DD Month YYYY
EMA/761520/2021

Data analysis plan

Prevalence of Acute Liver Injury

1. Rationale and background

A study on prevalence of acute liver injury was requested to support regulatory decision-making process on an orphan application in the indication of acute liver failure of any cause.

Acute liver injury is defined as a sudden appearance of liver test abnormalities (increased liver enzymes or markers of abnormal liver function) that range from mild abnormal biochemical liver values to acute liver failure (1-3).

It is of interest to obtain data on the yearly prevalence of acute liver injury in the European population, overall, and stratified by age group and gender. Drug-induced liver injury (DILI) is an important cause of acute liver injury. The incidence of DILI has been estimated at around 10-200 per million persons per year (4-8) with slightly higher rates of up to 260 per million persons per year in mainland China (9).

Acute liver injury can lead to liver failure that is generally estimated to occur in fewer than 10 cases per million persons per year (11) with a recent estimate from Germany at around 11.3 per million per year (12). According to Orphanet, acute liver failure is a rare disease with a prevalence of 1-5 per 10,000 persons Almost 50% of the patients died within 3 months of the diagnosis.

This study attempts to identify patients at risk of acute liver failure rather than patients already diagnosed with acute liver failure. Sensitivity analyses will be conducted, including and excluding diagnoses codes of acute liver failure.

2. Research question and objectives

The main objective of the study is to estimate the yearly prevalence of acute liver injury in five EU countries (Germany, France, Italy, Romania, and Spain), overall and stratified by gender and age group (0-17 years, 18-49 years, 50-79 years, 80+ years).

3. Research methods

3.1. Study design

This is a descriptive study of the yearly prevalence of acute liver injury, overall and stratified by gender and age group. No hypothesis will be tested.

3.2. Setting and study population

The study population will be patients visiting general practices in France, Italy, Germany, Romania, and Spain. In Germany paediatric practices will also be included as they are part of primary care for children. Patients will be considered observable during a year if they are observable for at least one day during the year. Patients are considered to be observable between their first and last visit dates or, alternatively, between the date of registering with the practice and until de-registration.

Study period will be between January 2016 and December 2020.

3.3. Variables

To distinguish acute from chronic liver injury it will be required that the patient has a disease duration of less than 26 weeks (13).

Case definition

A validation study of WHO ICD 10 diagnosis codes for acute liver injury in patients treated with antidepressants has been published [3]. A set of specific codes was identified that had a high positive predictive value for acute liver injury, including the WHO ICD 10 codes K71.0 (toxic liver disease with cholestasis), K71.1 (toxic liver disease with hepatic necrosis), K71.2 (toxic liver disease with acute hepatitis), K71.6 (toxic liver disease, not elsewhere classified), K71.9 (toxic liver disease, unspecified), K72.0 (acute and subacute hepatic failure) and K72.9 (hepatic failure, unspecified), K75.9 (inflammatory liver disease, unspecified), and K76.2 (central haemorrhagic necrosis of liver). However, of interest for this study were also specifically ischaemic, autoimmune, viral and other forms of acute and subacute liver injury, including Wilson's disease [10]. These conditions represent both patients without chronic liver disease that experience acute liver injury, and patients with chronic liver disease due to Wilson's disease, autoimmunity, or reactivation of chronic viral hepatitis who are at risk of acute liver failure [10]. For this reason, a broader set of diagnosis codes was used. However, to avoid acute-on-chronic liver injury, chronic hepatitis not elsewhere classified (WHO ICD 10 code K73), chronic liver failure (WHO ICD 10 code K72.1), and fibrosis and cirrhosis of liver (WHO ICD 10 code K74) were not considered.

Acute liver injury will be identified based on the WHO ICD 10 codes B15 (acute hepatitis A), B16 (acute hepatitis B), B17 (other acute hepatitis), B18 (chronic viral hepatitis), B19 (unspecified viral hepatitis), B25.1 (cytomegaloviral hepatitis), B58.1 (toxoplasma hepatitis), B94.2 (sequelae of viral hepatitis), E83.0 (disorders of copper metabolism), K71.0 (toxic liver disease with cholestasis), K71.1 (toxic liver disease with hepatic necrosis), K71.2 (toxic liver disease with acute hepatitis), K71.6 (toxic liver disease, not elsewhere classified), K71.9 (toxic liver disease, unspecified), K72.0 (acute and subacute hepatic failure), K72.9 (hepatic failure, unspecified), K75 (inflammatory liver disease), K76 (other diseases of liver), O14.2 (HELLP syndrome), T39.1 (poisoning by 4-aminophenol derivatives) and Y45.5 (adverse effects of therapeutic use of 4-aminophenol derivatives). Patients will be considered to have acute liver injury up to a maximum of 25 weeks and 6 days (181 days) after the acute liver injury diagnosis, except for patients with Wilson's disease (ICD 10 code E83.0) that will be considered to have acute liver injury during their entire follow-up time.

A sensitivity analysis will exclude diagnoses codes for acute or subacute hepatic failure (WHO ICD 10 codes K72.0 and K72.9).

This case definition is based on etiological factors for acute liver failure described by Stravitz *et al*. (10)

In the THIN® Spain and THIN® Italy databases, diagnoses are recorded using ICD 9 codes. Corresponding ICD 9 codes were identified, please see Annex 2.

3.4. Data sources

The following databases and versions will be used:

- IQVIA™Disease Analyzer France version December 2021
- IQVIA™Disease Analyzer Germany version December 2021
- THIN® Italy version February 2022
- THIN® Romania version February 2022
- THIN® Spain version February 2022

A brief description of these databases is provided in **Annex 1**.

3.5. Statistical analysis

3.5.1. Main statistical methods

The prevalence will be estimated as the number of patients with the condition anytime during the year according to the case definition (i.e. all patients that were observable during the year and had received a diagnosis of acute liver injury during the year or up to 181 days before* the start of the year) per million persons in the population that were observable during the year. The period of 181 days will only be applied to the first date when the specific diagnosis is recorded in the patient, but the same patient can contribute to more than one 181-day period if more than one acute liver injury diagnosis is recorded.

* For Wilson's disease no time limit will be applied.

All analyses will be performed using SAS Enterprise Guide version 7.15.

3.5.2. Sensitivity analysis

A sensitivity analysis will exclude diagnoses codes for acute or subacute hepatic failure (WHO ICD 10 codes K72.0 and K72.9).

3.6. Quality control

The study will be conducted according to the ENCePP code of conduct (European Medicines Agency 2018).

Standard operating procedures or internal process guidance will be adhered to for the conduct of the study. These procedures include rules for secure and confidential data storage, quality-control procedures for all aspects of the study from protocol development to the reporting of the results.

All documents will undergo at least one round a review by an experienced reviewer, while the results from the statistical analysis will be either reviewed or checked via double coding.

The quality control of the data is the responsibility of the data holder.

3.7. Limitations of the research methods

Many patients with acute liver injury are likely to be hospitalised and diagnosed in the secondary care hospital setting. However, since this study is based on data from patients visiting primary care, patients not visiting primary care are not included in the study and patients diagnosed only in secondary care may be been missed, which might limit how well the results are representative of the overall population.

In addition, the study may carry some misclassification of cases as some codes used to identify patients may not be specific enough.

4. Protection of human subjects

Patient confidentiality will be protected according to the EU General Data Protection Regulation (GDPR) on the protection of individuals.

For information entering the public domain such as publication in EU-PAS register, and in accordance with database rules on the management of low cell counts, cells with low numbers (<10 in IQVIA™

Disease Analyzer France, THIN® Spain, Italy and Romania) will be removed prior to publication of this report. Additional cells may be redacted (events/patients typically being rounded up to the nearest 10) if needed in order to ensure that the aforementioned low cell counts cannot be re-identified. This may include both events/patients and follow-up times.

5. Management and reporting of adverse events/adverse reactions

Pursuant to the requirements for reporting of adverse events for secondary data (GVP module VI, VI.C.1.2.1.2), adverse event reporting will not be conducted as part of this study given the study objectives will be met through the use of secondary data.

6. Plans for disseminating and communicating study results

The analysis plan and study results will be published in EUPAS registries upon completion.

7. References

1. Hussaini, S.H. and E.A. Farrington, *Idiosyncratic drug-induced liver injury: an update on the 2007 overview.* Expert Opin Drug Saf, 2014. **13**(1): p. 67-81.
2. Lee, W.M., R.T. Stravitz, and A.M. Larson, *Introduction to the revised American Association for the Study of Liver Diseases Position Paper on acute liver failure 2011.* Hepatology, 2012. **55**(3): p. 965-7.
3. Forns, J., et al., *Validity of ICD-9 and ICD-10 codes used to identify acute liver injury: A study in three European data sources.* Pharmacoepidemiol Drug Saf, 2019. **28**(7): p. 965-975.
4. Navarro, V.J. and J.R. Senior, *Drug-related hepatotoxicity.* N Engl J Med, 2006. **354**(7): p. 731-9.
5. Larrey, D., *Epidemiology and individual susceptibility to adverse drug reactions affecting the liver.* Semin Liver Dis, 2002. **22**(2): p. 145-55.
6. Watkins, P.B. and L.B. Seeff, *Drug-induced liver injury: summary of a single topic clinical research conference.* Hepatology, 2006. **43**(3): p. 618-31.
7. Bell, L.N. and N. Chalasani, *Epidemiology of idiosyncratic drug-induced liver injury.* Seminars in liver disease, 2009. **29**(4): p. 337-347.
8. Björnsson, E., *Review article: drug-induced liver injury in clinical practice.* Aliment Pharmacol Ther, 2010. **32**(1): p. 3-13.
9. Shen, T., et al., *Incidence and Etiology of Drug-Induced Liver Injury in Mainland China.* Gastroenterology, 2019. **156**(8): p. 2230-2241.e11.
10. Bernal, W. and J. Wendon, *Acute Liver Failure.* New England Journal of Medicine, 2013. **369**(26): p. 2525-2534.
11. Weiler, N., et al., *The Epidemiology of Acute Liver Failure.* Deutsches Arzteblatt international, 2020. **117**(4): p. 43-50.
12. Pievsky, D., N. Rustgi, and N.T. Pyrsopoulos, *Classification and Epidemiologic Aspects of Acute Liver Failure.* Clin Liver Dis, 2018. **22**(2): p. 229-241.

Annexes

Annex 1 - Information on Databases and Healthcare systems included

IQVIA™ Disease Analyzer Germany

IQVIA™ Disease Analyzer Germany collects computerised information from specialised and general primary care practices throughout Germany since 1992. Around 3% of general practitioners (GP) practices are included, which covers all patients consulting a practice. Data from IQVIA™ Disease Analyzer Germany have been shown to be reasonably representative of German healthcare statistics for demographics and certain diseases and is considered one of the largest national medical databases worldwide. IQVIA™ Disease Analyzer Germany includes more than 2,500 practices and 3,100 physicians (13 speciality groups) representing over 15,000,000 patients. This database used to be named IMS® Germany and some use of this terminology may persist.

The quality of IQVIA™ Disease Analyzer data is ensured by a series of continuous QA controls and data refinement. These include checking incoming data for criteria such as completeness and correctness, (e.g. linkage between diagnoses and prescriptions), and standardizing certain data values such as laboratory test results in order to enable reliable analysis.

IQVIA™ Disease Analyzer France

IQVIA™ Disease Analyzer France collects anonymised patient medical records since 1997 through a representative panel of GPs. The physician sample represents approximately 2% of physicians and is weighted by age and gender of the physician, doctor region and the SNIR of the physician (National Official Indicator of the GP volume of activity in terms of visits and consultations). Some 99% of the French population is insured, but there are differences regarding level of coverage. IQVIA™ Disease Analyzer France includes around 1,000 GPs and represents more than 4,000,000 of patients and considered representative for the French population. This database used to be named IMS France and some use of this terminology may persist.

The quality of IQVIA™ Disease Analyzer data is ensured by a series of continuous QA controls and data refinement. These include checking incoming data for criteria such as completeness and correctness, (e.g. linkage between diagnoses and prescriptions), and standardizing certain data values such as laboratory test results in order to enable reliable analysis.

The Health Improvement Network (THIN®) Italy

In THIN® Italy data collection started in 2000 and this database is currently able to provide clinical monitoring data of anonymised patients managed by 500 GPs in primary care (including patients' history). The data source of THIN® Italy is electronic health care records. The entire database reaches 900,000 patients (active and non-active), from which 500,000 are currently actively followed. In order to be representative at national and macroregional level, physicians have been recruited in accordance with their universe distribution in terms of geography, age and gender.

THIN® is an unobtrusive European medical data collection scheme that collects anonymized patient data from the Electronic Health Records of GPs and specialists, including information on patient's diagnoses, test results and medication. The databases follow a very strict anonymization process. In all countries patients are informed about the collection and anonymization of the data and they are able to opt out, in which case no data are subsequently transmitted to the THIN® database.

The Health Improvement Network (THIN®) Romania

THIN® Romania is a primary care healthcare database, including only General Practitioners (574 GPs). The source of data is electronic health care records. Enrolled GPs and their patients are representative of the whole Romanian population in terms of location, demographics and prevalence from the point of view of main chronic health pathologies. Data collection started in 2012.

In Romania, the insured population (background sampled population) numbered 17.1 million individuals (data from 2012). Among these, 8.5 million individuals benefited of healthcare services, in the public system. The number of GPs who worked in the public healthcare system, in 2017 was aproximately 11,000 physicians. They recorded 76 million consultations and issued 71 million prescriptions (data from 2017). The number of deceased patients was of 297,000 individuals, and number of newborns in 2020 was of 179,000 individuals.

THIN® is an unobtrusive European medical data collection scheme that collects anonymized patient data from the Electronic Health Records of GPs and specialists, including information on patient's diagnoses, test results and medication. The databases follow a very strict anonymization process. In all countries patients are informed about the collection and anonymization of the data and they are able to opt out, in which case no data are subsequently transmitted to the THIN® database.

The Health Improvement Network (THIN®) Spain

THIN® Spain is mainly a primary care healthcare database, including practitioners (GP), specialists and pediatricians & nurses. It contains data from approximately 2,000 GPs and 2,400 specialists (cardiology, pulmonology, urology, etc.). THIN® Spain also includes partial activities related to the hospital. THIN® Spain is globally representative of the whole national demographics and prevalence on the main chronic health pathologies. THIN® Spain includes 3,000,000 individuals out of the overall population. Among these, 1,050,000 are active in the previous year and 1,800,000 are active from 2014. Number of deceased patients globally varies between 8 and 9 thousand individuals per year, and number of new-borns ranges between 10 and 12 thousand individuals. New patients are automatically included into the database, and deceased patients identified in a specific field.

THIN® is an unobtrusive European medical data collection scheme that collects anonymized patient data from the Electronic Health Records of GPs and specialists, including information on patient's diagnoses, test results and medication. The databases follow a very strict anonymization process. In all countries patients are informed about the collection and anonymization of the data and are able to opt out, in which case no data are subsequently transmitted to the THIN® database.

The THIN® Spain Database has been approved by two Ethics Committees, one from the Community of Madrid (Hospital Ramón Cajal) and one from the Community of Catalonia (Hospital Clinic de Barcelona). These ethics committees reviewed the data collection, protection, and anonymization processes and positively approved THIN® Spain for observational research of medical products (upon protocol submission).

Annex 2 – Code lists

List of WHO ICD codes for acute liver injury

WHO ICD 10 code	Description
B15	Acute hepatitis A
B16	Acute hepatitis B
B17	Other acute hepatitis
B18	Chronic viral hepatitis
B19	Unspecified viral hepatitis
B25.1	Cytomegaloviral hepatitis
B58.1	Toxoplasma hepatitis
B94.2	Sequelae of viral hepatitis
E83.0	Disorders of copper metabolism
I82.0	Budd-Chiari syndrome
K71.0	Toxic liver disease with cholestasis
K71.1	Toxic liver disease with hepatic necrosis
K71.2	Toxic liver disease with acute hepatitis
K71.6	Toxic liver disease, not elsewhere classified
K72.0	Acute and subacute hepatic failure*
K72.9	Hepatic failure, unspecified*
K75	Other inflammatory liver diseases
K76	Other diseases of liver
O14.2	HELLP syndrome
T39.1	Poisoning by 4-aminophenol derivatives
Y45.5	Adverse effects of therapeutic use of 4-aminophenol derivatives

* A sensitivity analysis will exclude diagnoses codes for acute or subacute hepatic failure (WHO ICD 10 codes K720 and K72.9).

List of corresponding WHO ICD 9 codes for acute liver injury (THIN Italy and Spain)

WHO ICD 9 code *	Description
070 (all subcodes)	Viral hepatitis
130.5	Hepatitis due to toxoplasmosis

WHO ICD 9 code *	Description
275.1	Disorders of copper metabolism
453.0	Budd-chiari syndrome
570 **	Acute and subacute necrosis of liver
571.8	Other chronic nonalcoholic liver disease
572	Liver abscess and sequelae of chronic liver disease
572.0	Abscess of liver
572.1	Portal pyemia
572.2 **	Hepatic encephalopathy
572.3	Portal hypertension
572.4	Hepatorenal syndrome
573	Other disorders of liver
573.0	Chronic passive congestion of liver
573.1	Hepatitis in viral diseases classified elsewhere
573.2	Hepatitis in other infectious diseases classified elsewhere
573.3	Hepatitis, unspecified
573.4	Hepatic infarction
573.8	Other specified disorders of liver
573.9	Unspecified disorder of liver
642.5 (all subcodes)	Severe pre-eclampsia
965.4	Poisoning by aromatic analgesics, not elsewhere classified
E935.4	Pyrazole derivatives causing adverse effects in therapeutic use

* Please see http://www.icd9data.com/2015/Volume1/001-139/default.htm

** A sensitivity analysis excluded diagnoses codes for acute or subacute hepatic failure (WHO ICD 9 codes 570 and 572.2).