Analysis of Structured and Unstructured Data From a Large Electronic Health Record System Identifies Frequent Pruritus and Fatigue in Patients With PSC

Poster SAT-403

Anthony Perez,[1] Ashley Jaksa,[1] Sarah Sidhu,[2] Douglas B. Mogul,[2] Shannon Vandriel,[2] April LaRow,[2] Michael Fried[1]

[1]Target RWE, Durham, North Carolina, United States; [2]Mirum Pharmaceuticals, Inc., Foster City, California, United States

Introduction

- Primary sclerosing cholangitis (PSC) is a progressive, autoimmune, cholestatic liver disease characterised by scarring and destruction of intrahepatic and extrahepatic bile ducts that can lead to fibrosis, cirrhosis, hepatocellular carcinoma, or bile duct cancer.[1-3]
  - Key clinical manifestations include fatigue, pruritus, difficulty sleeping, brain fog, anxiety, pain, and jaundice.[1,2,4,5]
- Pruritus and fatigue may occur in patients with PSC and can lead to a reduced quality of life, but reported prevalence varies widely across studies.[1,4]
  - Pruritus is present in up to 38%-91% of patients with PSC and can be debilitating.[4,6]
  - Fatigue has been reported in 22%-71% of patients with PSC and can follow a persistent pattern in daily life.[6,7]
- While most physicians report assessing pruritus in patients with PSC, there is a consensus that standardised tools are needed to improve identification, capture, and assessment of pruritus.[8]
- TARGET-LD and TARGET-GASTRO are longitudinal observational electronic health record (EHR) databases that capture real-world clinical data for >800,000 patients with liver and gastrointestinal diseases from seven healthcare networks (HCNs) across the United States. Data are captured via waiver of informed consent; up to five years of historical EHR data are available.
  - All HCNs contributed structured data, with three also contributing unstructured data.

Objective

- To assess dataset robustness and frequency of patient-reported pruritus and fatigue in patients with PSC using structured and unstructured data from TARGET-LD and TARGET-GASTRO.

Methods

TARGET-LD and TARGET-GASTRO Network (N=856,254)

↓

Patients with PSC were identified based on ≥1 PSC ICD-10 code (K83.1) or PSC listed as diagnosis[a,b] (N=3,958)

↓

Analysis of EHR data

Structured data from 7/7 HCNs (All HCNs; N=3,958)

Descriptive analysis of:

- Patient demographics
- Prevalence of ICD-10 codes for pruritus and fatigue
- Patients with relevant labs or medications
- Patients with evidence of MRCP testing

↓

Unstructured data from 3/7 HCNs (UD HCNs; n=1,427)

Analysis using a purpose-built native language processing system to 'sweep' unstructured EHR text to extract key terms and synonyms related to[c]:

- Pruritus
- Fatigue
- MRCP

[a]Patients selected by ICD-10 code are automatically included under an institutional review board (IRB)-approved waiver of informed consent via an established, automated interface with participating sites' EHRs. [b]The index date is first ICD-10 code for PSC. [c]The models distinguished between mentions that indicate a patient symptom versus clinical exclusion of that symptom through a scoring system based on the probability of the recognised entity being true.

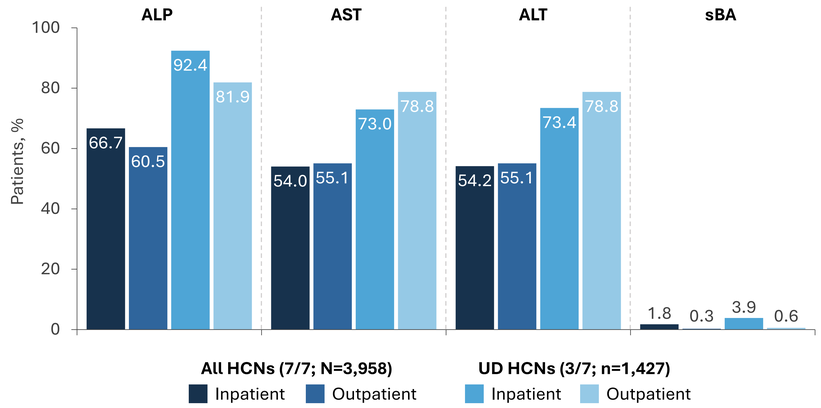
Results

Table 1. Patient Demographics

Parameter	All HCNs (7/7; N=3,958)	UD HCNs (3/7; n=1,427)
Age, n (%)[a]		
18 to <45 years	1,169 (29.5)	493 (34.6)
45 to 64 years	1,258 (31.8)	525 (36.8)
≥65 years	1,528 (38.6)	408 (28.6)
Sex, n (%)[a]		
Male	2,128 (53.8)	760 (53.3)
Age at first PSC code, years, mean (SD)	54.0 (17.7)	53.0 (17.3)

[a]Across both cohorts, a total of four patients (n=3 in all HCNs; n=1 for UD HCNs) were missing age information and one patient (n=1 in all HCNs) was missing sex information.

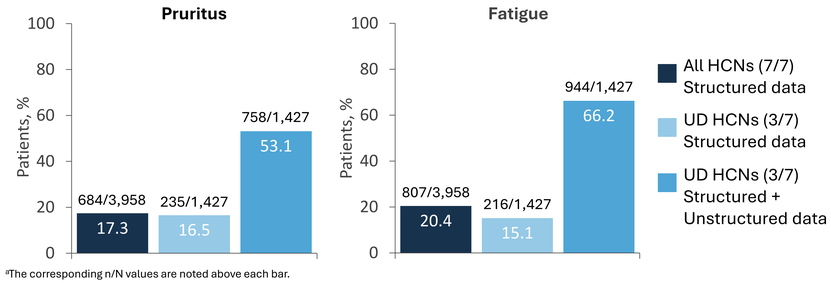
Figure 1. Percentage of Patients With ≥1 Labs in Both Inpatient and Outpatient Settings, Stratified by the Use of Structured or Both Structured and Unstructured Data



- More than 2,600 patients had ≥1 lab result for each major liver enzyme.
  - In this cohort of patients with PSC, sBA testing was rarely performed.
- Inclusion of unstructured data revealed higher rates of testing for all major liver enzymes in both the inpatient and outpatient settings compared to structured data alone.
- Over a third of patients had outpatient labs performed ≥4 times.

Improvements in Identification of Pruritus and Fatigue Were Observed When Including Unstructured Data

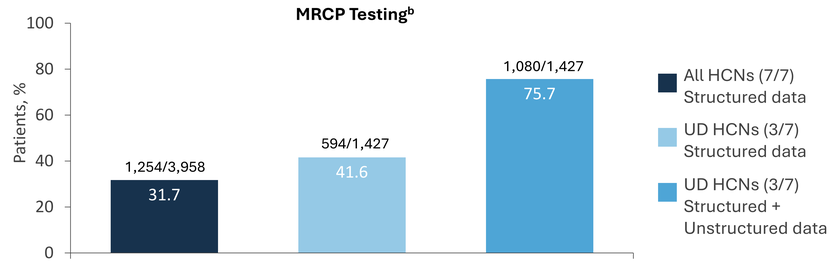
Figure 2. Comparison of Pruritus and Fatigue Symptoms Captured Through Analysis of Structured vs Unstructured Data[a]



[a]The corresponding n/N values are noted above each bar.

Improvements in Identification of MRCP Procedures Were Observed When Including Unstructured Data

Figure 3. Comparison of MRCP Procedures Captured Through Analysis of Structured vs Unstructured Data[a]



[a]The corresponding n/N values are noted above each bar. [b]Procedure searched based on keywords due to lack of CPT codes in data.

- Inclusion of unstructured data increased prevalence estimates of procedures and key symptoms relative to traditional code-based enumeration alone.
  - Both pruritus and fatigue were identified at a greater frequency with the addition of unstructured data.
  - MRCP testing was identified in twice as many patients in unstructured data compared to structured data.

Conclusions

- TARGET-LD and TARGET-GASTRO contain one of the largest cohorts of patients with PSC and include robust data on laboratory values and imaging.
- While pruritus and fatigue were infrequently coded, additional analysis of unstructured data markedly increased their identification and indicated these are important, prevalent symptoms in PSC.
- Use of both structured and unstructured elements may provide a new opportunity to understand the natural history of this rare disease.

Abbreviations

ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CPT, current procedural terminology; EHR, electronic health record; HCN, healthcare network; ICD-10, International Classification of Diseases, 10th Revision; IRB, institutional review board; MRCP, magnetic resonance cholangiopancreatography; PSC, primary sclerosing cholangitis; sBA, serum bile acid; UD HCNs, healthcare networks contributing both structured and unstructured data.

Disclosures

AP and MF are employees of Target RWE. AJ is an employee of Target RWE and holds stock in CVS Health. SS, DBM, SV, and AL are employees of and shareholders in Mirum Pharmaceuticals, Inc.

Acknowledgements

The authors would like to thank the physicians and patients for their participation in this study. This analysis was funded by Mirum Pharmaceuticals, Inc. Medical writing and editorial support for the development of this poster was provided by Precision AQ in Bethesda, Maryland, and was funded by Mirum Pharmaceuticals, Inc.

References

1. Karlsen TH, et al. *J Hepatol*. 2017;67:1298-1323. **2.** Eaton JE, et al. *Gastroenterology*. 2013;145:521-536. **3.** Fung BM, et al. *World J Gastroenterol*. 2019;25:659-671. **4.** Kowdley KV, et al. *Hepatol Commun*. 2026;10:e00904. **5.** PSC Partners Seeking a Cure. Pediatric PSC. https://pscpartners.org/patients-caregivers/living-with-psc/voice-of-the-patient-report-pfdd-forum.html Accessed April 22, 2026. **6.** van Munster KN, et al. *Liver Int*. 2022;42:1562-1570. **7.** Kruk B, et al. *Eur J Clin Invest*. 2025;55:e70041. **8.** Mayo MJ, et al. Presented at AASLD 2025.

Presented at European Association for the Study of the Liver (EASL) Congress; 27-30 May 2026; Barcelona, Spain