ORIGINAL ARTICLE

Chronic liver disease-associated severe thrombocytopenia in Spain: Results from a retrospective study using machine learning and natural language processing

Jose Luis Calleja-Panero [a,*], Rafael Esteban Mur [b], Isidro Jarque [c], Manuel Romero-Gómez [d], Savana Research Group [e], Lucia García Labrador [f], Jose González Calvo [f]

[a] Department of Hepatology, Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain
[b] Department of Hepatology, Hospital Universitario Vall d'Hebron, Barcelona, Spain
[c] Department of Hematology, Hospital Universitari i Politècnic La Fe, Valencia, Spain
[d] Department of Hepatology, Hospital Universitario Virgen del Rocío, Sevilla, Spain
[e] Medsavana and Savana Research S.L., Madrid, Spain
[f] Shionogi S.L.U., Madrid, Spain

Abstract

Background: Patients with chronic liver disease (CLD) often develop thrombocytopenia (TCP) as a complication. Severe TCP (platelet count < 50×10^9/L) can increase morbidity and complicate CLD management, increasing bleeding risk during invasive procedures.

Objectives: To describe the real-world scenario of CLD-associated severe TCP patients' clinical characteristics. To evaluate the association between invasive procedures, prophylactic treatments, and bleeding events in this group of patients. To describe their need of medical resource use in Spain.

Methods: This is a retrospective, multicenter study including patients who had confirmed diagnosis of CLD and severe TCP in four hospitals within the Spanish National Healthcare Network from January 2014 to December 2018. We analyzed the free-text information from Electronic Health Records (EHRs) of patients using Natural Language Processing (NLP), machine learning techniques, and SNOMED-CT terminology. Demographics, comorbidities, analytical parameters and characteristics of CLD were extracted at baseline and need for invasive procedures, prophylactic treatments, bleeding events and medical resources used in the follow up period. Frequency tables were generated for categorical variables, whereas continuous variables were described in summary tables as mean (SD) and median (Q1–Q3).

* Corresponding author.
  E-mail address: joseluis.calleja@uam.es (J.L. Calleja-Panero).

Results: Out of 1,765,675 patients, 1787 had CLD and severe TCP; 65.2% were male with a mean age of 54.7 years old. Cirrhosis was detected in 46% (*n* = 820) of patients and 9.1% (*n* = 163) had hepatocellular carcinoma. Invasive procedures were needed in 85.6% of patients during the follow up period. Patients undergoing procedures compared to those patients without invasive procedures presented higher rates of bleeding events (33% vs 8%, *p* < 0.0001) and higher number of bleedings. While prophylactic platelet transfusions were given to 25.6% of patients undergoing procedures, TPO receptor agonist use was only detected in 3.1% of them. Most patients (60.9%) required at least one hospital admission during the follow up and 14.4% of admissions were due to bleeding events with a hospital length of stay of 6 (3, 9) days.

Conclusions: NLP and machine learning are useful tools to describe real-world data in patients with CLD and severe TCP in Spain. Bleeding events are frequent in those patients who need invasive procedures, even receiving platelet transfusions as a prophylactic treatment, increasing the further use of medical resources. Because that, new prophylactic treatments that are not yet generalized, are needed.

Trombocitopenia grave asociada a enfermedad hepática crónica en España: Resultados de un estudio retrospectivo utilizando aprendizaje automático y procesamiento de lenguaje natural

Resumen

Antecedentes: Los pacientes con enfermedad hepática crónica (EHC) a menudo desarrollan trombocitopenia (TCP) como agravante de su enfermedad. La TCP grave (definida por un recuento de plaquetas < 50 x 10^9/L) puede aumentar la morbilidad y complicar el manejo de la EPC, incrementando el riesgo de hemorragia durante los procedimientos invasivos.

Objetivos: Describir el escenario de mundo real de las características clínicas de los pacientes con TCP grave asociado a EHC. Evaluar la asociación entre procedimientos invasivos, tratamientos profilácticos y eventos hemorrágicos en este grupo de pacientes, así como describir el uso de recursos médicos en España.

Métodos: Se plantea un estudio multicéntrico retrospectivo que incluye pacientes con diagnóstico confirmado de EHC y TCP grave en cuatro hospitales de la Red Nacional de Salud de España desde enero de 2014 hasta diciembre de 2018. Analizamos la información de texto libre de la Historia Clínica Electrónica (HCE) de pacientes que utilizan procesamiento de lenguaje natural (PLN), técnicas de aprendizaje automático y terminología de SNOMED-CT. Los datos demográficos, las comorbilidades, los parámetros analíticos y las características de la EHC se extrajeron al inicio del estudio, así como la necesidad de procedimientos invasivos, tratamientos profilácticos, eventos hemorrágicos y recursos médicos utilizados en el periodo de seguimiento. Se generaron tablas de frecuencia para las variables categóricas, mientras que las variables continuas se describieron en tablas resumen como media (SD) y mediana (Q1-Q3).

Resultados: De 1.765.675 pacientes identificados, 1.787 tenían EHC y TCP grave, siendo el 65,2% varones con una edad media de 54,7 años. Se detectó cirrosis en el 46% (n = 820) de los pacientes y el 9,1% (n = 163) de ellos presentaron un diagnóstico de carcinoma hepatocelular. Se necesitaron procedimientos invasivos en el 85,6% de los pacientes durante el periodo de seguimiento, presentando una mayor tasa de eventos hemorrágicos (33 vs. 8%, p < 0,0001) y mayor número de sangrados. El 25,6% de los pacientes sometidos a procedimientos invasivos recibieron transfusiones profilácticas de plaquetas, si bien el uso de agonistas del receptor de trombopoyetina (TPO) solo se detectó en el 3,1% de los casos. La mayoría de los pacientes (60,9%) requirieron al menos un ingreso hospitalario durante el periodo de seguimiento, debiéndose el 14,4% de ellos a eventos hemorrágicos con una estancia hospitalaria de 6 (3, 9) días.

Conclusiones: El PNL y el aprendizaje automático son herramientas útiles para describir datos del mundo real en pacientes con EHC y TCP grave en España. Los eventos hemorrágicos son frecuentes en aquellos pacientes que se someten a procedimientos invasivos, incluso recibiendo transfusiones de plaquetas como tratamiento profiláctico, lo que supone un incremento del uso de recursos médicos. Estos datos apuntan hacia la necesidad actual existente de generalización de nuevos tratamientos profilácticos.

Introduction

Thrombocytopenia (TCP) is a common complication in patients with chronic liver disease (CLD). Its prevalence in CLD has been reported to be up to 78% of patients with cirrhosis, varying according to various factors, such as the criteria used to define this hematologic abnormality and the severity of the underlying liver disease.[1,2] Multiple factors can contribute to the development of TCP in this group of patients, including splenic platelet sequestration, bone marrow suppression by chronic hepatitis C virus (HCV) infection, and antiviral treatments.[3] While the clinical impact of mild and moderate TCP is minimal and usually does not interfere with treatment or management decisions, severe TCP (defined as a platelet count $< 50 \times 10^9$/L) can be associated with significant morbidity. Severe TCP frequently complicates the medical management of patients with advanced liver disease and other disorders such as cancer, immune thrombocytopenic purpura (ITP), or HCV infection.[4–7]

Severe TCP can significantly increase the risk of bleeding, especially during surgery or other invasive procedures.[8,9] As such, CLD patients often require numerous medical procedures during diagnosis and therapy, and the presence of severe TCP can significantly complicate routine patient care, resulting in delayed or canceled procedures.[10] Platelet transfusion has generally been considered the gold standard for treatment of severe TCP prior to an invasive surgery, as it can correct low platelet counts and reduce the risk of bleeding.[11] Despite its widespread use, this approach has several limitations, such as its short-lived efficacy, risk of transfusion-related reactions, ineffectiveness in patients with liver disease, and development of antiplatelet antibodies (alloimmunization).[12–14] Therefore, alternative therapeutic options have been sought to safely and effectively raise platelet counts and to reduce the incidence of bleeding events in this group of patients.[15] Currently, in clinical practice, thrombopoietin receptor agonists (TPO-RA) are being used, e.g., romiplostim and eltrombopag, since there is mounting evidence that impaired hepatic production of TPO may be a major cause of TCP in liver disease.[16–19] In this regard, thrombocytopenic patients with advanced liver disease, in fact, have inappropriately low levels of TPO.[1,20,21]

To the best of our knowledge, the current clinical scenario for patients with CLD-associated severe TCP has not been described in Spain in a real-world setting. Because that, our main objectives were: (1) to describe the real-world scenario of CLD-associated severe TCP patients' clinical characteristics, (2) to evaluate the association between invasive procedures, prophylactic treatments and bleeding events in this group of patients and (3) to describe their need of medical resource use in Spain. We hence carried out this study using Natural Language Processing (NLP) and machine learning (ML) techniques to analyze the information included in Electronic Health Records (EHRs) of patients with severe TCP associated to CLD in Spain.

Methods

Study design

This was a retrospective, multicenter, non-interventional study using secondary free-text data from patients' EHRs of four hospitals representing different areas of the Spanish National Healthcare Network: Hospital Universitario Río Hortega (Valladolid), Hospital Universitario Puerta de Hierro-Majadahonda (Madrid), Hospital Universitario y Politécnico La Fe (Valencia), and Hospital Universitario Vall d'Hebron (Barcelona). Data from all available departments, including inpatient hospital, outpatient hospital, and emergency room, for virtually all types of provided services in each participating hospital were retrieved from January 1st, 2014 to December 31st, 2018.

Patient population and variable definitions

All patients who had a confirmed diagnosis of CLD were included when severe TCP was associated (index date). Patients were analyzed at baseline (3-month period before index date) and during a follow up (period including the index date until the end of the study period or last EHR available).

Severe TCP was defined by laboratory values of platelet levels $< 50 \times 10^9$/L. Variables related with clinical characteristics at baseline including demographics, toxic habits, comorbidities, and analytic values as well as CLD characteristics (time from CLD to severe TCP, etiology, and complications) were extracted. During the follow up period following variables were also extracted: need for invasive procedures and type, bleeding events, prophylactic treatments, and medical resources used (number of visits and length of hospital stay or LOS when admission occurred). Invasive procedures ordered from highest to lowest risk of bleeding, included surgical interventions, TIPS, endoscopic procedures, biopsies, paracentesis, thoracocentesis, and other procedures (including percutaneous, endovascular or endoluminal procedures). They were stratified into major (when performed under general or regional anesthesia and respiratory assistance) or minor (those that did not require postoperative resuscitation, were performed on an outpatient basis with local or troncular anesthesia, on superficial tissues, and/or easily accessible structures under regional anesthesia) and urgent (when included in the emergency department medical records) or non-urgent (when included in any other medical record). Prophylactic treatment was considered when administered prior to invasive procedures. In the case of TPO-RA drugs, only those initiated during previous 15 days and lasted days or weeks (not months, to exclude chronic treatments) were considered.

Most of the variables were included by direct detection but some of them were inferred from other variables, additionally to their direct detection. In this regard, ''Platelet transfusion'' was inferred by combining detection of ''transfusion'' and low platelet levels at the same timepoint.

Extracting free text from EHRs: EHRead®

An innovating data-driven system based on Natural Language Processing (NLP), machine learning, and deep learning (*EHRead*® technology) was used to conduct the study.[22–28] Natural language was extracted from EHRs and captured as free text (unstructured information). The free text was translated into medical concepts and related terms, e.g., laboratory parameters and acronyms, based on SNOMED-CT: a comprehensive, scientifically validated terminology. Regular expression rules and machine learning models were used to classify the concepts, and deep learning classification was used to reflect negation, affirmation, or speculation.

External validation

To evaluate the ability of EHRead® technology to correctly identify records mentioning CLD and related variables, an external validation was performed, as previously described.[22,29] Expert physicians generated an annotated corpus by marking a set of EHRs with metadata tags related to CLD and related variables which was used as a gold standard. By comparing EHRead®'s output with this corpus, performance was calculated and reflected in the following metrics: accuracy (P), recall (R), and their harmonic mean (F1-score).

Data analyses

Frequency tables were generated for categorical variables, whereas continuous variables were described by means of summary tables that included the mean, standard deviation (SD), median, and quartiles (Q1-Q3) of each variable. To test for possible statistically significant differences in the distribution of categorical variables between study groups (i.e., presence or absence of invasive procedures), we used Yates-corrected chi^2 tests.

Ethical considerations

This study was classified as a ''post-authorization study with other designs; not for prospective follow up'' (EPA-OD) by the Spanish Agency of Medicines and Health Products (AEMPS) and was approved by the Research Ethics Committee at Hospital Universitario Puerta de Hierro-Majadahonda (Spain). All methods and analysis were conducted in compliance with local legal and regulatory requirements, as well as generally accepted research practices described in the Helsinki Declaration in its latest edition, and Good Pharmacoepidemiology Practices. Data were analyzed from de-identified EHRs, which were aggregated in an irreversibly dissociated manner. Thus, individual patient consent was not required in the study.

Results

We used NLP and ML to analyze EHRs from a total of 1,765,675 patients. We finally detected and included in the study 1787 patients with CLD who had also severe TCP.

Baseline characteristics

Baseline characteristics of included patients are shown in Table 1. Briefly, 65.2% of CLD patients with severe TCP were male (n = 1165) with a mean age at index date of 54.7 (21.6) years. Within this group, 44.4% (n = 793) were regular alcohol users and 30.2% (n = 540) were smokers. Common comorbidities included arterial hypertension (28.6%; n = 511), dyslipidemia (21.5%; n = 385), and diabetes (21.4%; n = 383). Regarding laboratory values, the mean platelet count (n = 614) was 28.4×10^9/L (16.5), and the mean hemoglobin value (n = 1263) was 10.3 (2.4) g/dL.

Patient characteristics related to CLD at baseline are shown in Table 2. Severe TCP appeared after a median of 0.6 (1.6) years from the CLD confirmed diagnosis. CLD etiology was detected in 24.8% of included patients being chronic viral hepatitis B, C, D, or E the most reported cause (54.1%; n = 240), followed by hemochromatosis (13.9; n = 62), non-alcoholic steatohepatitis (9.7%; n = 43) and alcoholic liver disease (7.7%; n = 34). Among CLD complications, cirrhosis was detected in 46.0% (n = 820) of patients, whereas 9.1% (n = 163) had hepatocellular carcinoma. Some patients presented portal hypertension signs as splenomegaly (18.4%; n = 328) or gastroesophageal varices (13.2%; n = 236). We also detected that 2.7% (n = 49) had previously needed a transjugular intrahepatic portosystemic shunt (TIPS).

Invasive procedures and bleedings

Most patients (85.6%; n = 1529) underwent at least one invasive procedure, with a median of 4.0 (2.0–7.0) procedures per patient during the follow-up (Fig. 1A). A majority of them went through both major (82.1%; n = 1256) and minor (86.1%; n = 1317) invasive procedures and most of them were non-urgent (98.2%; n = 1502) while 43.8% (n = 670) were urgent (Fig. 1B). The most frequent types of invasive procedures were biopsies (62.8%; n = 960), surgical interventions (58.4%; n = 893), or endoscopic procedures (57.3%; n = 877) (Fig. 1C).

A total of 526 patients (29.4%) presented bleeding events (Fig. 2A). Bleeding events were presented in a higher percentage in those patients who needed any invasive procedure during the follow up compared with patients without them [33.0% vs 8.1%, p < 0.0001] (Fig. 2B). Details regarding the number of bleeding episodes in both groups are shown in Fig. 2C. A total of 39.8% (n = 201/505) of bleeding events in patients who underwent invasive procedures were occurred directly after the procedure. The median time between a bleeding event and the invasive procedure was 3 (2–6) days, using a 7-day post-procedure time window to capture information regarding bleeding.

Prophylactic treatments

We detected that 25.6% (n = 391) of included patients who needed invasive procedures received a platelet transfusion as a bleeding prophylaxis (Fig. 3A). Other prophylactic treatments as TPO-RA or antihemorrhagic agents were used in a lesser extent (3.1%, n = 47). We detected that 1.2% (n = 19) of patients undergoing an invasive procedure were treated with eltrombopag, a further 1.2% (n = 19) with etamsylate,

Table 1 Demographic and clinical characteristics of patients with CLD and severe TCP at baseline.

	CLD + severe TCP n = 1787	
Demographics		
Male, n (%)	1165 (65.2)	
Age (years)		
Mean (SD)	54.7 (21.6)	
Median (Q1, Q3)	58.0 (48.0, 69.0)	
Toxic habits, n (%)		
Alcohol		
Current	793 (44.4)	
Former	84 (4.7)	
Tobacco		
Current	540 (30.2)	
Former	386 (21.6)	
Other substances		
Current	211 (11.8)	
Former	59 (3.3)	
Comorbidities *n* (%)		
Arterial hypertension	511 (28.6)	
Dyslipidemia	385 (21.5)	
Diabetes	383 (21.4)	
Arrhythmias	334 (18.7)	
Valvopathies	253 (14.2)	
Cerebrovascular disease	97 (5.4)	
Congestive heart failure	48 (2.7)	
Chronic ischemic heart disease	11 (0.6)	
Other prothrombotic/procoagulant disorders	157 (8.8)	
Analytical parameters	Mean (SD)	Median (Q1, Q3)
Platelet count ($\times 10^9$/L), n = 614	28.4 (16.5)	31.0 (15.0–43.8)
White blood cell count ($\times 10^9$/L), n = 949	22.4 (102.3)	4.8 (2.6, 8.4)
Neutrophils ($\times 10^9$(L), n = 729	6.9 (24.4)	2.5 (1.1, 4.9)
Lymphocytes ($\times 10^9$/L), n = 703	7.4 (40.5)	0.8 (0.4, 1.6)
Eosinophils ($\times 10^9$/L), n = 624	0.6 (6.9)	0.0 (0.0, 0.1)
Hemoglobin (g/dL), n = 1263	10.3 (2.4)	10.0 (2.2–25)
INR, n = 990	1.5 (0.8)	1.3 (0.0–9.6)
Total bilirubin (g/dL), n = 903	3.1 (5.5)	1.4 (0.0–47.8)

and 0.6% (n = 9) with romiplostim (Fig. 3A). In the subgroup of patients who were treated with platelet transfusion prior to an invasive procedure, a median of 2 (1−4) invasive procedures per patient were observed. Invasive procedures involving platelet transfusions were classified as minor in 77.5% of patients (n = 303) and major in 62.7% (n = 245) and were predominantly surgical interventions (63.7%; n = 249), biopsies (46.3%; n = 181), and endoscopic procedures (41.9%, n = 164) (Fig. 3B). Among patients who needed an invasive procedure, a total of 599 (39.1%) received platelet transfusions during the follow up.

Hospital resources

Hospital admissions of included patients during all the study period were evaluated. Hospital LOS for these patients due to any cause or specifically due to bleeding events are shown in Table 3. Most patients were admitted to hospital (60.9%; n = 1089) throughout the study period, with a median LOS of 5.0 (3.0−9.0) days. A 14.4% (n = 157) of hospitalizations were due to bleeding events and those patients stayed a median of 6.0 (3.0−9.0) days. Departments or services most visited by CLD and severe TCP patients during the follow-up period are shown in Suppl. Table 1. The Accident and Emergency department was visited by 49.6% (n = 887) of the patients a median of 2.0 (1.0−4.0) times, followed by the Gastroenterology department, which was visited by 36.3% (n = 649) of patients, though more frequently, with a median of 11.0 (2.0−40.0) visits per patient. Visits to the Hematology department were recorded for 28.5% (n = 510) of patients, with a median of 5.0 (2.0−19.0) visits per patient. In turn, the Hepatology department was visited by 12.3% (n = 220) of the patients a median of 2.0 (1.0−4.0) visits.

Validation metrics

Validation metrics for the main variable of the study, namely, CLD, yielded an F1-score (harmonic mean between preci-

Table 2 Baseline characteristics related to CLD in included patients.

	CLD + severe TCP *n* = 1787
Time from CLD diagnosis since severe TCP, years	
Mean (SD)	0.6 (1.6)
Etiology of CLD, *n* (%)[a]	444 (24.8)
Chronic viral hepatitis B, C, D or E	240 (54.1)
Hemochromatosis	62 (13.9)
Non-alcoholic steatohepatitis (NASH)	43 (9.7)
Alcohol	34 (7.7)
Cryptogenic	27 (6.1)
Autoimmune hepatitis	27 (6.1)
Primary biliary cholangitis	7 (1.6)
Complications	
Cirrhosis, n (%)	820 (46.0)
Hepatocellular carcinoma, n (%)	163 (9.1)
Portal hypertension	
Signs, *n* (%)	
Splenomegaly	328 (18.4)
Gastroesophageal varices	236 (13.2)
Portal vein thrombosis	72 (4.0)
Hepatorenal syndrome	27 (1.5)
Hydrothorax	6 (0.3)
Other portosystemic collaterals	28 (1.6)
Transjugular intrahepatic portosystemic shunt (TIPS), n (%)	49 (2.7)
Time since TIPS (months)	
Mean (SD)	12.4 (35.7)
Median (Q1, Q3)	0.7 (0.1, 5.0)

[a] Percentages for this variable were calculated with respect to patients with recorded etiology of CLD, not the whole CLD + severe TCP group.

Table 3 Baseline characteristics related to CLD in included patients.

	CLD + severe TCP *n* = 1787
Patients with any hospital admission, *n* (%)	1089 (60.9)
Length of stay (days)	
Mean (SD)	6.2 (4.3)
Median (Q1, Q3)	5.0 (3.0, 9.0)
Patients with hospital admission due to bleeding events, *n* (%)	157(8.8)
Length of stay (days)	
Mean (SD)	6.7 (4.4)
Median (Q1, Q3)	6.0 (3.0, 9.0)

sion and recall) of 0.89, showing high agreement between physicians' annotations (using 1187 EHRs to obtain the gold standard) and *EHRead*® detection. Linguistic validation of other related variables including disease characterization (e.g., gastroesophageal varices, portal thrombosis, digestive bleeding), diagnosis (e.g., platelet count), and treatments (e.g., TIPS) returned an F-score above 0.80 (Suppl. Table 2). However, the F1-score for the term ''thrombocytopenia'' was 0.60, where precision was 0.82 but the recall was 0.47. TCP could be interpreted from the laboratory values even if the term was not contained in the EHRs, whereas validation refers only to terminology identification, such that inferences can make up for inadequate term detection. Similarly, ''platelet transfusion'' returned an F1-score of 0.44 since the term per se is rarely contained in the EHRs. Physicians interpreted the term ''platelet transfusion'' from the appearance of ''transfusion'' and the values in laboratory records. The temporality evaluation yielded a robust F1-score of 0.78, considering that temporality detection is one of the most complex NLP tasks, with F1-scores usually ranging between 0.551 and 0.795.[30,31]

Discussion

The key findings of this study are: (1) CLD associated severe TCP affects predominantly young men that are
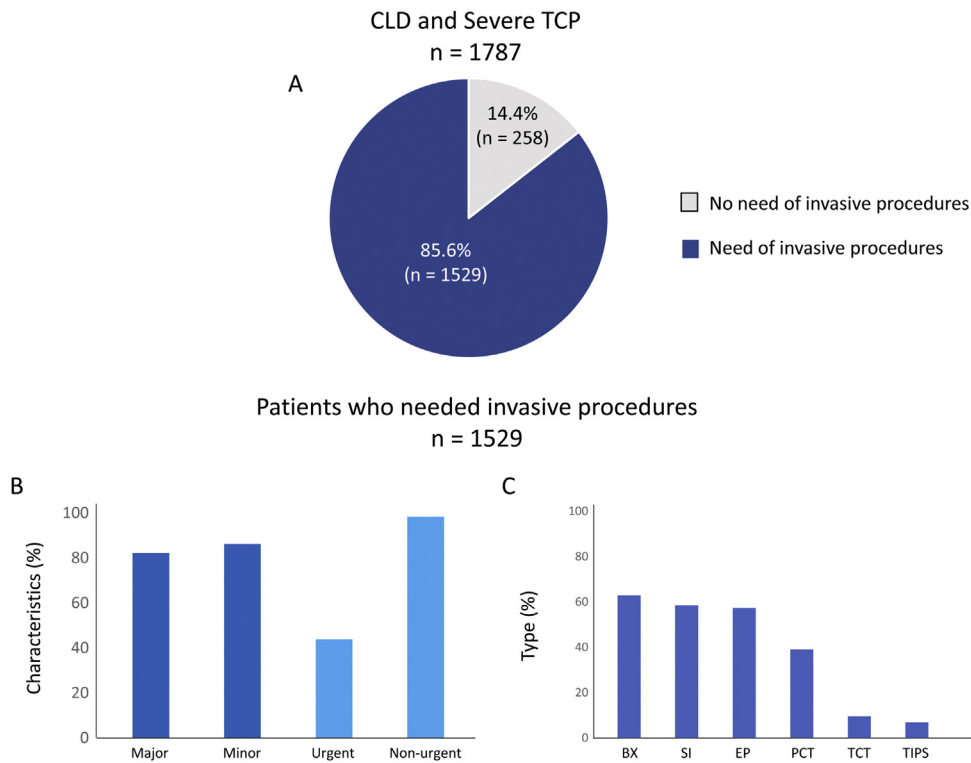
Figure 1 Invasive procedures realized in included patients. A shows the percentage of patients who needed an invasive procedure during the follow up. B and C show the characteristics and types respectively of invasive procedures performed on included patients. SI: surgical intervention; BX: biopsy; EP: endoscopic procedures; PCT: paracentesis; TCT: thoracocentesis; TIPS: transjugular intra-hepatic portosystemic shunt and OTH: others.
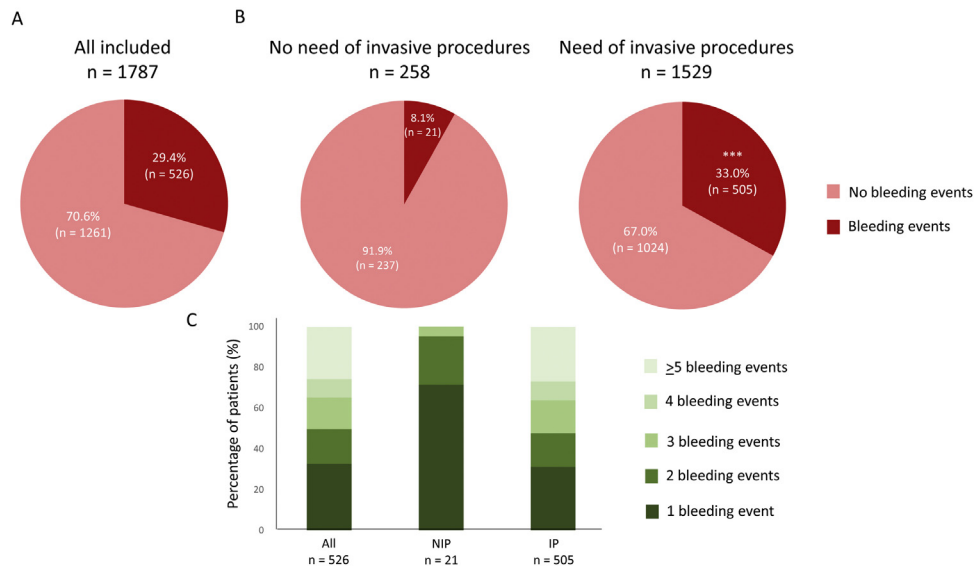

Figure 2 Bleeding events in all included patients (A) and regarding the need of invasive procedures (B) during the follow up. C shows the number of bleeding events in the three groups during the follow up. ***$p < 0.0001$ vs no need of invasive procedures. NIP: no need of invasive procedures; IP: need of invasive procedures.

alcohol users with other complications derived from the CLD; (2) Most of them need invasive procedures after the combination of CLD and severe TCP; (3) Patients who need invasive procedures present higher bleeding rates and higher number of bleeding events in comparison with those patients who don't; (4) Platelet transfusions are still the most common bleeding prophylaxis used and (5) CLD and severe TCP patients consume a signifi-cant number of healthcare resources after the combined diagnosis.
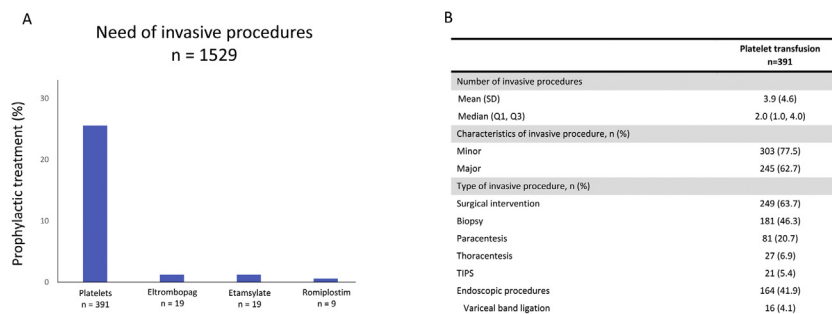
Figure 3 Bleeding prophylactic treatment in included patients who needed invasive procedures during the follow up (A). B shows the characteristics of invasive procedures in patients who received platelet transfusions as a bleeding prophylactic treatment.

To date, there is a lack of scientific evidence and publications, as well as clinical practice guidelines or recommendations on CLD associated severe TCP. In this regard, a previous work based on a Delphi study was recently published which gave rise to estimates for several parameters in this population.[32] However, in the present study we described the clinical characteristics of patients with CLD-associated severe TCP in Spain in a real-world setting, shedding light on the real clinical practice. The lack of prior published data or patient registries on this topic and our use of AI techniques such as NLP and machine learning makes this study a first of its kind. We found that CLD patients who developed severe TCP were mainly young men that were alcohol users, where the most frequent comorbidity observed was arterial hypertension. The most detected cause of CLD in this study was chronic viral hepatitis, followed by hemochromatosis, NASH and alcoholic origin, all them previously described as frequent etiologies for CLD.[33] Moreover, a higher percentage of included patients presented some CLD complication at baseline, most frequently a cirrhosis stage as previously described, and a small percentage had been subjected to TIPS during the year before.[34] All this could suggest an advanced state of the CLD when severe TCP occurs.

Our results showed that most patients with CLD-associated severe TCP underwent invasive procedures during the follow up period. This is according to previous reports that show that CLD ranges from stable phases, which are gradually followed by complications, to decompensated phases, where patients usually require an increasing number of invasive and potentially hemorrhagic interventions.[34] Interestingly, we found a median of 4 (2–7) invasive procedures per patient throughout the 5-year study period, equivalent to approximately one procedure per year, as estimated in the Delphi study.[32] Most interventions in these patients are planned, and patients in our study received prophylactic platelet transfusions in almost 30% of cases. Despite this fact, we detected that the percentage of patients with bleeding events and the number of events were higher in patients with invasive procedures within 7 days post-procedure. This is also in accordance with clinical guidelines, which describe a paucity of data to recommend a ''safe'' minimum platelet number for procedures in patients with CLD and whether platelet transfusion is effective.[35] In fact, there is no clear evidence regarding the efficacy and safety of platelet transfusion under these circumstances and increments may be poor and short-lived in patients with portal hypertension.[35]

On the contrary, more recent studies of TPO-RA have been shown to reduce the need for platelet transfusion in patients with cirrhosis requiring elective invasive procedures. However, as previously described in the Delphi study, we found that only about 3% of patients received TPO-RA drugs.[32] Despite the evidence, the low rate of use of these drugs could be related to their recent approval by regulatory agencies. Because of this, these drugs were used off-label and most still sparingly during the study period.[36–38] It would be beneficial to establish risk-adapted regimens to reduce the use of transfusions and their associated complications.[39]

Finally, we found that CLD patients with associated severe TCP were frequently admitted to the hospital during the follow up period, where 14% of cases were due to bleeding events. It can represent a meaningful economic burden for the National Health Service, in addition to causing social problems and an important impact on these patients' quality of life.[40] As expected, the accident and emergency department, gastroenterology, and hematology were the most frequently visited departments by this group of patients.

Strengths and limitations

Automated data abstraction from EHR through NLP is highly accurate, faster and on a larger scale than manual abstraction. It can facilitate real-world evidence studies at a greater scale than what could be achieved with manual data extraction. Moreover, unlike traditional randomized trials and registries which include patients that are non-representative of the target population due to strict eligibility criteria, real world data can better address storing data from daily clinical practice. In this regard, the use of NLP and ML technologies allows for direct access to real-world evidence and large sample sizes. However, results must be interpreted considering several limitations related to the study methods. First, as with other observational and retrospective studies, causal inferences must be carefully interpreted. Second, we obtained a low recall metrics for the variables ''thrombocytopenia'' and ''platelet transfusion'' but it does not affect the quality of the result shown here, for precision metrics are within range. Third, raw data used for this study are free text in EHRs; therefore, the quality of the results for some variables depends directly on the quality of the information written in EHRs. This can

be the case for variables of interest to the study that are not reflected in all EHRs, precluding their analysis. Besides, it could explain an underestimation of some results included in the present study. Finally, specific invasive procedures as well as some CLD related variables such as severity indexes scores, were not well detailed or appeared in a low percentage within the free text of the EHRs. Because that, no conclusions can be drawn regarding bleeding events and invasive procedures depending on their risk of bleeding nor the need of invasive procedures regarding CLD severity.

Conclusions

In conclusion, this study presents the current situation of CLD patients developing severe TCP based on real-world data analysis. Young male alcohol users with complications related to CLD seem to be the most affected population. They usually and periodically require invasive procedures, which are associated with bleeding events, even when prophylactic treatment with platelet transfusions is performed. Alternative methods to improve platelet counts such as TPO-RA drugs in patients with CLD-associated severe TCP scheduled to undergo invasive procedures could help to reduce dependence on platelet transfusions, avoiding associated adverse events. This scenario implies a burden for the healthcare system during the years after diagnosis that could have important economic and social consequences, given the young population affected. Standardizing management of CLD-associated severe TCP patients and inclusion of specific recommendations in clinical practice guidelines would be necessary to improve outcomes.

Authors' contributions

Conception and design: J.L. Calleja-Panero, J. Gonzalez Calvo, L. García.
Administrative support: L. García.
Collection and assembly of data: Savana Research Group.
Data analysis and interpretation: All authors.
Manuscript writing: All authors.
Final approval of manuscript: All authors.
Accountable for all aspects of the work: All authors.

Conflicts of interest

Dr. Calleja-Panero has served as a speaker for Shionogi, SOBI, AbbVie, Gilead Sciences, and Roche, and as an advisor for COI. Dr. Romero-Gómez has served as scientific advisor and speaker for Allergan, Alpha-sigma, AbbVie, BMS, MSD, Janssen-Cilag, Intercept, Genfit, NovoNordisk, Gilead Sciences, Prosciento, Kaleido, Boehringer-Ingelheim, Pfizer, SOBI, Shionogi, and Zydus. He has also been a scientific advisor in diagnostic methods for Siemens, Genfit, and Rubió, as well as co-owner of DeMILI (a non-invasive MR-based method for NASH diagnostic). He has also received grants from AbbVie, Gilead, and Intercept. Dr. Jarque has served as a scientific advisor and speaker for AbbVie, Alexion, Amgen, AstraZeneca, Gilead, Janssen, Novartis, and Takeda. He has been a speaker for Servier and scientific advisor for Apellis, Beigene, Eusa, Grifols, Incyte, Shionogi, and Sobi.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gastrohep.2023.05.010.

References

1. Giannini E, Botta F, Borro P, Malfatti F, Fumagalli A, Testa E, et al. Relationship between thrombopoietin serum levels and liver function in patients with chronic liver disease related to hepatitis C virus infection. Am J Gastroenterol. 2003;98:2516–20, http://dx.doi.org/10.1111/j.1572-0241.2003.08665.x.
2. Giannini EG. Review article: thrombocytopenia in chronic liver disease and pharmacologic treatment options. Aliment Pharmacol Ther. 2006;23:1055–65, http://dx.doi.org/10.1111/j.1365-2036.2006.02889.x.
3. Afdhal N, McHutchison J, Brown R, Jacobson I, Manns M, Poordad F, et al. Thrombocytopenia associated with chronic liver disease. J Hepatol. 2008;48:1000–7, http://dx.doi.org/10.1016/j.jhep.2008.03.009.
4. Gangireddy VG, Kanneganti PC, Sridhar S, Talla S, Coleman T. Management of thrombocytopenia in advanced liver disease. Can J Gastroenterol Hepatol. 2014;28:558–64.
5. Heyman MR, Schiffer CA. Platelet transfusion therapy for the cancer patient. Semin Oncol. 1990;17:198–209.
6. George JN. Idiopathic thrombocytopenic purpura: current issues for pathogenesis, diagnosis, and management in children and adults. Curr Hematol Rep. 2003;2:381–7.
7. Dienstag JL, McHutchison JG. American Gastroenterological Association technical review on the management of hepatitis C. Gastroenterology. 2006;130:231–64, http://dx.doi.org/10.1053/j.gastro.2005.11.010, quiz 214–7.
8. De Gottardi A, Thevenot T, Spahr L, Morard I, Bresson-Hadni S, Torres F, et al. Risk of complications after abdominal paracentesis in cirrhotic patients: a prospective study. Clin Gastroenterol Hepatol. 2009;7:906–9, http://dx.doi.org/10.1016/j.cgh.2009.05.004.
9. Seeff LB, Everson GT, Morgan TR, Curto TM, Lee WM, Ghany MG, et al. Complication rate of percutaneous liver biopsies among persons with advanced chronic liver disease in the HALT-C trial. Clin Gastroenterol Hepatol. 2010;8:877–83, http://dx.doi.org/10.1016/j.cgh.2010.03.025.
10. Friedman LS. Controversies in liver biopsy: who, where, when, how, why? Curr Gastroenterol Rep. 2004;6:30–6.
11. Stroncek DF, Rebulla P. Platelet transfusions. Lancet. 2007;370:427–38, http://dx.doi.org/10.1016/s0140-6736(07)61198-2.
12. Ning S, Barty R, Liu Y, Heddle NM, Rochwerg B, Arnold DM. Platelet transfusion practices in the ICU: data from a large transfusion registry. Chest. 2016;150:516–23, http://dx.doi.org/10.1016/j.chest.2016.04.004.

13. Perrotta PL, Snyder EL. Non-infectious complications of transfusion therapy. Blood Rev. 2001;15:69–83, http://dx.doi.org/10.1054/blre.2001.0151.
14. Slichter SJ. Evidence-based platelet transfusion guidelines. Hematol Am Soc Hematol Educ Progr. 2007:172–8, http://dx.doi.org/10.1182/asheducation-2007.1.172.
15. Poordad F. Review article: thrombocytopenia in chronic liver disease. Aliment Pharmacol Ther. 2007;26 Suppl 1:5–11, http://dx.doi.org/10.1111/j.1365-2036.2007.03510.x.
16. Amgen Europe BV. Nplate (romiplostim) powder for solution for injection: summary of product characteristics; 2009 [8.4.20].
17. Novartis Europharm Limited. Revolade (eltrombopag) film-coated tablets: summary of product characteristics. 2010. [8.4.20].
18. Mitchell O, Feldman DM, Diakow M, Sigal SH. The pathophysiology of thrombocytopenia in chronic liver disease. Hepat Med. 2016;8:39–50, http://dx.doi.org/10.2147/hmer.S74612.
19. Peck-Radosavljevic M. Thrombocytopenia in chronic liver disease. Liver Int. 2017;37:778–93, http://dx.doi.org/10.1111/liv.13317.
20. Adinolfi LE, Giordano MG, Andreana A, Tripodi MF, Utili R, Cesaro G, et al. Hepatic fibrosis plays a central role in the pathogenesis of thrombocytopenia in patients with chronic viral hepatitis. Br J Haematol. 2001;113:590–5.
21. Peck-Radosavljevic M. Thrombocytopenia in liver disease. Can J Gastroenterol. 2000;14 Suppl. D, p60d-66d.
22. Izquierdo JL, Morena D, Gonzalez Y, Paredero JM, Perez B, Graziani D, et al. Clinical management of COPD in a real-world setting. A big data analysis. Arch Bronconeumol (Engl Ed). 2021;57:94–100, http://dx.doi.org/10.1016/j.arbres.2019.12.025.
23. Izquierdo JL, Almonacid C, González Y, Del Rio-Bermúdez C, Ancochea J, Cárdenas R, et al. The impact of COVID-19 on patients with asthma. Eur Respir J. 2021, http://dx.doi.org/10.1183/13993003.03142-2020. p. 2003142.
24. Ancochea J, Izquierdo JL, Medrano IH, Porras A, Serrano M, Lumbreras S, et al. Evidence of gender differences in the diagnosis and management of COVID-19 patients: an analysis of Electronic Health Records using Natural Language Processing and machine learning. J Womens Health (Larchmt). 2021;30:393–404, http://dx.doi.org/10.1089/jwh.2020.8721.
25. Izquierdo J, Ancochea J, Savana COVID-19 Research Group, Soriano J. Clinical characteristics and prognostic factors for intensive care unit admission of patients with COVID-19: retrospective study using machine learning and natural language processing. J Med Internet Res. 2020;22:e21801, http://dx.doi.org/10.2196/21801.
26. Graziani D, Soriano JB, Del Rio-Bermudez C, Morena D, Díaz T, Castillo M, et al. Characteristics and prognosis of COVID-19 in patients with COPD. J Clin Med. 2020;9:3259.
27. Montoto C, Gisbert JP, Guerra I, Plaza R, Pajares Villarroya R, Moreno Almazán L, et al. Evaluation of natural language processing for the identification of Crohn disease-related variables in Spanish electronic health records: a validation study for the PREMONITION-CD project. JMIR Med Inform. 2022;10:e30345, http://dx.doi.org/10.2196/30345.
28. Gomollón F, Gisbert JP, Guerra I, Plaza R, Pajares Villarroya R, Moreno Almazán L, et al. Clinical characteristics and prognostic factors for Crohn's disease relapses using natural language processing and machine learning: a pilot study. Eur J Gastroenterol Hepatol. 2022;34:389–97, http://dx.doi.org/10.1097/MEG.0000000000002317.
29. Canales L, Menke S, Marchesseau S, D'Agostino A, Del Rio-Bermudez C, Taberna M, et al. Assessing the performance of clinical natural language processing systems: development of an evaluation methodology. JMIR Med Inform. 2021;9:e20492, http://dx.doi.org/10.2196/20492.
30. Bethard SD, Savova L, Pustejovsky G, Verhagen JM. SemEval-2015 Task 6: Clinical TempEval. Proceedings of the 9th International Workshop on Semantic Evaluation. Association for Computational Linguistics. 2015.
31. Bethard SS, Chen G, Derczynski W-TL, Pustejovsky J, Verhagen M. SemEval-2016 Task 12: Clinical TempEval. Proceedings of the 10th International Workshop on Semantic Evaluation. Association for Computational Linguistics. 2016.
32. Calleja-Panero JL, Andrade RJ, Banares R, Crespo J, Esteban R, Jarque I, et al. Management of chronic liver disease-associated severe thrombocytopenia in Spain: a view from the experts. Rev Esp Enferm Dig. 2020;112:778–83, http://dx.doi.org/10.17235/reed.2020.6895/2020.
33. Sharma A, Nagalli S. Chronic liver disease, in: StatPearls. 2021: Treasure Island (FL).
34. Garcia-Tsao G, Friedman S, Iredale J, Pinzani M. Now there are many (stages) where before there was one: in search of a pathophysiological classification of cirrhosis. Hepatology. 2010;51:1445–9, http://dx.doi.org/10.1002/hep.23478.
35. Neuberger J, Patel J, Caldwell H, Davies S, Hebditch V, Hollywood C, et al. Guidelines on the use of liver biopsy in clinical practice from the British Society of Gastroenterology, the Royal College of Radiologists and the Royal College of Pathology. Gut. 2020;69:1382–403, http://dx.doi.org/10.1136/gutjnl-2020-321299.
36. Kim ES. Lusutrombopag: first global approval. Drugs. 2016;76:155–8, http://dx.doi.org/10.1007/s40265-015-0525-4.
37. Hidaka H, Kurosaki M, Tanaka H, Kudo M, Abiru S, Igura T, et al. Lusutrombopag reduces need for platelet transfusion in patients with thrombocytopenia undergoing invasive procedures. Clin Gastroenterol Hepatol. 2018, http://dx.doi.org/10.1016/j.cgh.2018.11.047.
38. Peck-Radosavljevic M, Simon K, Iacobellis A, Hassanein T, Kayali Z, Tran A, et al. Lusutrombopag for the treatment of thrombocytopenia in patients with chronic liver disease undergoing invasive procedures (L-PLUS 2). Hepatology. 2019;70:1336–48, http://dx.doi.org/10.1002/hep.30561.
39. Scharf RE. Thrombocytopenia and hemostatic changes in acute and chronic liver disease: pathophysiology. Clinical and laboratory features, and management. J Clin Med. 2021;10, http://dx.doi.org/10.3390/jcm10071530.
40. Solé C, Solà E. Update on acute-on-chronic liver failure. Gastroenterol Hepatol. 2018;41:43–53, http://dx.doi.org/10.1016/j.gastrohep.2017.05.012.