup> Table 2 copied from Reference [18]

In the "Possible" category there is at least one other reasonable diagnosis that is more likely than the drug to be the cause of liver injury, while in the "Unlikely" category the temporal relationship is atypical for DILI or another etiology is considered to be responsible for the reaction. The "Unassessable" or "Indeterminate" category is reserved for cases that have insufficient data to arrive at a reasonable causality assessment.

The 5-categories scale (consisting of "Unlikely," "Possible," "Probable," "Highly likely" and "Definite") has been favored by DILIN and other academic investigators. However, several drug makers have preferred the use of a 3-category scale, namely "Unlikely," "Possible" and "Probable" (Table 3), for the following reasons: (1) Due to various practical limitations, cases documented during drug development are often not evaluated as completely as the DILIN cases, and are often missing some data. It is rarely possible to reach the same granularity in causality-classification as DILIN with the data available to drug makers. Therefore, it is often impossible to differentiate between "Probable," "Highly likely" and "Definite." (2) From the regulatory perspective, there may be no significant difference in the approach towards cases, which are classified as "Probable" versus those that are classified as "Highly likely" or "Definite." From the practical standpoint, the future implications are often the same. (3) The differentiation between "Probable," "Highly likely" and "Definite" requires reviewers with competence in the fields of liver disease and knowledge of DILI. In many cases, the level of proficiency among assessors of causality in pharmaceutical companies may not be as high as that of DILIN investigators, and the differentiation between "Probable," "Highly likely" and "Definite" may not be a realistic expectation for these assessors. (4) From the statistical standpoint, the 5-categories classification system may be viewed as an **unbalanced rating scale**, which, although suitable for studies such as that of DILIN's, may introduce a potential bias when used during drug development. In the

3-categories system, there is one low category ("Unlikely") one intermediated category ("Possible") and one high category ("Probable"). In contrast, in the 5-category system there is one low category ("Unlikely") one intermediate category ("Possible") and three high categories ("Probable", "Highly likely" and "Definite"). It has been shown that a scale which is balanced towards higher categories is more likely to provoke an apparently higher response level [19, 20]. The unbalanced 5-category scale might therefore provoke higher causality assessment responses, which may lead to skewed assessments, especially when used by non-experts. Although formal recommendations are lacking, it may be preferable for pharmaceutical companies to use a 3-category scale (Probable, Possible, Unlikely) for causality assessment during drug development.

#### 4.1 Recommendations

1. Causality assessment for suspected DILI during clinical trials should utilize expert opinion as much as possible.
2. The experts should, where possible, be blinded to treatment assignment and should perform the causality assessments independently prior to seeking consensus.
3. In milder liver injury cases (meeting criteria for categories 1–2) (See Sect. 3.1), and in the absence of hepatology experts, causality assessment should be performed by two or three physicians who are trained and experienced in the process of causality assessment and are blinded to each other's assessment and to treatment assignment.
4. A cross-pharma training program on causality assessment of liver injury cases should be organized.
5. In most cases pertaining to drug development, a 3-category causality assessment scale may be preferable to a 5-category scale.

**Table 3** A three-category causality assessment scale for suspected drug-induced liver injury (DILI)

Causality score	Likelihood (%)	Description
1 = Probable	50–100	The causality is supported by the preponderance of evidence, and the drug is more likely than not to be the causal agent. Other likely causes have been ruled out with appropriate tests
2 = Possible	25–49	The causality is not supported by the preponderance of evidence; however, one cannot definitively exclude the possibility. Another etiology is more likely to be the cause of liver injury
3 = Unlikely	<25	The evidence for causality is ‘highly unlikely’ based upon the available information. Another etiology is likely to be the cause of abnormal liver tests
Insufficient data	Not applicable	Key elements of the drug exposure history, initial presentation, alternative diagnoses and/or diagnostic evaluation prevent one from determining a causality score

6. In its present form, RUCAM cannot be recommended for causality assessment in clinical trials. There is a great need for a systematic revision of RUCAM that will render it more suitable for the setting of clinical trials and drug development.

### 5 Factors to Consider in a Structured Causality Assessment

Factors to consider during causality assessment should include latency, pattern and course of the reaction as well as resolution on dechallenge. Considering that in phase III trials comparable data for both treatment and control groups are available, it is important to compare frequency and pattern of liver injury between the groups. It is advisable to avoid rigid causality criteria in assessing new drugs, since different mechanisms of liver injury and treatment population susceptibility factors may be at play, which eventually will explain phenotypic differences among hepatotoxic agents in drug development programs. The following should be described for all study enrollees when liver injury is identified:

#### 5.1 Temporal Relationship with the Drug Treatment

Latency: The ‘time of DILI onset’ will be the time from the first exposure to the drug to the first qualifying laboratory tests (see below) except in cases where symptoms directly related to DILI (i.e., right upper quadrant abdominal pain/discomfort, nausea, vomiting, increasing fatigue, jaundice) clearly have preceded the laboratory test [4].

#### 5.2 Pattern and Severity

This is described as hepatocellular, cholestatic or mixed based on the current definition of DILI using the first set of tests when the event meets the threshold for DILI. These patterns can be established using an equation referred to as the ‘R value.’ R is the ratio of serum activity of ALT/serum

activity of ALP. Each activity is expressed as a multiple of ULN, and both should be measured together at the time of recognition of liver injury. Liver injury is designated ‘hepatocellular’ when there is an increase above 3X ULN in ALT alone or when  $R \geq 5$ . Liver injury is designated ‘cholestatic’ when there is an increase above 2X ULN in ALP alone or when  $R \leq 2$ . Liver injury is designated ‘mixed’ when there is an increase above 2X ULN in ALT and an increase in ALP and  $2 < R < 5$ .

It is also important to record the peak of the reaction and define severity according to the existing definitions.

#### 5.3 Resolution on De-Challenge

The degree of decrease in liver enzymes should be characterized. This can be described as the time to reach a decrease in serum enzymes of >50 % from peak value above the ULN or the time to reach baseline levels.

#### 5.4 Exclusion of Other Disorders

Exclusion of other potential causes of liver injury is an essential part of causality assessment. The frequency of specific liver disorders varies widely according to populations and geographies. In general, the frequency of non-drug related causes is often considerably higher than that of DILI. For example, while the incidence of DILI for specific drugs have typically ranged between 1:10,000 and 1:100,000, the US prevalence of hepatitis C alone is 1.8 %. Similar to hepatitis C, the prevalence of hepatitis B and E differ widely between various populations. In Eastern Asia, the prevalence of hepatitis B may be as high as or even higher than 10 %. Although the general prevalence in the US is less than 0.5 %, it has been reported to range between 6–10.5 % among Asian Americans [21, 22]. Importantly, in studies with immune modulating or immune suppressant agents in patients previously infected with HBV (HBsAg or anti-HBc positive), HBV may undergo reactivation, which may lead to severe hepatitis, and occasionally to liver failure [23]. Hepatitis E has been

considered endemic to India and certain areas in Africa and Central America, however, recent evidence suggest that the incidence of hepatitis E is significantly higher in Western countries than had previously been assumed, and may reach 21 % in the US [24]. The likelihood that abnormal liver tests are related to viral hepatitis is therefore relatively high, so that if specific tests for viral hepatitis have not been performed, assigning high causality to a drug is problematic. In one analysis, two thirds of cases initially reported as drug-induced chronic hepatitis were subsequently attributed to chronic hepatitis C [25]. In the U.S. DILIN registry, 3 % of the cases initially believed to be due to DILI were later determined to be a result of acute hepatitis E infection [26]. Viral hepatitis A, B, C and E must therefore be excluded in every patient with a significant abnormality in liver tests during clinical trials. In HBsAg positive patients, acute HDV infection should be ruled out as well. Autoimmune hepatitis should be evaluated by testing for anti-nuclear antibody (ANA), smooth muscle antibody (SMA), immunoglobulin profile, and in certain circumstances (e.g., acute hepatitis in pediatric patients) liver-kidney microsomal antibody (LKM) and soluble liver/liver-pancreas antibody (SLA/LP) [27, 28]. Other infectious agents such as cytomegalovirus (CMV IgM serology) and Epstein-Barr Virus (EBV IgM serology) should also be excluded, particularly when extra-hepatic manifestations are present. Gallstone disease is common in certain populations, and as high as 24 % of cases of AST elevation over 400 U/L have been shown to have an underlying pancreaticobiliary disorder on investigation [29]. Imaging to exclude biliary obstruction is therefore essential in cases of suspected DILI during clinical trials. In addition, hepatic ischemia and hypoxia due to circulatory or cardiac failure can present with acute hepatocellular form of liver injury indistinguishable from DILI [29], and systemic sepsis should also be considered as an alternative diagnosis in an appropriate clinical scenario [30]. In cases with elevated bilirubin, it is imperative to consider the possibility of Gilbert's syndrome, which may affect up to 10 % of the population, and may be mistakenly diagnosed as severe DILI or a Hy's Law case. This requires assessment of both total and direct bilirubin values. In addition, testing for alleles known to be associated with Gilbert's disease should be considered.

However, exclusion of all of these conditions alone does not guarantee that a drug is the causative agent underlying liver injury as acute sero-negative hepatitis of unknown etiology indistinguishable from DILI accounts for substantial proportion of patients developing acute liver failure even in extensively investigated cohorts [31, 32]. Similarly, significant proportion of sero-negative chronic hepatitis is indistinguishable from auto-immune hepatitis in clinical characteristics, response to steroids and natural history [28,

33–35]. Yet, drugs have been associated with otherwise unexplained chronic hepatitis [36]. Herbal products and food supplements are emerging as an important cause of DILI, including severe liver injury and liver failure leading to liver transplantation.

A thorough evaluation for other causes of abnormal liver tests is therefore critical to the process of causality assessment for suspected DILI. It is also good practice to save samples in suspected DILI cases to allow further investigations in the event that uncertainty persists after initial evaluation.

### 5.5 Risk Factors

Generally, the role of most potential risk factors has not been clearly established and is still a matter of debate [37, 38]. Furthermore, information regarding potential risk factors that may increase the susceptibility to DILI is unlikely to be known at the time of the clinical trial. Nevertheless, information regarding factors such as age, gender, ethnicity, body mass index, diabetes, metabolic syndrome, alcohol intake, smoking and other comorbidity should be collected whenever possible.

### 5.6 Extra-Hepatic Features

Clinical symptoms occur at a variable frequency in persons who develop DILI, but the form of presentation can be helpful in defining the 'phenotype' or 'signature' of a particular drug. For example, the presence of fatigue, nausea, vomiting and right upper quadrant pain characteristic of symptomatic hepatitis, points the way to a group of drugs known to cause acute hepatocellular injury. Other drugs might be considered when there are features suggestive of an inflammatory or immune mediated process, such as rash, fever, arthralgia, or eosinophilia. Occasionally other organ involvements (kidney, lungs or pancreas) may accompany the reaction and point to yet other drugs responsible for the injury. Prospective collection of information regarding the presence or absence of these manifestations can aid in the recognition of signatures of a particular drug and may also provide clues to the underlying mechanisms of DILI.

## 6 Conclusions and Recommendations

Causality assessment for suspected DILI in clinical trials should include careful assessment and laboratory tests for viral hepatitis A, B, C and E, autoimmune hepatitis, alcoholic liver disease, gallstone disease, other hepatotoxic drugs (including herbal and over the counter products) and extrahepatic conditions which may lead to abnormal liver



tests, such as systemic infection, hypotension, and heart failure. Based on the clinical picture and laboratory findings the evaluation may also include tests for Epstein Barr Virus (EBV), Cytomegalovirus (CMV) and metabolic or hereditary liver diseases such as Wilson's disease and hemochromatosis. Meticulous data collection during the hepatic event is critical for effective causality assessment and to establish a diagnosis of DILI.

**Acknowledgments** The Innovative Medicines Initiative and the Hamner-University of North Carolina Institute for Drug Safety Sciences sponsored the workshop, part of which is summarized in this article. This article is part of a supplement entitled *Liver Safety Assessment in Clinical Drug Development: A Best Practices Workshop report*, which was guest edited by Drs. Paul B. Watkins, Michael Merz and Mark I. Avigan. The guest editing by Dr. Avigan does not reflect the position of, nor imply endorsement from, the US Food and Drug Administration or the US Government. Dr Paul B. Watkins is chair of the steering committee of the U.S. Drug-Induced Liver Injury Network that uses expert opinion as the method to assess causality. Drs. Watkins, Merz and Avigan did not receive any honoraria for guest editing the supplement. All manuscripts were peer reviewed by Dr. Rolf Teschke. Dr. Rolf Teschke has no conflicts of interest to declare and did not receive any honoraria for peer reviewing the supplement; however, he received a free yearly online subscription to the journal *Drug Safety*.

The Innovative Medicines Initiative (<http://www.imi.europa.eu/>) is a public-private partnership set up by the European Commission in 2008 to relieve the bottlenecks in drug development and to provide economic stimulus. With a €2 billion commitment, the IMI now has an important portfolio of projects where experts from academia, industry and regulatory bodies collaborate on an unprecedented scale and at a non-competitive level to develop tools and technologies. Drug-induced liver injury has been a focus of several projects including the SAFE-T (Safer and Faster Evidence-based Translation) consortium, which is working on clinical qualification of new biomarkers to better detect and characterize liver toxicity, and MIP-DILI, which is working to determine the optimal preclinical testing to detect potential of liver injury in patients.

The Hamner-University of North Carolina Institute for Drug Safety Sciences (IDSS – <http://www.thehamner.org/idss/>), located in Research Triangle Park, NC, is dedicated to solving drug safety challenges through a variety of innovative approaches including mouse genetics, mechanistic biomarkers, and culture models derived from induced pluripotent stem cells. Efforts in drug-induced liver injury include the DILI-sim Initiative, a public-private partnership developing computer models to explain and predict drug-induced liver injury.

**Conflict of interest** Arie Regev, Leonard B. Seeff, Michael Merz, Sif Ormarsdottir, Guruprasad P. Aithal, and Jim Gallivan, have no conflicts of interest that are directly relevant to the content of this article. Paul B. Watkins chairs the Steering Committee for the US Drug-Induced Liver Injury Network, which employs expert opinion as the preferred causality assessment method.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Chitturi S, Farrell GC. Drug induced liver disease. In: Schiff ER, et al., editors. Schiff's diseases of the liver. 11th ed. New York: Wiley-Blackwell; 2012. p. 703–82.
- FDA. Drug-induced liver injury: premarketing clinical evaluation. In: Guidance for industry. 2009. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM174090.pdf>. Accessed 21 January 2013.
- Health Canada. Guidance document: pre-market evaluation of hepatotoxicity in health products. 2012. [http://www.hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/hepatotox\\_guide\\_ld-eng.php](http://www.hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/hepatotox_guide_ld-eng.php). Accessed 13 February 2012.
- Aithal GP, Watkins PB, Andrade RJ, et al. Case definition and phenotype standardization in drug-induced liver injury. *Clin Pharmacol Ther.* 2011;89(6):806–15.
- Danan G, Benichou C. Causality assessment of adverse reactions to drugs: I. A novel method based on the conclusions of international consensus meetings: applications to drug induced liver injury. *J Clin Epidemiol.* 1993;46:1323–30.
- Maria VA, Victorino AJ. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. *Hepatology.* 1997;26:664–9.
- Kindmark A, Jawaid A, Harbron CG, et al. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *Pharmacogenomics J.* 2008;8(3):186–95.
- Spraggs CF, Budde LR, Briley LP, et al. HLA-DQA1\*02:01 is a major risk factor for lapatinib-induced hepatotoxicity in women with advanced breast cancer. *J Clin Oncol.* 2011;29(6):667–73.
- Singer JB, Lewitzky S, Leroy E, et al. A genome-wide study identifies HLA alleles associated with lumiracoxib-related liver injury. *Nat Genet.* 2010;42(8):711–4.
- Rochon J, Protiva P, Seef LB, et al. Reliability of the Roussel Uclaf causality assessment method for assessing causality in drug-induced liver injury. *Hepatology.* 2008;48:1175–83.
- Lucena MI, Camargo R, Andrade RJ, et al. Comparison of two clinical scales for causality assessment in hepatotoxicity. *Hepatology.* 2001;33:123–30.
- Teschke R, Frenzel C, Schulze J, Schwarzenboeck A, Eickhoff A. Herbalife hepatotoxicity: evaluation of cases with positive reexposure tests. *World J Hepatol.* 2013;5(7):353–63.
- Teschke R, Frenzel C, Wolff A, Eickhoff A, Schulze J. Drug induced liver injury: accuracy of diagnosis in published reports. *Ann Hepatol.* 2014;13(2):248–55.
- Björnsson ES, Bergmann OM, Björnsson HK, Kvaran RB, Olafsson S. Incidence, presentation, and outcomes in patients with drug-induced liver injury in the general population of Iceland. *Gastroenterology.* 2013;144(7):1419–25.
- Benichou C, Danan G, Flahault A. Causality assessment of adverse reactions to drugs-II. An original model for validation of drug causality assessment method: case reports with positive rechallenge. *J Clin Epidemiol.* 1993;4(6):1331–6.
- Papay JI, Clines D, Rafi R, et al. Drug-induced liver injury following positive drug rechallenge. *Regul Toxicol Pharmacol.* 2009;54(1):84–90.
- Hunt CM. Mitochondrial and immunoallergic injury increase risk of positive drug rechallenge after drug-induced liver injury: a systematic review. *Hepatology.* 2010;52(6):2216–22.
- 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.

19. Brown G, Copeland T, Willward M, et al. Monadic testing of new products—an old problem and some partial solutions. *J Market Res Soc.* 1973;15:112–31.
20. Friedman HH, Amoo T. Rating the rating scales. *J Market Manag.* 1999;9:114–23.
21. Kowdley KV, Wang CC, Welch S, et al. Prevalence of chronic hepatitis B among foreign-born persons living in the United States by country of origin. *Hepatology.* 2012;56:422–33.
22. Hann HW, Hann RS, Maddrey WC. Hepatitis B virus infection in 6,130 unvaccinated Korean–Americans surveyed between 1988 and 1990. *Am J Gastroenterol.* 2007;102(4):767–72.
23. Mindikoglu AL, Regev A, Schiff ER. Hepatitis B virus reactivation after cytotoxic chemotherapy: the disease and its prevention. *Clin Gastroenterol Hepatol.* 2006;4:1076–81.
24. Kunihom MH, Purcell RH, McQuillan GM, et al. Epidemiology of hepatitis E virus in the United States: results from the third National Health and National Health and Nutritional Examination survey, 1988–1994. *JID.* 2009;200:48–56.
25. Laurent-Puig P, Dussaix E, de Paillette L, et al. Prevalence of hepatitis C RNA in suspected drug induced liver injury (letter). *J Hepatol.* 1993;19:487–9.
26. Davern TJ, Chalasani N, Fontana RJ, et al. Acute hepatitis E infection accounts for some cases of suspected drug-induced liver injury. *Gastroenterology.* 2011;141:1665–72.
27. Desmet VJ, Gerber M, Hoofnagle JH, et al. Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology.* 1994;19:1513–20.
28. Czaja AJ. Acute and acute severe (fulminant) autoimmune hepatitis. *Dig Dis Sci.* 2013;58:897–914.
29. Whitehead MW, Hawkes ND, Hainsworth I, et al. A prospective study of the causes of notably raised aspartate aminotransferase of liver origin. *Gut.* 1999;45(1):129–33.
30. Aithal GP, Rawlins MD, Day CP. Accuracy of hepatic adverse drug reaction reporting in one English health region. *BMJ.* 1999;319(7224):1541.
31. Bernal W, Ma Y, Smith HM, et al. The significance of autoantibodies and immunoglobulins in acute liver failure: a cohort study. *J Hepatol.* 2007;47(5):664–70.
32. Marudanayagam R, Shanmugam V, Gunson B, Mirza DF, et al. Aetiology and outcome of acute liver failure. *HPB (Oxford).* 2009;11(5):429–34.
33. Kaymakoglu S, Cakaloglu Y, Demir K, et al. Is severe cryptogenic hepatitis similar to autoimmune hepatitis? *J Hepatol.* 1998;28(1):78–83.
34. Gassert D, Garcia H, Tanako K, et al. Corticosteroid-responsive cryptogenic chronic hepatitis: evidence for seronegative autoimmune hepatitis. *Dig Dis Sci.* 2007;52(9):2433–7.
35. Heringlake S, Schütte A, Flemming P, et al. Presumed cryptogenic liver disease in Germany: high prevalence of autoantibody-negative autoimmune hepatitis, low prevalence of NASH, no evidence for occult viral etiology. *Z Gastroenterol.* 2009;47(5):417–23.
36. Mathiesen UL, Franzén LE, Frydén A, et al. The clinical significance of slightly to moderately increased liver transaminase values in asymptomatic patients. *Scand J Gastroenterol.* 1999;34(1):85–91.
37. Rosenberg P, Urwitz H, Johannesson A, et al. Psoriasis patients with diabetes type 2 are at high risk of developing liver fibrosis during methotrexate treatment. *J Hepatol.* 2007;46:1111–8.
38. Chalasani N, Björnsson E. Risk factors for idiosyncratic drug-induced liver injury. *Gastroenterology.* 2010;138:2246–59.