Drug Saf (2013) 36:651–661
DOI 10.1007/s40264-013-0060-8

ORIGINAL RESEARCH ARTICLE

Assessment of Case Definitions for Identifying Acute Liver Injury in Large Observational Databases

Aaron J. Katz · Patrick B. Ryan · Judith A. Racoosin · Paul E. Stang

Published online: 14 May 2013

Abstract
Background Determining the aetiology of acute liver injury (ALI) may be challenging to both clinicians and researchers. Observational research is particularly useful in studying rare medical outcomes such as ALI; however, case definitions for ALI in previous observational studies lack consistency and sensitivity. ALI is a clinically important condition with various aetiologies, including drug exposure.
Objective The aim of this study was to evaluate four distinct case definitions for ALI across a diverse set of large observational databases, providing a better understanding of ALI prevalence and natural history.
Data Sources Seven healthcare databases: GE Healthcare, MarketScan® Lab Database, Humana Inc., Partners HealthCare System, Regenstrief Institute, SDI Health (now IMS Health, Inc.), and the National Patient Care Database of the Veterans Health Administration.
Methods We evaluated prevalence of ALI through the application of four distinct case definitions across seven observational healthcare databases. We described how laboratory and clinical characteristics of identified case populations varied across definitions and examined the prevalence of other hepatobiliary disorders among identified ALI cases that may decrease suspicion of drug-induced liver injury (DILI) in particular.
Results This study demonstrated that increasing the restrictiveness of the case definition resulted in fewer cases, but greater prevalence of ALI clinical features. Considerable heterogeneity in the frequency of laboratory testing and results observed among cases meeting the most restrictive definition suggests that the clinical features, monitoring patterns and suspicion of ALI are highly variable among patients.
Conclusions Creation of four distinct case definitions and application across a disparate set of observational databases resulted in significant variation in the prevalence of ALI. A greater understanding of the natural history of ALI through examination of electronic healthcare data can facilitate development of reliable and valid ALI case definitions that may enhance the ability to accurately identify associations between ALI and drug exposures. Considerable heterogeneity in laboratory values and frequency of laboratory testing among individuals meeting the criteria for ALI suggests that the evaluation of ALI is highly variable.

Electronic supplementary material The online version of this article (doi:10.1007/s40264-013-0060-8) contains supplementary material, which is available to authorized users.

A. J. Katz (✉)
UNC Eshelman School of Pharmacy, Division of Pharmaceutical Policy and Outcomes, University of North Carolina at Chapel Hill, CB# 7573, Chapel Hill, NC 27599, USA
e-mail: aj_katz@unc.edu

P. B. Ryan · P. E. Stang
Janssen Research and Development, Titusville, NJ, USA

P. B. Ryan · J. A. Racoosin · P. E. Stang
Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, Bethesda, MD, USA

J. A. Racoosin
Center for Drug Evaluation and Research, USFDA, Silver Spring, MD, USA

1 Introduction

Acute liver injury (ALI) related to drug exposure is an important complication of pharmacotherapy and it is

essential that it is detected as quickly as possible. However, diagnosis of ALI and identification of the aetiologic agent can be distinctively challenging, particularly when ALI is associated with medication use. For example, most cases of acute drug-induced liver injury (DILI) are idiosyncratic, with no association to the dose, duration or pharmacological action. High variability in patient presentation and the lack of standardized diagnostic criteria and confirmatory biomarkers further contribute to the difficulty in diagnosing DILI [1].

Observational studies using administrative claims data and electronic health records permit the assessment of associations between medical products and rare health outcomes such as DILI that may not have been observed in randomized clinical trials. However, there is little consensus on how to conduct retrospective analyses of observational databases with DILI as the outcome of interest. Previous retrospective studies of DILI have relied heavily on the presence of coded information from computerized patient records [2–7] and administrative databases [2, 8–21], but there is significant variation among these studies in regards to the algorithms and codes that were used to define and identify DILI cases. Furthermore, most retrospective database studies on DILI definitively excluded patients with diagnoses associated with chronic liver disease or more common aetiologies of ALI, such as viral hepatitis or alcohol, thus compromising the sensitivity of the algorithms used to identify DILI in favour of improving specificity. Although greater specificity minimizes the false-positive detection of non-cases and may improve the accuracy of associations identified between ALI and drug exposures of interest, poor sensitivity may result in the underestimation of true cases and a reduction in the statistical power needed to identify true associations.

Variation in the algorithms and data elements that have been used to define and identify DILI in previous studies, along with a limited assessment of the validity of the various algorithms, has hindered the ability of observational studies to describe DILI epidemiology and natural history in a consistent manner. Thus, identification of additional clinical and laboratory characteristics present around the time potential DILI cases are identified, as well as the development of a standard definition of DILI, are essential to the validity and reproducibility of case identification in observational data.

In an effort to inform the development of a postmarket risk identification and analysis system by the US FDA Sentinel Initiative, the Observational Medical Outcomes Partnership (OMOP) [22] has set out to evaluate and describe the identification of important health outcomes through the application of distinct outcome definitions across a disparate set of observational healthcare databases. However, although there is additional value in determining the legitimacy of outcome definitions utilized in observational research, many of the OMOP databases do not have the capability to access source medical records and, thus, estimating the validity of the outcome definitions in the present study was not feasible. As such, the aims of our study are to assess the use of multiple definitions of ALI across several observational healthcare databases and to better understand how automated healthcare data alone could enhance the understanding of the clinical presentation of ALI. Specifically, the objectives of this study are to (i) describe the prevalence of ALI across multiple observational databases using distinct case definitions of ALI; (ii) compare the demographic and clinical characteristics of each distinct group of cases across the subset of observational databases that contain a full range of clinical and laboratory data; (iii) examine the prevalence of other hepatobiliary conditions that may decrease suspicion that an identified case of ALI was possibly related to drug exposure.

2 Methods

Four distinct case definitions of ALI (Fig. 1) were created, taking into consideration the diagnosis codes, procedure codes and laboratory result parameters used to identify ALI in previous observational studies [23]. International Classification of Diseases 9th edition, Clinical Modification (ICD-9-CM) diagnosis codes as well as ICD-9-CM, Current Procedural Terminology (CPT®) and Logical Observation Identifiers Names and Codes (LOINC®) procedure codes indicative of ALI (see Appendix table SI [Online Resource 1]) formed the criteria of the diagnosis (D) and diagnostic or therapeutic procedure (P) components of case definitions, respectively. Following the FDA guidance document for pre-clinical evaluation of DILI [24], laboratory report of serum alanine aminotransferase (ALT) or aspartate aminotransferase (AST) activities threefold or greater than the upper limit of normal (ULN), and serum concentration of total bilirubin (TBL) greater than 2-fold the ULN constituted the criteria for the laboratory (L) component of case definitions.

Seven observational healthcare databases (Table 1) were available within the OMOP network at the time of the study, each of which was standardized to the OMOP common data model (http://omop.finh.org/ETLProcess). These seven observational databases vary in regards to the patient population represented as well as the types, sources and consistency of data captured. For example, the Humana database contains administrative claims from a wide range of health plans administered under a single private insurer and allows for identification of ALI in both the inpatient and outpatient settings among a diverse population of beneficiaries. However, the Humana database, along with the National Patient Care Database (NPCD) of the

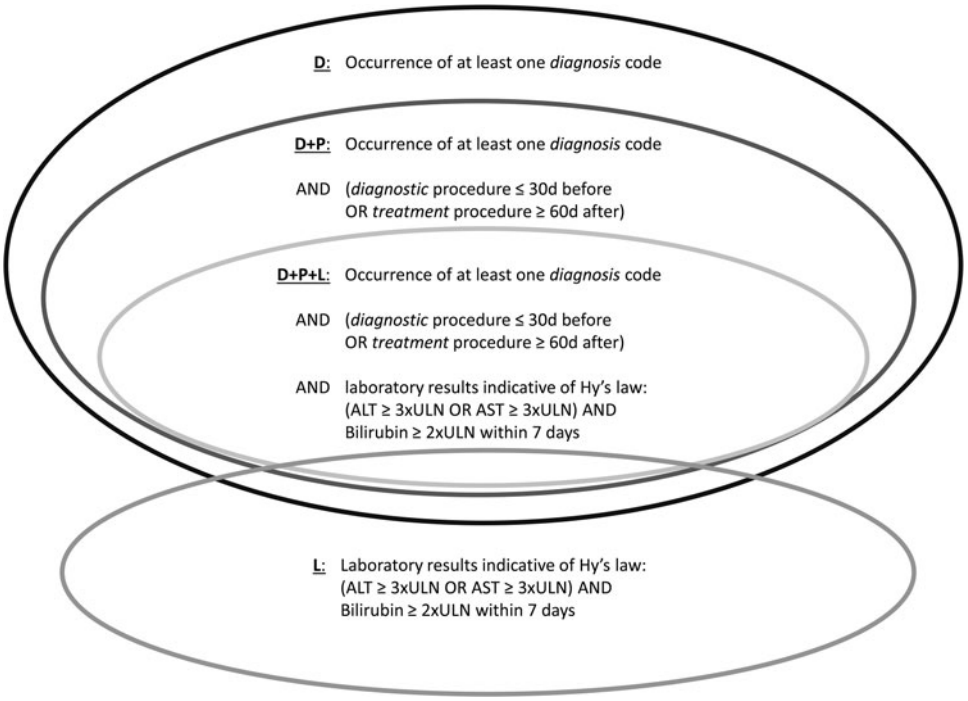

Fig. 1 Definitions of acute liver injury (ALI). The figure shows how the four distinct case definitions of ALI were defined. The broadest definition, D, only required the occurrence of at least one diagnosis code. Definition L required elevated laboratory results, including ALT or AST levels ≥3 times the ULN and bilirubin ≥2 times the ULN. Definition D+P required the occurrence of at least one diagnosis code as well as the occurrence of either a diagnostic procedure code ≤30 days or a treatment procedure code ≥60 days from the day of the diagnosis code. The most restrictive definition, D+P+L, added the requirement of elevated laboratory results to the D+P definition. *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *D* diagnosis, *L* laboratory, *P* diagnostic or therapeutic procedure, *ULN* upper limit of normal

Veterans Health Administration and the SDI Health (now IMS Health, Inc.) database, lacks laboratory result information. In consequence, identification of ALI within these three databases can only be done through the detection of D and P codes. Conversely, the GE Healthcare (GE), Regenstrief Institute (RI), Partners HealthCare System (PHS) and MarketScan® Lab (MSLR) databases all contain information on laboratory results, including measures of ALT or AST and TBL expressed in multiples of the ULN. However, because the GE database is composed of electronic medical record information reflecting clinical events documented in the ambulatory care setting, the capture of inpatient experiences, particularly inpatient procedures, is generally incomplete. Thus, since many diagnostic and therapeutic procedures pertaining to ALI are performed in an inpatient setting, the GE database may not be considered ideal for the evaluation of ALI when using case definitions that include these procedures as a necessary criterion. As a result, given the varying capabilities of the disparate data sources to identify cases under each of the operational definitions of ALI, we applied definitions D and D+P across all seven databases, restricted application of definition L to the four databases with laboratory information (GE, RI, PHS and MSLR), and limited application of definition D+P+L to the three databases that more completely capture diagnosis, procedure and laboratory result information (RI, PHS and MSLR). The prevalence of ALI for each given combination of definition and observational database was then calculated as the number of individuals with at least one occurrence of ALI divided by the total population within the database.

Once ALI cases were identified for a specified case definition, an index date was established for each patient, representing the date on which the case definition criteria were met. Patients identified as cases under more than one definition had separate index dates established for each definition of ALI under which they met case criteria. The index date was created to represent a timepoint when ALI was detected in a given individual under specified definition criteria of ALI. From this reference point, we then looked at the clinical events occurring in the time period prior to ALI detection in order to evaluate and characterize the prevalence of clinical observations (diagnoses, signs and symptoms denoted by ICD-9-CM diagnosis codes) and utilization of other diagnostic activities (procedures and tests denoted by ICD-9-CM procedure codes) that took place during ALI work-up and prior to case detection. In addition, the prevalence of various hepatobiliary diseases

Table 1 Description of seven observational healthcare databases within the Observational Medical Outcomes Partnership research network

Database	Description	Population	Number of records (millions)			
			Drugs	Conditions	Procedures	Observations
GE Healthcare	Derived from data pooled by providers using GE Centricity Office (an ambulatory electronic health record) into a data warehouse in an HIPAA-compliant manner	Total: 11.2 million % Male: 42 Mean age (SD): 39.6 (22.0)	182.6	66.1	110.6	1,121.1
MarketScan® Lab Database	Represents privately insured population that has at least one recorded laboratory value, with administrative claims from inpatient, outpatient and pharmacy services supplemented by laboratory results	Total: 1.2 million % Male: 35 Mean age (SD): 37.6 (17.7)	37.6	49.5	68.5	41.8
Humana, Inc.	Contains medical (inpatient, outpatient and emergency room), pharmacy and laboratory data (including test results) from Humana's administrative claims database of medical members	Total: 5.2 million % Male: 48 Mean age (SD): 48 (24.3)	208.8	323.7	307.3	498.4
Partners HealthCare System	Includes data from the Partners HealthCare System clinical transaction-based data repository as well as inpatient and outpatient billing feeds collected in the Research Patient Data Registry, an analytic-structured database	Total: 2.9 million % Male: 45 Mean age (SD): 37.8 (22.1)	100.1	123.6	179.2	828.6
Regenstrief Institute/INPC	Includes healthcare data from the INPC containing population-based, longitudinal and structured coded and text data captured from hospitals, physician practices, public health departments, laboratories, radiology centres, pharmacies, pharmacy benefit managers and payers	Total: 2 million % Male: 48 Mean age (SD): 32.1 (23.1)	98	170.7	219.7	440.8
SDI Health (now IMS Health, Inc.)	Contains HIPAA-compliant, de-identified and encrypted patient-level data from hospitals, clinics, physician offices, and retail and specialty pharmacies from all 50 US states	Total: 90.5 million % Male: 41 Mean age (SD): 42 (23.5)	3,248.2	2,048.5	2,108	0
National Patient Care Database of the Veterans Health Administration	Includes data for US Veterans from prescription dispensing records and electronic health records across facilities within the VA healthcare system, as part of the VA Center for Medication Safety (VA MedSAFE)	Total: 3.2 million % Male: 96 Mean age (SD): 65.9 (12)	421.3	358.9	506.6	54

HIPAA Health Insurance Portability and Accountability Act, *GE* GE Healthcare, *INPC* Indiana Network for Patient Care, *SD* standard deviation, *VA* Veterans' Affairs

that may decrease the likelihood that an ALI case was drug-induced was also evaluated for each distinct case definition through recognition of diagnosis codes that occurred prior to the index dates of cases.

Individual electronic patient profiles of cases meeting D+P+L definition criteria in the MSLR database were reviewed for potential trends in the frequency of laboratory monitoring and laboratory test results, both prior to and following the index date of cases. ALT, AST and TBL values for each identified case were plotted longitudinally as standardized values by dividing observed values by the ULN.

Descriptive statistics, including counts with percentages and means with standard deviations, were calculated to characterize patient demographics (age, sex) and to explore prior conditions and comorbidities, patterns in health service utilization (visits and procedures) and trends in laboratory values among patients satisfying the four distinct definitions of ALI.

3 Results

Table 2 highlights the number of patients identified as ALI cases according to the 'ALI' definition and observational database utilized. The prevalence of ALI under the broadest definition, D, ranged from 4.6 to 12.6 % across the seven databases. Imposing the additional requirement of a diagnostic or treatment procedure reduced the proportion of potential cases to a range of 0.1–0.4 % in six of the databases; the D+P definition identified only 75 cases in the GE database, most likely due to its reduced capture of procedures performed outside of the ambulatory care setting, such as liver biopsy. Among the four databases that contained the laboratory results necessary to identify cases under definition L criteria, the prevalence of ALI ranged from 0.02 to 1.2 %. The three databases able to identify cases satisfying the most restrictive definition of ALI, D+P+L, detected 28 (0.002 %), 384 (0.02 %) and 2,568 (0.09 %) cases in MSLR, RI and PHS, respectively.

In the MSLR, RI and PHS databases, the sex distribution of potential cases shifted substantially as the case definition of ALI became more restrictive. Under the D definition of ALI, females predominated, representing 53 % of cases in PHS to 58 % of cases in MSLR. Under the D+P+L definition, males predominated, representing 61 % of cases in each of the three databases. There were also notable differences in the age distribution across definitions. The D cohort had a lower average age and larger variability than

Table 2 The prevalence and demographic characteristics of patients satisfying four distinct definitions of acute liver injury, by database

Database	Total population [n (000s)]	Patients satisfying acute liver injury definitions			
		D [n (%)]	L [n (%)]	D+P [n (%)]	D+P+L [n (%)]
MarketScan® Lab Database	1,229	154,357 (12.6)	191 (0.02)	3,743 (0.3)	28 (0.002)
Regenstrief Institute	2,002	215,930 (10.8)	7,349 (0.4)	2,735 (0.1)	384 (0.02)
Partners HealthCare System	2,942	264,083 (9.0)	36,267 (1.2)	11,032 (0.4)	2,568 (0.09)
National Patient Care Database of the Veterans Health Administration	3,202	345,519 (10.8)	NA	6,453 (0.2)	NA
Humana Inc.	5,197	447,886 (8.6)	NA	8,972 (0.2)	NA
GE Healthcare	11,216	514,118 (4.6)	6,461 (0.1)	75 (0.001)	NA
SDI Health (now IMS Health, Inc.)	90,485	6,491,416 (7.2)	NA	67,102 (0.1)	NA
Sex (% female)					
MarketScan® Lab Database	NA	57.8	48.7	52.6	39.3
Regenstrief Institute	NA	54.8	50.5	51.7	39.3
Partners HealthCare System	NA	53.2	42.8	47.7	38.8
Age at index date; mean (SD); [25–75 %]					
MarketScan® Lab Database	NA	32.6 (20.7) [12–50]	45.6 (14.6) [39–55]	48.3 (11.3) [42–56]	50.7 (15.1) [45–58]
Regenstrief Institute	NA	25.1 (26.3) [3–45]	46.5 (31.4) [32–62]	47.5 (14.4) [40–56]	45.9 (17.7) [40–56]
Partners HealthCare System	NA	38.6 (21.7) [24–55]	51.9 (18.9) [39–67]	47.1 (16.0) [36–58]	48.6 (16.9) [39–60]

ALT alanine aminotransferase, *AST* aspartate aminotransferase, *D* occurrence of at least one diagnosis code, *D+P* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code; *L* ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *D+P+L* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code AND ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *NA* not applicable, *ULN* upper limit of normal

Table 3 The prevalence of signs, symptoms, laboratory tests and diagnostic procedures present during acute liver injury work-up prior to case index dates, according to defined cohorts and observational database (%)

Signs, symptoms, laboratory tests, and diagnostic procedures	MarketScan® Lab Database				Regenstrief Institute				Partners HealthCare System			
	D	L	D+P	D+P+L	D	L	D+P	D+P+L	D	L	D+P	D+P+L
Signs and symptoms												
Abdominal pain	14	32	30	43	17	28	38	48	15	16	25	33
Jaundice	0	12	3	29	1	4	4	21	1	2	4	15
Liver enzyme tests abnormal	2	9	19	25	2	6	26	29	3	6	20	26
Malaise and fatigue	8	12	14	14	9	13	24	23	11	8	16	15
Nausea and vomiting	2	5	5	4	6	12	13	20	5	7	9	14
Laboratory tests and diagnostic procedures												
Hepatic function panel	10	34	38	68	5	8	31	22	14	19	39	40
Bilirubin; direct	9	56	24	61	3	6	25	40	11	17	37	40
ALT	4	10	11	14	8	13	42	53	19	21	48	45
AST	4	9	10	14	7	12	40	53	23	25	54	51
Alkaline phosphatase	3	9	8	11	5	9	33	49	20	24	53	51
Prothrombin time	9	38	61	79	4	10	41	51	23	39	79	76
Complete, real-time abdominal ultrasound with image documentation	4	27	33	54	1	4	13	14	5	11	37	37
Percutaneous liver biopsy	0	5	4	0	0	1	4	8	0	2	5	8

ALT alanine transaminase, *AST* aspartate transaminase, *D* occurrence of at least one diagnosis code, *D+P* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code, *L* ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *D+P+L* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code AND ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *ULN* upper limit of normal

the other cohorts, in part because the D cohort contains a sizeable proportion of patients <18 years of age. Conversely, the D+P+L cohort is older on average compared with the other definition cohorts, with at least 75 % of the population aged 39 years or older in all three databases.

Table 3 shows the proportion of patients within the identified cohorts who had diagnosis or procedure codes for signs and symptoms, laboratory tests and procedures expected to be observed during the work-up of an ALI case. Within all three databases, jaundice was recorded for <1 % of patients who met definition D, but was seen in 15–29 % of patients in the D+P+L cohort. The prevalence of abdominal pain also increased substantially across definition cohorts as the restrictiveness of the definition of ALI increased. Within MSLR, 10 % of patients in the D cohort had a hepatic function panel recorded prior to the case index date. This proportion increased to 38 % in the D+P cohort, and to 68 % in the D+P+L cohort. Within the MSLR database, tests of prothrombin time were observed prior to the index date in 9, 38, 61 and 79 % of the D, L, D+P and D+P+L cohorts, respectively. Similar trends in the occurrence of hepatic function panels and tests of prothrombin time were observed in the PHS and RI databases. Ultrasound of the liver was recorded in fewer than 10 % of patients in all databases who satisfied definition D, but was more commonly observed among patients in the more restrictive definition cohorts. Liver biopsy was recorded in <8 % of patients across all definitions and all databases.

Table 4 describes the prevalence of pre-existing hepatobiliary disorders that may reduce suspicion of medication use as the aetiologic factor of ALI. In general, the prevalence of pre-existing hepatobiliary disorders increased within all three databases as the restrictiveness of the ALI definition increased. In the D cohort, the prevalence of most pre-existing hepatobiliary disorders was less than 1 %. Conversely, some pre-existing hepatobiliary conditions were present in 20 % or more of the cases meeting D+P+L definition criteria. For example, within the RI database, 20 % of patients meeting D+P+L criteria had a prior diagnosis record of acute hepatitis C, 25 % had a prior record of viral hepatitis C and 32 % had a diagnosis code for chronic hepatitis C. Cirrhosis of the liver (without mention of alcohol) was identified in 25 % of MSLR patients, 26 % of PHS patients and 43 % of RI patients. Across all three databases, the prevalence of a prior code for primary malignant neoplasm of the liver was 10–11 %. Evidence of prior liver transplantation (e.g. a code for 'complication of transplanted liver') was highly variable and observed most often among suspected cases within the RI database (15 %). Patient profile review of the 28 qualifying patients in the MSLR cohort revealed that all 28

Table 4 The prevalence of additional hepatobiliary conditions prior to case index dates that may decrease the likelihood an acute liver injury is drug-induced, according to defined cohorts and observational database (%)

Condition	MarketScan® Lab Database				Regenstrief Institute				Partners HealthCare System			
	D	L	D+P	D+P+L	D	L	D+P	D+P+L	D	L	D+P	D+P+L
Total cohort size (n)	154,357	191	3,743	28	215,930	7,349	2,735	384	264,083	36,267	11,032	2,568
Alcoholic cirrhosis	0	5	3	4	0	3	4	17	0	2	3	10
Alcoholic liver damage	0	0	1	0	0	1	1	3	0	0	1	2
Acute hepatitis C	0	5	28	18	0	3	17	20	0	2	17	15
Viral hepatitis C	0	0	8	14	0	5	18	25	0	1	4	4
Chronic hepatitis C	0	7	31	25	0	4	31	32	0	3	16	16
Chronic hepatitis	0	3	7	7	0	2	7	16	0	1	5	7
Cirrhosis of liver without mention of alcohol	0	10	8	25	0	6	12	43	0	5	11	26
Cholelithiasis AND cholecystitis without obstruction	1	5	3	7	1	4	4	6	1	2	2	3
Calculus of gallbladder without mention of cholecystitis, without mention of obstruction	1	10	8	21	1	9	12	20	2	6	11	13
Disorder of biliary tract	0	6	2	11	0	4	5	18	0	3	4	12
Chronic non-alcoholic liver disease	0	6	9	7	0	7	19	41	1	3	15	16
Complication of transplanted liver	0	0	1	4	0	1	2	9	0	0	1	2
History of liver transplant	0	0	3	0	0	1	4	15	0	0	1	5
Primary malignant neoplasm of liver	0	4	2	11	0	1	2	11	0	2	4	10
Secondary malignant neoplasm of liver	0	0	2	4	0	2	2	2	0	6	9	12

ALT alanine transaminase, *AST* aspartate transaminase, *D* occurrence of at least one diagnosis code, *D+P* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code, *L* ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *D+P+L* occurrence of at least one diagnosis code AND either a diagnostic procedure code ≤30 days before diagnosis code or a treatment procedure code ≥60 days after diagnosis code AND ALT or AST ≥3 × ULN + total bilirubin ≥2 × ULN within 7 days, *ULN* upper limit of normal

cases had at least one of these pre-existing hepatobiliary disorders listed in their electronic health record prior to the index date, decreasing the likelihood that these cases of ALI were exclusively due to drug exposure.

The longitudinal patterns in liver-related laboratory tests were explored for the 28 patients in the MSLR D+P+L cohort (Fig. 2). A representative subset of the plots were selected to display the substantial heterogeneity in the density and duration of data capture, as well as the magnitude of observed elevations and difference from prior measures. Patient A had a series of laboratory tests occurring before, on and after their index date in which aminotransferase and bilirubin levels exceeded 10 times the ULN. Patient B had multiple laboratory evaluations throughout the period of interest that started with an abrupt elevation of all laboratory values between the time just prior to and on the index date followed by a gradual reduction in values toward the ULN. In contrast, patient F had a small number of laboratory tests recorded, each indicative of elevated values, but without any data to inform baseline or follow-up periods. Patient C had data available to display dramatic increases between the laboratory values taken prior to the index date and those observed on or just following the index date; however, the laboratory results that followed the index date were relatively short in duration and did not show evidence of resolution. Patients D and E displayed elevated laboratory values on the index date that appeared to subside thereafter, but no baseline information was available to assess whether or for how long laboratory values were elevated prior to the index date. Longitudinal laboratory test patterns of patients not included in the subset generally followed one of the six configurations in the figure.

4 Discussion

Through the evaluation of four distinct case definitions for ALI across a diverse set of large observational databases, the OMOP study provides an improved understanding of ALI prevalence and natural history. As would be expected, we observed that increasing the restrictiveness of a case definition for ALI resulted in a reduction in the number of cases within each of the databases studied. Under the broadest definition, D, the prevalence of ALI cases was as high as 12.6 %. When the requirement for a diagnostic or treatment procedure was added to the case definition (D+P), the prevalence of cases was no more than 0.4 %.

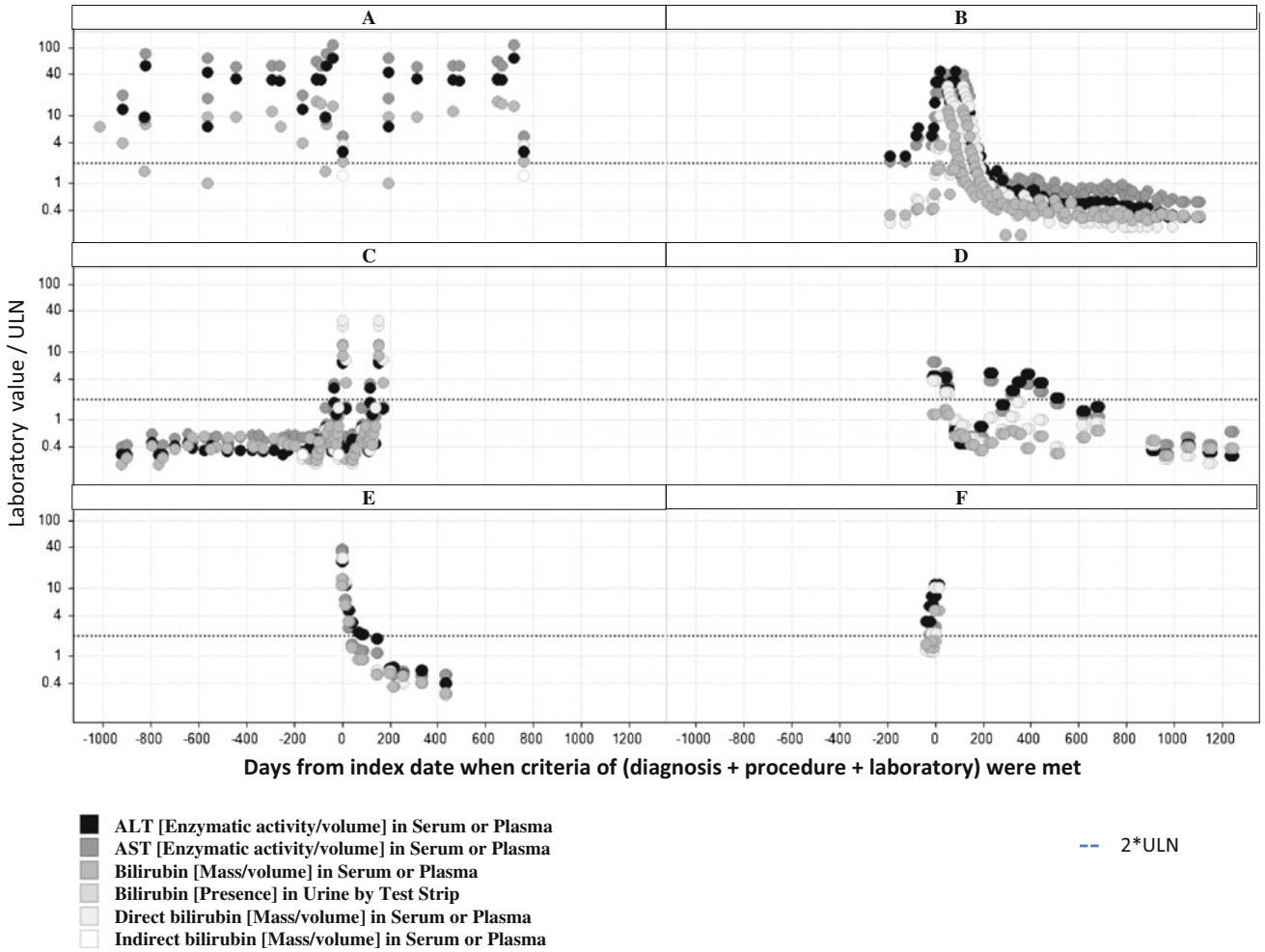

Fig. 2 Longitudinal patterns in liver-related laboratory values. The *trellis plot* shows a sample of six individual patients' laboratory values over time. Days are plotted on the x-axis, with day 0 representing the index date, i.e. the date on which the patient met case definition criteria (diagnosis code, diagnostic or treatment procedure code, and elevated laboratory results, including ALT or AST levels ≥3 times the ULN and bilirubin ≥2 times the ULN. Laboratory values standardized as the observation value divided by the ULN are plotted on the y-axis on the log scale, with 2 times the ULN (2*ULN) indicated by the *dashed horizontal line*. *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *ULN* upper limit of normal

Adding the requirement of elevated serum aminotransferase and bilirubin laboratory values (D+P+L), further reduced the prevalence of cases to less than or equal to 0.1 % in each of the three databases that had diagnostic, procedure and laboratory test information. Based on these results, it appears that using diagnosis codes only to define ALI may result in a significant amount of "noise" around the prevalence estimate, particularly if the diagnosis codes are used more often to rule out hepatobiliary conditions than they are to represent a bona fide case of ALI. However, whether the use of the D+P definition or the D+P+L definition should be promoted over the other is not as obvious. We recognize there is a limit to what we can learn looking at automated healthcare data only. Given that several OMOP databases do not have the capability of accessing source medical records, validation of the case definitions of ALI was not feasible in the current study. However, it will be useful to validate these case definitions in future research to determine whether the reduction in the number of cases is a result of changes in the sensitivity and/or specificity of the algorithms used to define cases. Variation in the sensitivity and specificity of the algorithms could have important implications for future research of ALI using observational databases; in particular, changes in these performance measures could alter the accuracy and/or precision of estimates made in studies evaluating associations between ALI and medications, and thereby affect our understanding about the safety of those drugs.

A notable study finding was the substantial shift in the distributions of sex and age of ALI cases as the case definition became more restrictive and more specific to identifying DILI. Under the broadest ALI definition (D), females

consistently represented over half of the cases and the average age was less than 40 years in each of the three databases. However, with the additional requirement of procedures and abnormal laboratory values in the case definition, ALI cases were predominantly male and the average age was greater than 45 years. This finding is inconsistent with a study from the DILI Network where females represented 60 % of 300 suspected cases of DILI [25]. However, it is important to highlight that patients identified in this study likely represent a range of causes of ALI and not just those that are drug-induced. In addition, the shift in sex distribution among cases in the most restrictive definition likely reflects the increased prevalence of certain clinically severe hepatobiliary disorders in men [26–28]. The older mean age and narrower age range of the more restrictive definition cohorts may reflect differences in clinical practice patterns as they relate to shifts in the suspicion of ALI and other hepatobiliary disorders in older patients.

Another key objective of the study was to characterize the clinical presentation and work-up of ALI cases in each of the defined cohorts. We found that increasing the restrictiveness of our case definition results in an increase in the prevalence of clinical features consistent with ALI that are present prior to the index dates of identified cases. Abdominal pain was the predominant symptom present among suspect cases in each of the four definitions and across each of three observational databases, MSLR, RI and PHS. In addition, the presence of jaundice increased considerably across the three databases as the restrictiveness in the definition of ALI also increased; jaundice was nearly non-existent among patients meeting criteria for definition D, whereas it was present in over 25 % of patients meeting the most restrictive definition, D+P+L.

We found that the diagnostic work-up of suspected cases also varied according to case definition and observational database. For example, compared with cases under definitions D or L in each of the databases, cases identified under definition D+P or definition D+P+L were more likely to have received a 'hepatic function panel' or aminotransferase or bilirubin tests prior to their index dates as cases. The frequency of the prothrombin time test also varied with the restrictiveness of the case definition, and was particularly common in the definition cohorts requiring procedure codes (D+P and D+P+L), reflecting either the likelihood that bleeding risk was assessed prior to a procedure or as a marker of hepatic dysfunction. One unexpected finding was the relatively low prevalence of diagnostic procedures, including laboratory tests, among each of the identified cohorts. Although, as noted, the prevalence of these procedures increased with the restrictiveness of the case definition, it was somewhat surprising that only 25 % or fewer cases identified with definition D received laboratory testing or had laboratory testing recorded prior to meeting definition criteria. It is conceivable that laboratory tests would be performed following diagnosis in order to rule out certain conditions; in this study we did not evaluate the extent to which these tests occurred among cases *after* definition criteria had been met.

When cases were evaluated for evidence of hepatobiliary diseases that may decrease the likelihood that an identified case of ALI was drug-induced, we found that the prevalence of these hepatobiliary diagnoses varied by ALI definition, disease and occasionally by database. For instance, among cases meeting D+P+L criteria, diagnoses of hepatitis C infection were commonly reported, whereas the code for 'alcoholic liver damage' was rarely observed. Over 40 % of cases meeting D+P+L criteria in the RI database had the code for 'cirrhosis of the liver without mention of alcohol' compared with approximately 25 % of D+P+L cases in both the MSLR and PHS databases. This finding highlights the usefulness of utilizing multiple observational databases when assessing the epidemiology and natural history of health outcomes such as ALI, as each database may represent different populations of potential cases and/or alternative methods of data collection, coding and presentation. Furthermore, as with our evaluation of laboratory tests and procedures, we did not ascertain whether other hepatobiliary conditions were present among identified cases after definition criteria were met. It is possible that, for some cases, further clinical evaluation resulted in the recognition and diagnosis of other hepatobiliary conditions soon after the index date. Thus, the percentage of identified ALI cases with evidence of other hepatobiliary diseases that may reduce suspicion of drug-induced injury could be underestimated in this study.

The longitudinal presentation of liver enzyme and bilirubin tests for cases meeting definition D+P+L in the MSLR database is a unique aspect of this study. Interestingly, although results for each of the 28 patients generally followed one of the six configurations represented in Fig. 2, we found considerable variation among suspect cases with respect to laboratory value patterns and frequency of testing around the index date on which the patients met case definition criteria. Some suspect cases only had laboratory test results on the day of and immediately before and/or after the index date, while other suspect cases had numerous laboratory tests reported over a considerable duration of time either before or after meeting case definition criteria. Additionally, the laboratory results of suspect cases typically followed one of the three following longitudinal patterns: (i) normal or close to normal baseline values prior to a sudden elevation on the index date;(ii) no baseline values followed by elevated values on the index date and then a decline toward normal values thereafter; or (iii) elevated values prior to the index date with values remaining elevated following the index date.

A possible explanation for the variation in the laboratory value patterns and frequency of laboratory testing is that the majority of test results contained in the MSLR database are from standalone outpatient laboratories; attainment of diagnostic testing results from laboratories within inpatient facilities is likely inconsistent at best. Thus, the capture of laboratory tests relevant to the identification of ALI may be incomplete and, as a result, it is possible that our assessments of ALI using definitions L and D+P+L resulted in underestimations of the prevalence of ALI within the MSLR database. While we could not conduct chart reviews to confirm diagnosis of ALI among the D+P+L cases identified in the MSLR database, variation in the longitudinal patterns of laboratory values among these cases suggests that evaluation of liver enzyme and bilirubin test history may have significant utility in discerning bona fide cases of ALI in observational studies when combined with diagnosis and procedure data. The sparse pattern of laboratory tests observed in some D+P+L patients is a reminder that data quality in large observational databases must always be evaluated, and unusual patterns of laboratory findings must be corroborated by other data sources (e.g. diagnosis or procedure codes) whenever possible. A key limitation of this study is the lack of ability to conduct source record verification. Further exploration and chart review may provide additional insights about the patterns observed and the clinical accuracy of the cases identified by these definitions.

In addition, understanding the utility of laboratory results in observational database studies of ALI requires appropriate interpretation and acknowledgement of the limitations of the laboratory measures. For example, an elevated measure of the serum activities of ALT may indicate hepatocellular injury; however, ALT is not exclusive to the liver and an elevated measure of ALT is not specific enough to determine the severity or aetiology of an ALI. An elevated measure of serum TBL is highly specific for liver complications and a useful indicator of decreased liver function, but increased serum TBL is an insensitive marker of liver injury because it often occurs late in the disease process. When used together, the combination of elevated serum ALT and TBL measures provides a sensitive and specific marker of liver injury but does not differentiate aetiology. Other potential biomarkers of liver complications exist [29], but their value in detecting and distinguishing specific forms of ALI relative to the combination of serum aminotransferase and TBL measures has not been established. Including these additional tests in laboratory-based definitions of ALI may provide greater sensitivity, but also create more variance in the types of liver injury identified; this is an important consideration if the eventual goal is to detect specific causes of liver injury, such as drug exposure.

5 Conclusions

ALI is a health condition with multiple aetiologies that presents a significant challenge to both clinicians and researchers. The lack of standardized and well-accepted diagnostic criteria and variation of case definitions across retrospective studies makes consistent identification of true ALI cases in observational studies extremely difficult and thus contributes to inconsistencies in epidemiologic measures of the condition. To better understand this heterogeneity, we developed several case definitions of ALI using different data elements and then assessed the prevalence of these distinct case definitions across multiple observational databases. In addition, to explore the capability of automated healthcare data to inform the understanding of the natural history of ALI, we characterized cohorts of identified cases with regard to clinical presentation and work-up. Finally, we displayed the potential utility of longitudinal laboratory data in observational studies to discern true ALI cases. Advancing the understanding of the epidemiology and natural history of ALI can facilitate the development of more valid and reliable case definitions that may improve our ability to monitor for ALI in a postmarket risk identification and analysis system in order to better understand the safety profile of newly approved drugs.

Acknowledgements We would like to thank Jesse A. Berlin, ScD, of the OMOP Scientific Advisory Board and Craig Cheetham, PharmD, of Kaiser Permanente Southern California for their critical and thorough review of the manuscript.

Financial support The OMOP is funded by the Foundation for the National Institutes of Health through generous contributions from the following: Abbott, Amgen Inc., AstraZeneca, Bayer Healthcare Pharmaceuticals, Inc., Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Johnson & Johnson, Lundbeck, Inc., Merck & Co., Inc., Novartis Pharmaceuticals Corporation, Pfizer Inc., Pharmaceutical Research Manufacturers of America (PhRMA), Roche, Sanofi, Schering-Plough Corporation, and Takeda.

Conflict of interest Patrick Ryan and Paul Stang are employees of Janssen Research and Development, which provides funding for the OMOP, and have received OMOP-funded travel support. They both hold stock in multiple pharmaceutical companies. Aaron Katz and Judith Racoosin have no conflicts of interest to declare that are directly relevant to the content of this study.

References

1. Fontana RJ. Approaches to the study of drug-induced liver injury. Clin Pharmacol Ther. 2010;88(3):416–9.
2. Meropol SB, Chan KA, Chen Z, et al. Adverse events associated with prolonged antibiotic use. Pharmacoepidemiol Drug Saf. 2008;17(5):523–32.
3. Garcia-Rodriguez LA, Masso-Gonzalez EL, Wallander MA, et al. The safety of rosuvastatin in comparison with other statins in

over 100,000 statin users in UK primary care. Pharmacoepidemiol Drug Saf. 2008;17(10):943–52.
4. Garcia Rodriguez LA, Williams R, Derby LE, et al. Acute liver injury associated with nonsteroidal anti-inflammatory drugs and the role of risk factors. Arch Intern Med. 1994;154(3):311–6.
5. Garcia Rodriguez LA, Duque A, Castellsague J, et al. A cohort study on the risk of acute liver injury among users of ketoconazole and other antifungal drugs. Br J Clin Pharmacol. 1999;48(6):847–52.
6. de Abajo FJ, Montero D, Madurga M, et al. Acute and clinically relevant drug-induced liver injury: a population based case-control study. Br J Clin Pharmacol. 2004;58(1):71–80.
7. Clifford GM, Logie J, Farmer RD. No risk of drug-associated liver injury with alpha1-adrenoreceptor blocking agents in men with BPH: results from an observational study using the GPRD. Pharmacoepidemiol Drug Saf. 2005;14(2):75–80.
8. Suissa S, Ernst P, Hudson M, et al. Newer disease-modifying antirheumatic drugs and the risk of serious hepatic adverse events in patients with rheumatoid arthritis. Am J Med. 2004;117(2):87–92.
9. Perez Gutthann S, Garcia Rodriguez LA. The increased risk of hospitalizations for acute liver injury in a population with exposure to multiple drugs. Epidemiology. 1993;4(6):496–501.
10. McAfee AT, Ming EE, Seeger JD, et al. The comparative safety of rosuvastatin: a retrospective matched cohort study in over 48,000 initiators of statin therapy. Pharmacoepidemiol Drug Saf. 2006;15(7):444–53.
11. Lee CH, Wang JD, Chen PC. Increased risk of hospitalization for acute hepatitis in patients with previous exposure to NSAIDs. Pharmacoepidemiol Drug Saf. 2010;19(7):708–14.
12. Jinjuvadia K, Kwan W, Fontana RJ. Searching for a needle in a haystack: use of ICD-9-CM codes in drug-induced liver injury. Am J Gastroenterol. 2007;102(11):2437–43.
13. Heaton PC, Fenwick SR, Brewer DE. Association between tetracycline or doxycycline and hepatotoxicity: a population based case-control study. J Clin Pharm Ther. 2007;32(5):483–7.
14. Graham DJ, Drinkard CR, Shatin D. Incidence of idiopathic acute liver failure and hospitalized liver injury in patients treated with troglitazone. Am J Gastroenterol. 2003;98(1):175–9.
15. Goettsch WG, Heintjes EM, Kastelein JJ, et al. Results from a rosuvastatin historical cohort study in more than 45,000 Dutch statin users, a PHARMO study. Pharmacoepidemiol Drug Saf. 2006;15(7):435–43.
16. Garcia-Rodriguez LA, Gonzalez-Perez A, Stang MR, et al. The safety of rosuvastatin in comparison with other statins in over 25,000 statin users in the Saskatchewan Health Databases. Pharmacoepidemiol Drug Saf. 2008;17(10):953–61.
17. Enger C, Gately R, Ming EE, et al. Pharmacoepidemiology safety study of fibrate and statin concomitant therapy. Am J Cardiol. 2010;106(11):1594–601.
18. Duh MS, Vekeman F, Korves C, et al. Risk of hepatotoxicity-related hospitalizations among patients treated with opioid/acetaminophen combination prescription pain medications. Pain Med. 2010;11(11):1718–25.
19. Cziraky MJ, Willey VJ, McKenney JM, et al. Statin safety: an assessment using an administrative claims database. Am J Cardiol. 2006;97(8A):61C–8C.
20. Chan KA, Truman A, Gurwitz JH, et al. A cohort study of the incidence of serious acute liver injury in diabetic patients treated with hypoglycemic agents. Arch Intern Med. 2003;163(6):728–34.
21. Carson JL, Strom BL, Duff A, et al. Acute liver disease associated with erythromycins, sulfonamides, and tetracyclines. Ann Intern Med. 1993;119(7 Pt 1):576–83.
22. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med. 2010;153(9):600–6.
23. Stang PE, Ryan PB, Dusetzina SB, et al. Health outcomes of interest in observational data: issues in identifying definitions in the literature. Health Outcomes Res Med. 2012;3(1):e37–44.
24. Food and Drug Administration. Guidance for industry on drug-induced liver injury: premarketing clinical evaluation. Silver Spring: US Department of Health and Human Services; 2009. http://www.fda.gov/downloads/Drugs/.../Guidances/UCM174090.pdf (Accessed 26 June 2011).
25. Chalasani N, Fontana RJ, Bonkovsky HL, et al. Causes, clinical features, and outcomes from a prospective study of drug-induced liver injury in the United States. Gastroenterology. 2008;135(6):1924–34, 1934.e1–4.
26. Clark JM. The epidemiology of nonalcoholic fatty liver disease in adults. J Clin Gastroenterol. 2006;40(Suppl. 1):S5–10.
27. Daniels D, Grytdal S, Wasley A. Surveillance for acute viral hepatitis—United States, 2007. MMWR Surveill Summ. 2009;58(3):1–27.
28. Cohen SM, Ahn J. Review article: the diagnosis and management of alcoholic hepatitis. Aliment Pharmacol Ther. 2009;30(1):3–13.
29. Ozer JS, Chetty R, Kenna G, et al. Recommendations to qualify biomarker candidates of drug-induced liver injury. Biomark Med. 2010;4(3):475–83.