Recommendations for Analyses and Displays Associated with Hepatotoxicity with a Focus on Phase 2-4 Clinical Trials and Integrated Submission Documents Stage 1: Initial Assessment and Screening for Potential for Hepatotoxicity

20 NOVEMBER 2024

Revision History

Version	Date	Summary
1.0	20 NOV 2024	Guideline and Best Practice for Key Analyses and Displays Associated with Hepatotoxicity with a Focus on Phase 2-4 Clinical Trials and Integrated Submission Documents Stage 1: Initial Assessment of Potential for Hepatoxicity

Contents

1 Overview: Purpose of this document

The purpose of this document is to provide recommendations for displaying, summarizing, and analyzing measures of hepatotoxicity in tables, figures, and listings (TFLs). The suggested TFLs will have implications for what and how liver-related data should be collected. If the data required to populate these TFLs are not collected, then there may be insufficient information to adequately assess the potential for a drug to cause or contribute to the cause of hepatotoxicity. Of note, threshold analyses, such as analyses that focus on excesses beyond specified limits relative to the upper limit and/or lower limit of normal, play an important role relative to analyses of central tendency. However central tendency analyses should also be considered in the overall assessment of hepatic changes in the lab parameters. The development of standard TFLs and associated analyses will lead to improved standardization of hepatotoxicity data from collection through data display and analysis. It will also harmonize product lifecycle management across therapeutic areas by ensuring that reviewers receive clinically relevant and meaningful analyses of patient safety for benefit-risk assessment.

Another purpose of this document is to improve expertise in liver safety assessment across the multiple disciplines involved with planning, interpreting, and reporting safety analyses. This requires increased cross-functional engagement and cross-disciplinary safety management teams to think more quantitatively and requires nonstatistical disciplines to obtain a higher level of analytical knowledge. Solid understanding of the underlying statistical methods, assumptions, and data limitations is essential for appropriate application to safety data analyses. As such, strong statistical guidance throughout the planning and review of DILI safety data is highly recommended. Throughout this document, we shall use the terms hepatotoxicity and drug- induced liver injury (DILI) interchangeably.

2 Scope

The need to carefully plan, execute analyses, and assess hepatotoxicity is widely discussed in the literature and some agencies have issued guidelines on the assessment of hepatotoxicity (Merz *et al.*, 2014, Avigan *et al.*, 2014, Health Canada Guidance on pre-market evaluation of hepatotoxicity, 2012, FDA Guidance on Premarketing clinical evaluation of DILI, 2009, FDA Draft Guide on Advancing Premarket Safety Analytics, 2022, and CIOMS X Guideline, 2020). The advice provided in this document is relevant to all therapeutic areas and also applies to Phase 2-4 clinical trials and integrated analyses as well as some Phase 1 studies or other types of medical research (e.g., observational studies). Detailed variable specifications for TFLs or dataset development are out of scope and will not be discussed.

It is recommeded that liver safety assessment be broken down into two stages (Table 1). This document will focus on Stage 1 (screening stage) recommendations.

Table 1: Stages of Liver Safety Assessment

Stage	Description
Stage 1	Initial assessment/screening of hepatotoxicity: • Create a set of analyses and outputs along with medical judgement to assess a drug's potential for DILI • Focus on aggregate statistical outputs in Phase 2 and later phases and integrated

	analyses • Assess and interpret these data and determine if there is a need to go to Stage 2
Stage 2	Additional analyses/outputs to further characterize potential drug-induced hepatotoxicity: • Additional analyses and outputs will need to be generated to characterize the potential DILI issue and assess causality • Assessments with both aggregate data and subject-level data • Assessments may need to be done in real time once a potential DILI case has been identified

3 Acronyms, Abbreviations, and Definitions

Table 2: Acronyms

Term	Definition
ADaM	Analysis Data Model
AE	adverse event
AESI	adverse event of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
AT	aminotransferase
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CIOMS	Council for International Organizations of Medical Sciences
CTCAE	Common Terminology Criteria for Adverse Events
DILI	drug-induced liver injury
EAIR	exposure-adjusted incidence rate
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
MedDRA	Medical Dictionary for Regulatory Activities
PHUSE	Pharmaceuticals User Software Exchange
PT	Preferred Term
SAE	serious adverse event
SAP	statistical analysis plan
SDTM	Study Data Tabulation Model
SOC	System Organ Class
SMQ	Standardised MedDRA Query
TBL	total bilirubin
TEAE	treatment-emergent adverse event
TFL	tables, figures, and listings
TRT	Treatment
ULN	upper limit of normal
WHO	World Health Organization

4 Problem Statement

Industry standards for data collection and storage have evolved over time resulting in the following: Clinical Data Acquisition Standards Harmonization (CDASH), Study Data Tabulation Model (SDTM), and Analysis Data Model (ADAM) across data domains. Additionally, some suggestions have been made to ascertain that appropriate data are collected for downstream analysis and assessment a drug's potential on liver safety (Avigan *et al*, 2014, Pei *et al*, 2023). However, efforts to create industry standards for analyses and displays are a recent phenomenon. Global harmonization of these efforts can lead to more optimal and consistent displays for all stakeholders in the drug development and data analytic space. As scientific thinking evolves, recommendations outlined in the white paper will require further modifications and/or addendums using a collaborative approach.

5 General Background

The PHUSE Computational Science Collaboration is an initiative involving PHUSE, the FDA, and industry stakeholders that identifies computational science priorities that could be addressed by collaboration, crowd sourcing, and innovation. Several working groups have been created within PHUSE to address many of these challenges (https://advance.phuse.global/display/WEL/Working+Groups). The Safety Analytics Working Group (https://advance.phuse.global/display/WEL/Safety+Analytics) within which this project falls under had overall responsibility of the development of this white paper.

Members of the Analyses and Displays for Hepatotoxicity White Papers Project Team reviewed regulatory guidances and shared ideas and lessons learned from their combined experience. They also reviewed a wide variety of common practices in assessing hepatotoxicity. Contributors and reviewers of the white paper include a cross functional team of physicians and statisticians from PHUSE, the FDA, and industry stakeholders. Additional input to this white paper was obtained via public review of the document prior to its final release.

6 Hepatotoxicity Background

DILI is one of the most frequent causes of drug withdrawals from the US market, and several drugs have not been approved in the US because European marketing experience revealed hepatotoxicity. Drugs can cause liver injury by many different mechanisms, and clinical presentation can resemble almost all known liver diseases. There are no objective blood, imaging, or histologic findings diagnostic of DILI. Therefore, the diagnosis relies on clinical judgement and exclusion of other liver injury causes by thorough clinical evaluation.

Most commonly, severe or acutely fatal DILI is hepatocellular. Hepatocellular injury is typically associated with substantial increases in serum ALT (alanine aminotransferase) and/or AST (aspartate aminotransferase) levels reflecting the release of these aminotransferases (AT) from the cytoplasm of damaged or necrotic hepatocytes. However, many drugs can cause transient rises in serum AT activity without causing severe DILI, even if drug administration is continued. The challenge is discerning such less concerning drugs from those that cause progressive injury resulting in hepatic failure signified by jaundice, coagulopathy, and symptoms (e.g., encephalopathy). Identification of more progressive liver injury risk is usually based on the severity of AT elevations. Increase in bilirubin is even more specific for severe injury risk because it indicates a decline in the liver's ability to process bilirubin (i.e., loss of

function). However, timely identification of more worrisome liver injury and discontinuance of the drug does not always prevent hepatic failure.

Usually New Drug Applications (NDA) or Biologics License Application (BLA) will not have any cases of severe (fatal or transplant requiring) DILI because the clinical trials are not powered to detect low rates of severe injury (e.g., 1/10,000 or less) which have been associated with drug withdrawals from the market. However, such drugs will often cause elevations in AT and bilirubin that are harbingers of severe DILI in a larger post-market population. Estimating this risk based on such lab abnormalities is based on Dr. Hy Zimmerman's astute observation that hepatocellular liver injury with consequent jaundice carries at least a 10% mortality risk (Zimmerman 1978, 1999). This observation, later coined by others as Hy's Law (FDA, 2009), has been a benchmark for DILI risk assessment by the FDA since the early 2000s (Temple 2001; Reuben 2004, FDA Guidance on Premarketing clinical evaluation of DILI, 2009). Retrospective analyses of NDAs for drugs later removed from the market indicated that just one to two subjects meeting Hy's Law was enough to raise concerns for liver safety post-market.

FDA's current guidance on criteria for assessing potential Hy's Law (2009) are as follows:

1. The drug causes hepatocellular injury, generally shown by a higher incidence of > 3 × ULN or greater elevations above the ULN of ALT or AST than the (non-hepatotoxic) control drug or placebo.
2. Among trial subjects showing such AT elevations, often with AT elevations much greater than 3xULN, one or more also show elevation of serum total bilirubin (TBL) to > 2 × ULN, without initial findings of cholestasis (elevated serum alkaline phosphatase (ALP)).
3. At the subject-level, no other reason can be found to explain the combination of increased AT and TBL, such as viral hepatitis A, B, or C; preexisting or acute liver disease; or another drug capable of causing the observed injury.

Table 3: Hy's Law

Term	Definition
Hy's Law	A term based on the clinical observation that a patient is at high risk of fatal drug-induced liver injury if given a medication that causes hepatocellular injury with jaundice. The law is based on observations by Hyman Zimmerman, a major scholar of drug-induced liver injury (DILI). According to the current FDA guidance document, Hy's Law cases have three components: • The drug causes hepatocellular injury, generally defined as an elevated ALT or AST by greater than 3 above the upper limit of normal (> 3 × ULN). Often the aminotransferases are increased much greater than 5-10 × ULN. • Among subjects showing such aminotransferase elevations, they also have elevation of their serum total bilirubin (TBL) of > 2 × ULN, without findings of predominant cholestatic injury (defined as substantial elevations of serum alkaline phosphatase). • No other reason can be found to explain the combination of increased aminotransferase and serum total bilirubin (TBL), such as viral hepatitis, alcohol abuse, ischemia, preexisting liver disease, or another drug capable of causing the observed injury.

The finding of a higher rate of AT elevations in drug-treated subjects than in a control group is supportive of a potential to cause severe DILI risk, but detection of such an imbalance of Hy's Law cases is more difficult due to under-powering.

7 General Considerations and Recommendations

7.1 P-values and Confidence Intervals

Detailed discussions of appropriate utility of p-values and confidence intervals in safety assessments are out of scope for this white paper. While we report out p-values and confidence intervals, they should not be taken as definitive for making regulatory decisions or interpretations of their displays on account of the many challenges that come with statistical inference in drug safety data (Singh and Loke, 2012). For example, an imbalance in Temple's quadrant rarely reaches statistical significance, but this may still be considered a finding medically important.

7.2 Importance of Visual Displays

Communicating information effectively and efficiently is crucial to enable decision-making. Current practice, which focuses on tables, has not always allowed us to communicate information effectively, since tables and listings may be very long and repetitive. Graphics, on the other hand, can provide more effective presentation of complex data. They may also better identify outlier values. Standardized presentation of visual information is encouraged. While this white paper focuses on static displays, we include notes for areas where interactive visual capabilities would be beneficial, and the displays in this white paper may serve as starting points for such future interactive programming.

7.3 Integrated Analyses

Principles for pooling data across studies and some integrated analysis methods are well established and also apply to assessment of liver safety (Reviewer Guidance Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review, 2005). In the FDA DILI guidance, it is noted that controlled studies should be more relied upon than open-label studies. When randomization ratios across studies and treatment exposure are imbalanced, it is important to carry out appropriate analysis adjustments and use appropriate metrics. Assessing results within each study, in addition to assessing the pooled results, is critical because studies can vary in several ways including size, duration of exposure, length of randomized controlled portions and comparator arms (e.g., marketed drug versus placebo).

7.4 Number of Therapy Arms

In this version of the white paper, the example TFLs show one or two treatment arms versus a placebo (or comparator) arm. Most TFLs can be easily adapted to include multiple arms or a single arm, as needed. The example TFLs for individual studies show two treatment arms and a comparator arm within a controlled phase of a study. The example TFLs for integrated summaries show one treatment arm (assumes all of the treated arms are pooled) and a comparator arm within the controlled phase of

the studies. Many of the TFLs recommended in this white paper can be adapted to display data from additional phases and/or treatment arms. Appropriate discussions among key subject matter experts and if necessary regulatory agencies should take place whether to include or exclude data from open label extension studies.

7.5 Planned versus Unplanned Measurements

Planned measurements of clinical laboratory data refer to those measurements that are collected based on the timelines specified in the protocol. Unplanned laboratory measurements are any additional laboratory measurements that are collected and not specified in the protocol. Unplanned safety measurements can arise for various reasons. During a study, the clinical investigator sometimes due to clinical concerns orders a repeat test, or retest, of a laboratory test, especially if the investigator has received an unexpected value. In general, retests are repeat tests performed because an initial test result had an unexpected value. The repeat result may either confirm the initial test results or (less commonly) suggest that a laboratory error occurred in the case of the initial result. Retests are often performed to verify that the action taken by the investigator (e.g., changing the dose of the study drug as allowed by the protocol) has the desired effect (e.g., test results have returned to within reference limits). If such retests are conducted until desired measurement results have been reached, analyses from baseline to last observation would be biased toward normality. Another common scenario is that the patient developed symptoms prior to scheduled visit or needs to go to a hospital emergency room for new onset liver safety-related symptoms. These labs are sometimes and often not included in the submission. Thus, we recommend including both planned and unplanned measurements when possible and where it makes sense when creating displays or conducting analyses over time and when assessing change from baseline to endpoint. Both planned and unplanned measurements should be included when conducting analyses that focus on outliers (categorical analyses) or shifts across an entire period, because these are intended to focus on the most extreme changes.

7.6 Central Versus Local Laboratories Data

In recent years, most large studies have utilized a central laboratory to ensure consistency in laboratory assessments across institutions. However, there are times when this is not feasible and which may be important for overall assessment of liver safety such as when different units are used or when final laboratory results cannot be made available in electronic format. For example, some studies may need to utilize local laboratories due to the nature of the study. There are also cases where the scheduled laboratory tests are done using a central laboratory, but ad hoc laboratory results are performed locally for patient care. Generally, results from different laboratories should be combined for DILI risk analyses. In this setting, there should be careful review of laboratory assay methods, and laboratory reference limits determination methods are clinically comparable, and the reference limits for all analytes are comparable and consistent.

7.7 Adverse Events of Interest

Although the focus of the Stage 1 assessment of hepatotoxicity is on clinical laboratory data, this can be supplemented by an assessment of adverse events data through the use of a Standardised MedDRA Query (SMQ) or FDA Medical Query (FMQ). Displays will show incidence for each treatment arm and preferred terms using the SMQ or FMQ.

8 Thresholds and List of Tables and Figures

8.1 Thresholds for Clinical Laboratory Data in Hepatotoxicity Assessment

Below are recommended thresholds that can be used across therapeutic areas, particularly in the spirit of standardization and harmonization. It is understood that teams can add to these thresholds as deemed appropriate for some of the therapeutic areas and/or drug classes, see for example, Parks et al (2013).

Lab Parameter	Thresholds
ALT	• > 3 × ULN • > 5 × ULN • > 10 × ULN • > 20 × ULN
AST	• > 3 × ULN • > 5 × ULN • > 10 × ULN • > 20 × ULN
ALP	• > 1.5 × ULN • > 2 × ULN • > 3 × ULN
Total Bilirubin	• > 1.5 × ULN • > 2 × ULN • > 5 × ULN • > 8 × ULN

8.2 Recommended Tables and Figures for Stage 1

As noted previously, Stage 1 is focused on initial assessment of hepatotoxicity and on analyses and displays that can be used, along with medical judgement, to help assess for potential DILI. This assessment leverages aggregate statistical outputs in Phase 2-3 studies and integrated analyses. This initial assessment and interpretation of the data is then used to determine if there is need to proceed to Stage 2. For this reason, the scope of the outputs is focused on key primary aggregate tables and figures. This set of tables does not preclude the team from adding and/or exploring additional tables and figures if that becomes necessary. In essence these tables and figures are intended to address both quantitative and medical assessment and questions on liver safety in Stage 1. There can include questions such as:

- Are there patients meeting biochemical criteria for evaluation of Hy's law (e.g., ALT or AST > 3 × ULN and total bilirubin > 2 × ULN)?
- Are there markedly abnormal values? Are there marked shifts from baseline?
- Is there medical etiology in unusual findings?

In the remaining part of this section, a list of core tables and figures intended for Stage 1, along with suggested content, are provided.

8.3 List of Tables and Figures

Type	Title
Table 1	Overview of Reported Liver Safety Findings
Table 2	Elevations in Liver Biochemical Tests Based on Thresholds
Table 3	Shifts in Liver Biochemical Tests Based on Thresholds
Table 4	Summary of Treatment Emergent Adverse Events of Adverse Events Meeting the Liver Safety SMQ (or FMQ)
Figure 1	Plot of Maximum Total Bilirubin Versus Maximum ALT or AST or ALP
Figure 2	Scatterplot of Baseline Versus Maximum Values for Liver Biochemical Tests
Figure 3	Boxplots of for Liver Biochemical Tests
Figure 4	Forest Plot of DILI Adverse Events Based on SMQ (or FMQ)
Figure 5	Patient Profiles of Subjects of Interest

Table 1
Overview of Reported Potential Liver Safety Findings
<Analysis Set>

	PL (N=XXX) n/N (%)	T1 (N=XXX) n/N (%)	T2 (N=XXX) n/N (%)	Risk Difference T1-PL (95% CI)	Risk Difference T2-PL (95% CI)	Risk Difference T2-T1 (95% CI)
Number of Subjects with Liver Safety Findings [a]						
Any Potential Liver Safety Findings	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
Findings Related to Adverse Events	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
Findings Related to Biochemical Tests	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
Findings Related to Both Adverse Events and Biochemical Tests	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)

PL=Placebo, T1=Treatment 1, T2=Treatment 2; N = number of applicable subjects in treatment arm; n = actual applicable observed number with values.
Note [q]: A potential liver safety finding is defined as meeting any of the laboratory thresholds.

Table 2
Elevations in Liver Biochemical Tests Based on Thresholds
<Analysis Set>

Criteria	PL (N=XXX) n/N (%)	T1 (N=XXX) n/N (%)	T2 (N=XXX) n/N (%)	Risk Difference T1-PL (95% CI)	Risk Difference T2-PL (95% CI)	Risk Difference T2-T1 (95% CI)
ALT						
> 3× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 5× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 10× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 20× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
AST						
> 3× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 5× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 10× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 20× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
Total Bilirubin						
> 2× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 5× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 8× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
ALP						
> 2× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
> 3× ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
ALT or AST and Total Bilirubin						
ALT > 3× ULN and Total Bilirubin > 2 × ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
AST > 3× ULN and Total Bilirubin > 2 × ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)
ALT or AST > 3× ULN and Total Bilirubin > 2 × ULN	XX/XXX (XX.X)	XX/XXX (XX.X)	XX/XXX (XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)	XX (XX.X, XX.X)

PL=Placebo, T1=Treatment 1, T2=Treatment 2, ALT=alanine aminotransferase, AST=aspartate aminotransferase, ALP= alkaline phosphatase; N = number of applicable subjects in treatment arm; n = actual applicable observed number with values.

Table 3
Shifts in Liver Biochemical Tests Based on Thresholds
<Analysis Set>

<ALT><AST><Total Bilirubin><ALP>	Maximum Baseline Value	Maximum Post-Baseline Result					
		< 3× ULN	**≥ 3 to < 5× ULN**	**≥ 5 to < 10 × ULN**	**≥ 10 to < 20 × ULN**	**≥ 20 × ULN**	**Missing**
PL (N=XXX)	< 3× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 3 to < 5× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 5 to < 10 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 10 to < 20 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 20× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	Missing	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
T1 (N=XXX)	< 3× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 3 to < 5× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 5 to < 10 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 10 to < 20 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 20× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	Missing	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
T2 Comparator (N=XXX)	< 3× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 3 to < 5× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 5 to < 10 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 10 to < 20 × ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	≥ 20× ULN	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)
	Missing	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX (XX.X)

	N	n	%	p-value
PL	XXX	XXX	XX (XX.X)	0.XXX
T1	XXX	XXX	XX (XX.X)	0.XXX
T2	XXX	XXX	XX (XX.X)	0.XXX

PL=Placebo, T1=Treatment 1, T2=Treatment 2, ALT=alanine aminotransferase, AST=aspartate aminotransferase, ALP= alkaline phosphatase; N = number of applicable subjects in treatment arm; n = actual applicable observed number with values.

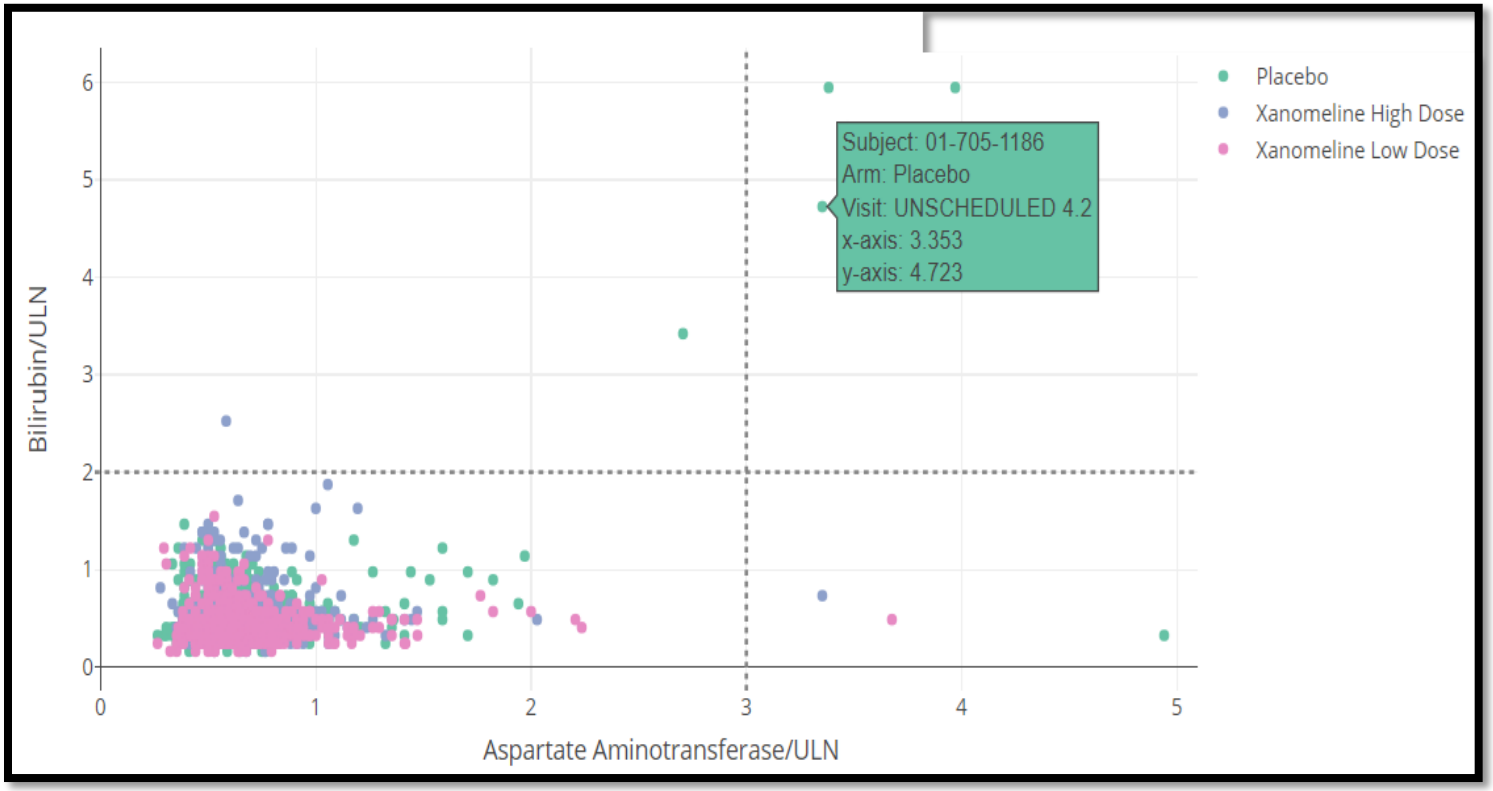
Table 5

Summary of Adverse Events Meeting DILI-Related SMQs Defined Using Standardized MedDRA Query

<Analysis Set>

Preferred Term (N/	Placebo (N=XXX)	Drug (N=XXX)	Comparator (N=XXX)	Risk Difference Drug – Placebo (95% CI)	Risk Difference Comparator – Placebo (95% CI)
Number of patients reporting at least one Narrow or Broad Scope PT	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
Number of patients reporting at least one Narrow or Broad Scope PT	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
Narrow Scope PTs					
[Preferred Term #1]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
[Preferred Term #2]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
[Preferred Term #n]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
Number of patients reporting at least one Narrow or Broad Scope PT	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
Narrow Scope PTs					
[Preferred Term #1]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
[Preferred Term #2]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)
[Preferred Term #n]	XX (XX.X)	XX (XX.X)	XX (XX.X)	XX.X (XX.X, XX.X)	XX.X (XX.X, XX.X)

Figure 1
Plot of Maximum Total Bilirubin Versus Maximum ALT or AST
<Analysis Set>

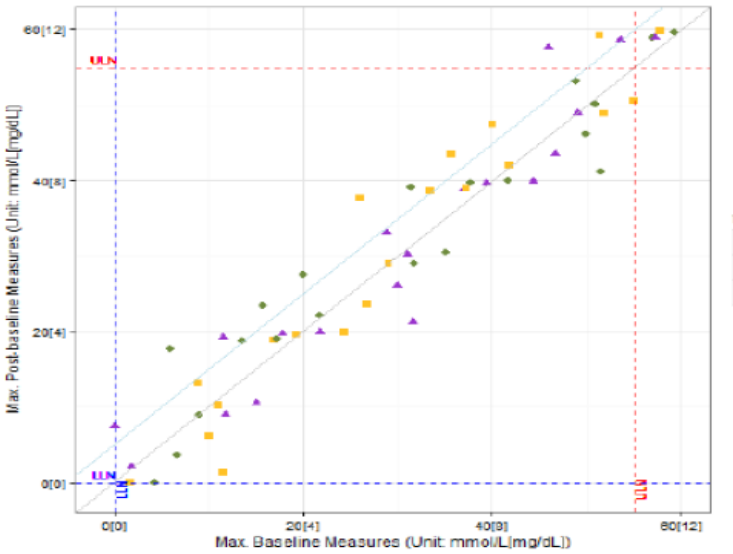


Recommend an interactive version of this plot.

Figure 2
Scatterplot of Baseline Versus Maximum Values for Liver Biochemical Tests
<Analysis Set>

Figure 17.5 Scatterplot and Shift Summary for Quantitative Safety Measures Assessing High Value with Change Criteria: Individual Study

Scatterplot and Shift Table Summary of Absolute Lab Values for Lab Test 1
Max baseline vs. Max Post-baseline



Treatment	Treatment Emergent High			
	N	n	%	P value*a
T1	xxx	xxx	xx.x	.xxx
T2	xxx	xxx	xx.x	.xxx
PL	xxx	xxx	xx.x	

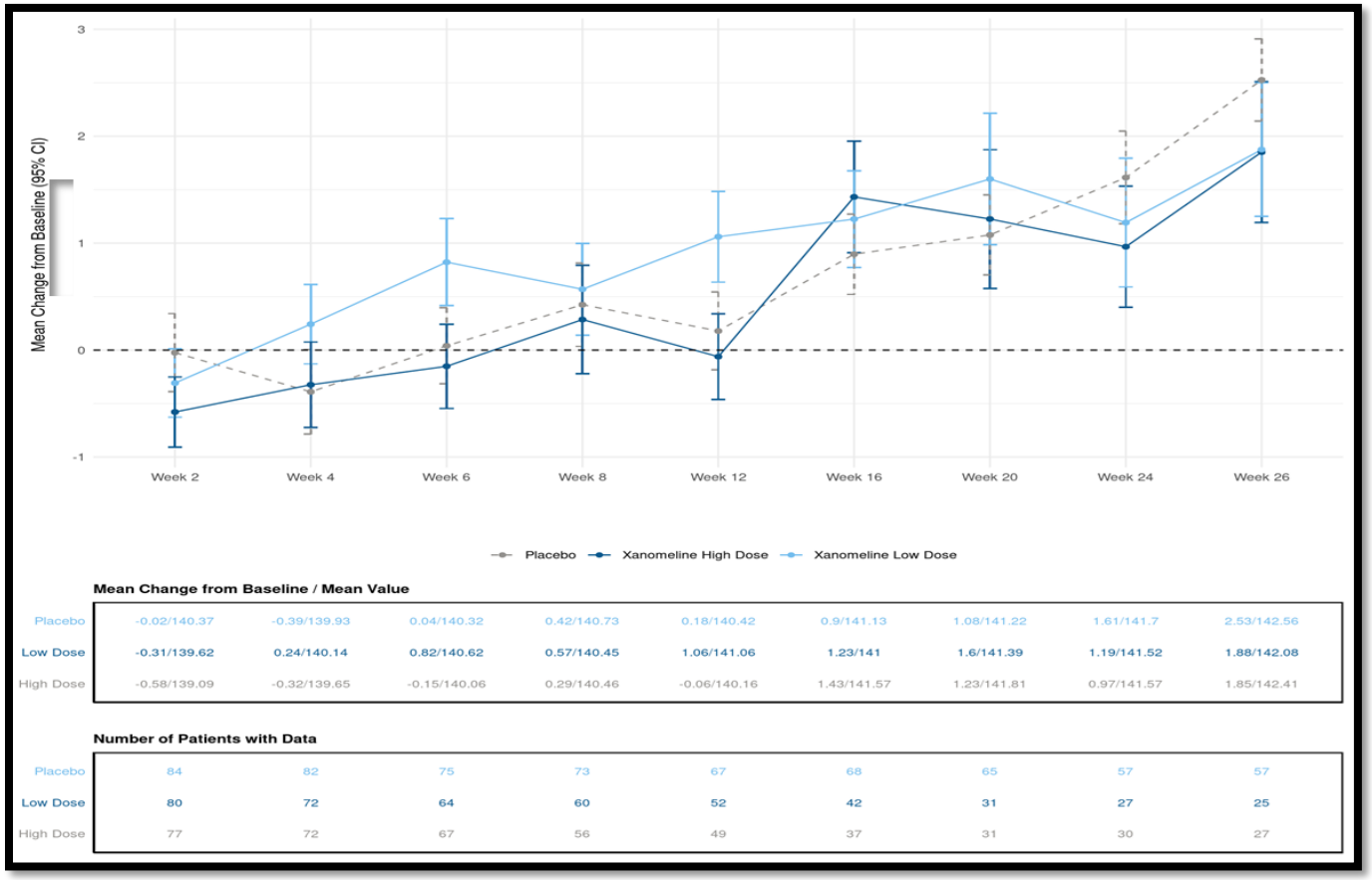
Treatment
PL
T1
T2

N = N = number of subjects with Max. baseline 'normal' or 'low' and at least one post-baseline value; n = among the 'N', number of subjects with Max. post-baseline 'high'; High and low are determined using the reference limits based on the individual subject's demographics.
*a – P values are from Fisher's Exact test, compared with PL

LLN=Lower limit of normal; ULN=Upper limit of normal; If the limits (LLN and ULN) vary across the population, the limits that apply to the majority of subjects are displayed in the scatter plot. Axes values are labeled in both SI unit and CN unit (in '[]'). The black diagonal line represents no change from max. baseline to max. post-baseline. The blue diagonal line represents an increase of xx units from max. baseline to max. post-baseline.

Recommend an interactive version of this plot.

Figure 3
Boxplots of for Liver Biochemical Tests
<Analysis Set>



Mean Change from Baseline / Mean Value

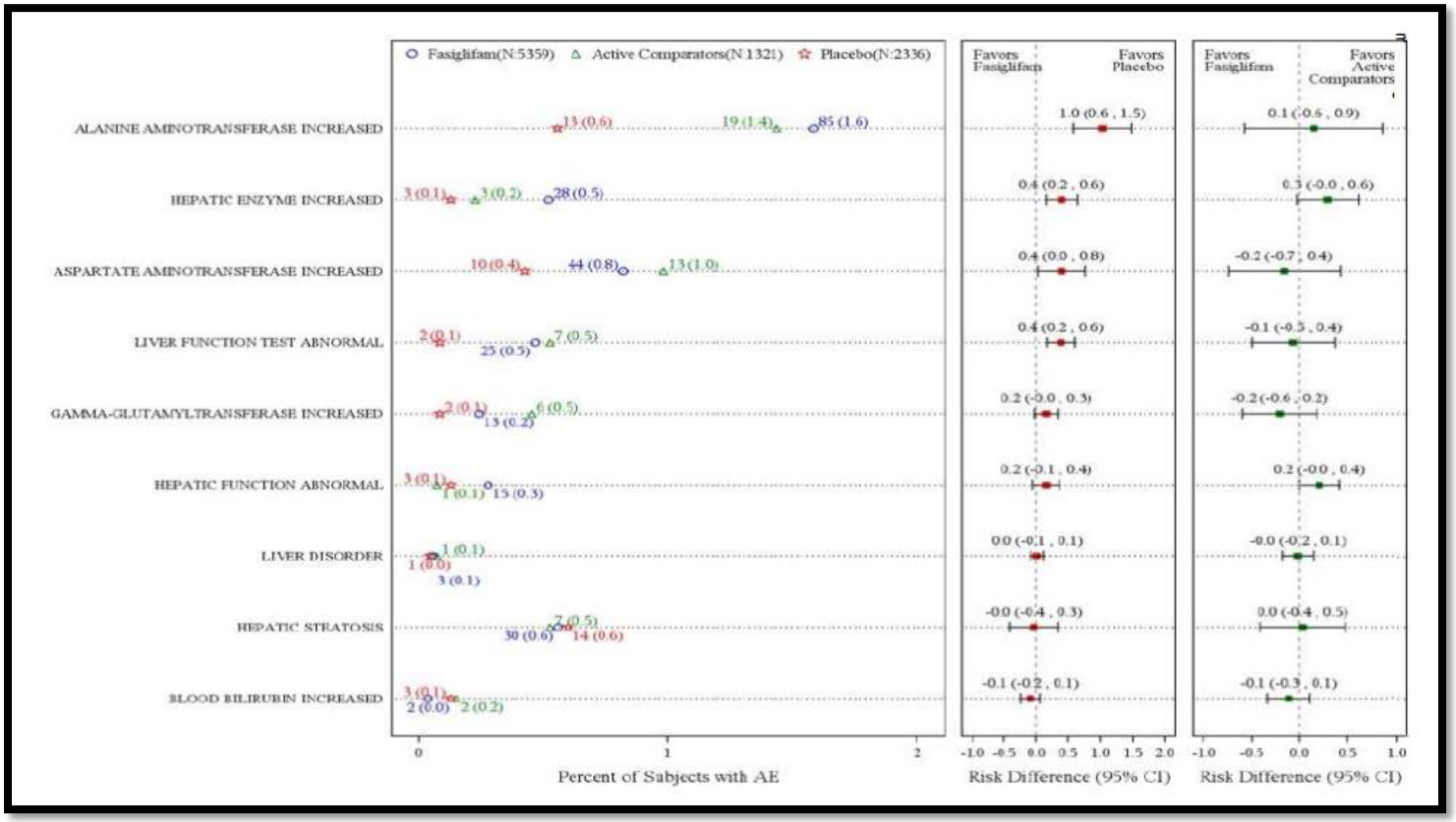
Placebo	-0.02/140.37	-0.39/139.93	0.04/140.32	0.42/140.73	0.18/140.42	0.9/141.13	1.08/141.22	1.61/141.7	2.53/142.56
Low Dose	-0.31/139.62	0.24/140.14	0.82/140.62	0.57/140.45	1.06/141.06	1.23/141	1.6/141.39	1.19/141.52	1.88/142.08
High Dose	-0.58/139.09	-0.32/139.65	-0.15/140.06	0.29/140.46	-0.06/140.16	1.43/141.57	1.23/141.81	0.97/141.57	1.85/142.41

Number of Patients with Data

Placebo	84	82	75	73	67	68	65	57	57
Low Dose	80	72	64	60	52	42	31	27	25
High Dose	77	72	67	56	49	37	31	30	27

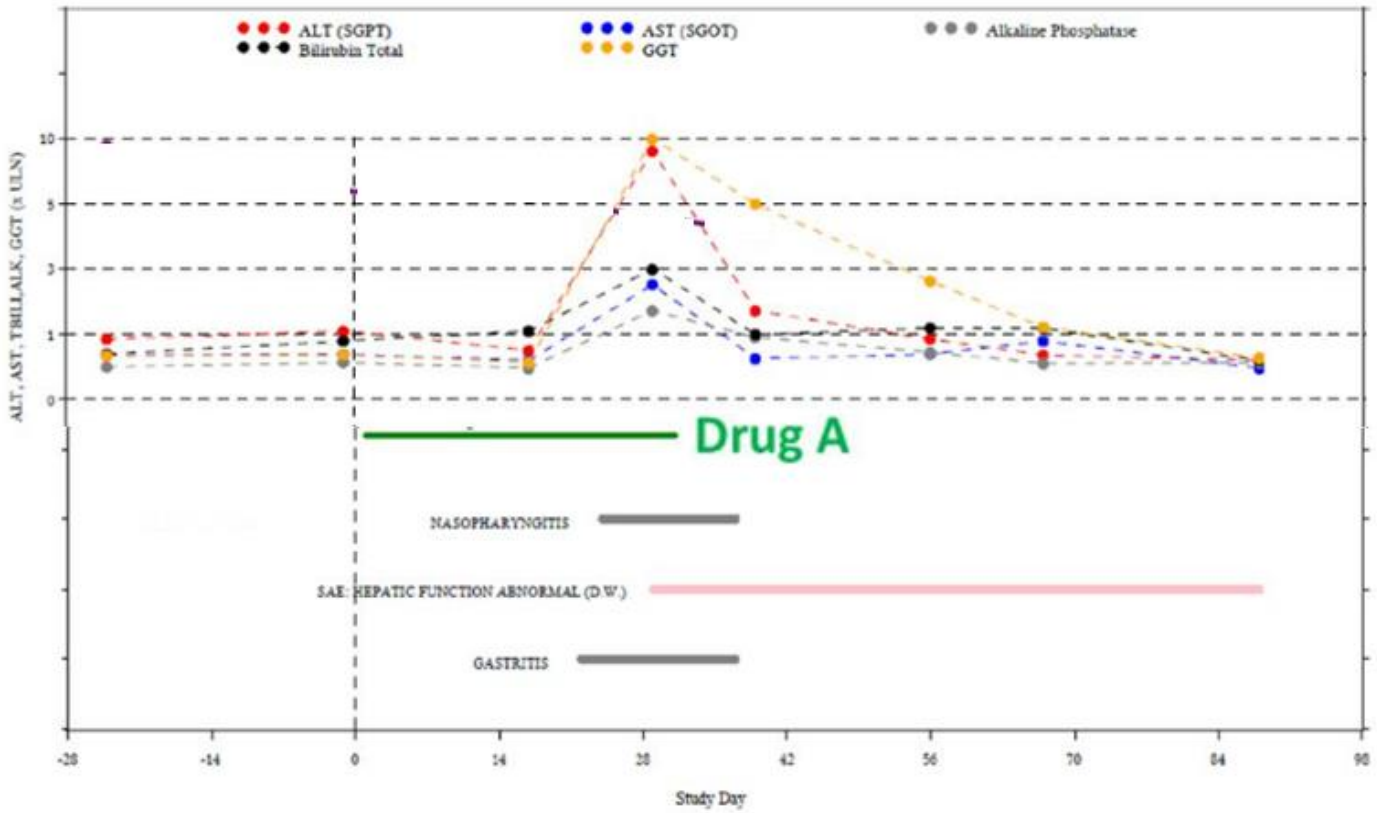
Recommend an interactive version of this plot. The CIs at each visit can be replaced by boxplots.

Figure 4
Forest Plots of DILI Adverse Events Based on Custom Query for DILI
<Analysis Set>



Recommend an interactive version of this plot.

Figure 5
Patient Profile of Interest
Categorical Increases in Hepatic Biochemistry and Enzymes
<Analysis Set>



Recommend an interactive version of this plot.

9 Acknowledgments:

The opinions expressed in this document are those of the authors and should not be construed to represent the opinions of PHUSE members, respective companies/organizations or Regulator's views or policies. The content in this document should not be interpreted as a data standard and/or information required by Regulatory Authorities.

The primary contributors include:

- Greg Ball, ASA Process Consulting
- Charles Beasley, Eli Lilly
- Paul "Skip" Hayashi, FDA
- John Landry, Eli Lilly
- John Marcinak, AbbVie
- Melvin Munsaka, AbbVie (Co-Lead)
- Mary Nilsson, Eli Lilly
- William Palo, AbbVie
- Swarnalatha Reddy, Arena Pharma
- Yang Veronica Pei, FDA
- Terry Walsh, Astellas (Co-Lead)

Thank you to the anonymous reviewers and apologies to any reviewer we may have missed. Thank you to the contributors and reviewers of the previous white papers.

10 Project Contact Information

Email: workinggroups@phuse.global

11 References

Avigan, M.I., Bjornsson, E.S., Pasanen, M. et al. Liver Safety Assessment: Required Data Elements and Best Practices for Data Collection and Standardization in Clinical Trials. Drug Saf 37 (Suppl 1), 19–31 (2014). https://doi.org/10.1007/s40264-014-0183-6.

CIOMS X: Drug-induced liver injury (DILI): Current status and future directions for drug development and the post-market setting: A consensus by a CIOMS Working Group Council for International Organizations of Medical Sciences (CIOMS). Geneva 2020, https://cioms.ch/wp-content/uploads/2020/06/CIOMS_DILI_Web_16Jun2020.pdf.

FDA Guidance for Industry Drug-Induced Liver Injury: Premarketing Clinical Evaluation, 2009. https://www.fda.gov/media/116737/download.

FDA Draft Guide on Advancing Premarket Safety Analytics Workshop Meeting Material Standard Safety Tables and Figures Integrated Guide, 2022. https://healthpolicy.duke.edu/events/advancing-premarket-safety-analytics.

FDA Reviewer Guidance Conducting a Clinical Safety Review of a New Product Application and Preparing a Report on the Review, 2005. https://permanent.fdlp.gov/LPS116705/fda/www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072275.pdf

Green, R. M., Flamm, S. AGA technical review on the evaluation of liver chemistry tests, Gastroenterology (2002), 123, 367-1384.

Health Canada: Guidance Document: Pre-market Evaluation of Hepatotoxicity in Health Products. https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/hepatotoxicity-pre-market-evaluation.html

Merz, M., Lee, K.R., Kullak-Ublick, G.A. et al. Methodology to Assess Clinical Liver Safety Data. Drug Safety, 37 (Suppl 1), 33–45 (2014). https://doi.org/10.1007/s40264-014-0184-5

Parks, D, Lin, X., Painter, J. L., Jie Cheng, Hunt, C. M, Spraggs, C. F. , Nelson, J. J., Curtis, L, Menius, J. A., and Lee, K. R. A proposed modification to Hy's law and EDISH criteria in oncology clinical trials using aggregated historical data. (2013), 22, 571-578. doi: 10.1002/pds.3405.

Pei, Y., Popat, V., Belowich, A. & Conley, C., (2023). Developing Technical Specifications for Submitting Clinical Trial Data Sets for Treatment of Noncirrhotic Nonalcoholic Steatohepatitis (NASH) Liver Fibrosis", Journal of the Society for Clinical Data Management 2(S1). doi: https://doi.org/10.47912/jscdm.209.

Reuben, A. (2004), Hy's Law, Hepatology, 39, 574-578.

Senior, J. R. Evolution of the Food and Drug Administration Approach to Liver Safety Assessment for New Drugs: Current Status and Challenges. Drug Safety, (2014), 37, 9–17.

Singh, S., Loke, Y.K. (2012) Drug safety assessment in clinical trials: methodological challenges and opportunities. Trials 13, 138 (2012). https://doi.org/10.1186/1745-6215-13-138

Temple, R. (2001), Hepatotoxicity Through the Years: Impact on the FDA, http://www.fda.gov/downloads/Drugs/ScienceResearch/ResearchAreas/ucm122149.pdf.

Zimmerman HJ. Drug-induced liver disease. Drugs. (1978), 16, 25-45. doi: 10.2165/00003495-197816010-00002. PMID: 352664

Zimmerman H. J. Hepatotoxicity. Dis Mon. (1993), 39, 675-787. PMID: 8223092.

Zimmerman, H. J. (1999). Hepatotoxicity: The Adverse Effects of Drugs and Other Chemicals on the Liver Subsequent Edition. Lippincott Williams & Wilkins.