MSR10

Incidence, Prevalence, and Clinical Characteristics of Patients With Chronic Lymphocytic Leukemia in Spain: Natural Language Processing Analysis of Electronic Health Records

José Ángel Hernández Rivas,[1] Nasim Bahar,[2] Cristina Bas,[3] Alberto Prieto,[2] Macarena Ortiz,[4] Eva Castillo,[5] Antonio Gutiérrez[6]

[1]University Hospital Infanta Leonor, Madrid, Spain; [2]BeiGene Switzerland GmbH, Basel, Switzerland; [3]BeiGene ESP, SL, Madrid, Spain; [4]Regional University Hospital Malaga, Malaga, Spain; [5]Savana Research, Madrid, Spain; [6]Lymphoma Unit, Department of Hematology, Son Espases University Hospital/IdISBa, Palma, Spain

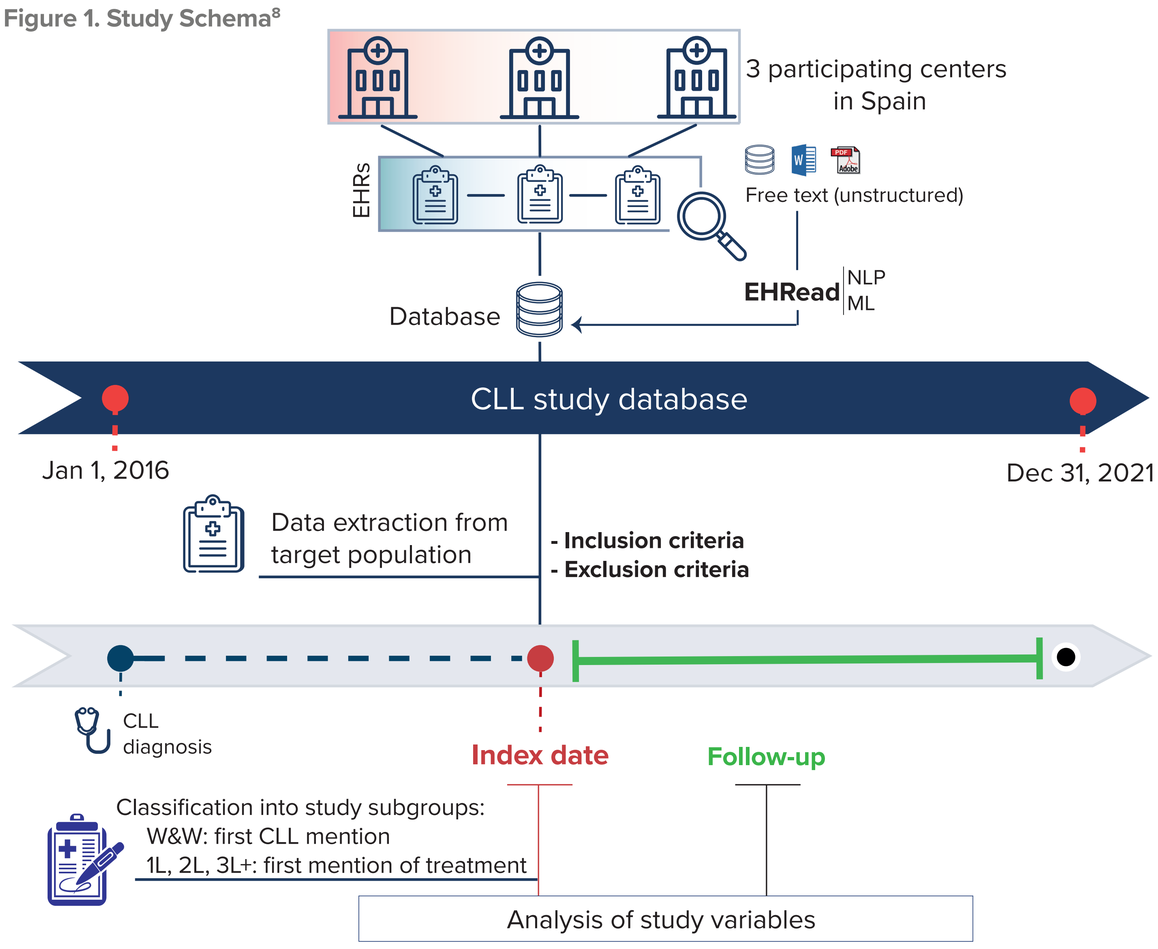
INTRODUCTION

- Chronic lymphocytic leukemia (CLL) is the most common leukemia in the Western world,[1] but a significant knowledge gap regarding the epidemiology of this disease prevents the optimal allocation of necessary resources for treating these patients
- The incidence of CLL, as reported in the literature, shows significant variation across age groups and genders, ranging from 3.6 to 6.9 cases per 100,000.[3,4] Similarly, the prevalence of CLL is estimated to be between 48 and 52 cases per 100,000 person-years[5-7]
- Unlike traditional registries or real-world studies based on claims, natural language processing (NLP) of electronic health records (EHRs) allows for analysis of heterogeneous datasets, mitigating selection bias and yielding precise patient data
- In this study, we used free-text data extracted with NLP from EHRs to determine the incidence, prevalence, and primary clinical characteristics of CLL in Spanish patients

METHODS

- **Study design:** This was an observational, multicenter, retrospective study based on secondary use of EHR data
- **Study population and study period:** This study included all adult patients with a CLL diagnosis between January 1, 2016, and December 31, 2021, at 3 Spanish hospitals (Hospital Universitari Son Espases, Mallorca; Hospital Regional Universitario Carlos Haya, Málaga; Hospital Universitario Infanta Leonor, Madrid). Patients diagnosed with Richter syndrome or prolymphocytic leukemia were excluded
- **Methods:** EHRs were evaluated using NLP based on clinical terminology (SNOMED-CT) and machine learning. Unstructured clinical information related to clinical characteristics and treatment was assessed; 205 clinical variables were extracted and summarized at the patient level using descriptive statistics (**Figure 1**)
- **Data privacy:** Data ownership remained with the participant sites. EHR data were anonymized before analysis, in compliance with corresponding data protection and privacy laws

Figure 1. Study Schema[8]



This study used EHRead,[2] an innovative technology that applies NLP and ML to extract, organize, and analyze the unstructured clinical information jotted down by health professionals in patient EHRs from the 3 participating hospitals.
1L, first-line; 2L, second-line; 3L, third-line; CLL, chronic lymphocytic leukemia; EHR, electronic health record; ML, machine learning; NLP, natural language processing; W&W, watch and wait.

RESULTS

- A total of 697 patients with CLL were included in the study, out of a population of 2,069,341 patients with a total of 88,872,628 EHRs
- The overall age-standardized (2013 European population) incidence and prevalence for the study period were 3.37 (95% CI, 2.54-4.1922) and 49.59 (95% CI, 47.4-51.77) cases per 100,000 person-years, respectively (**Table 1**)

Table 1. Age-Adjusted Prevalence and Incidence of CLL Based on the European Standard Population (per 100,000 Person-Years)

	Adjusted Incidence (95% CI)			Adjusted Prevalence (95% CI)		
	Overall	Male	Female	Overall	Male	Female
Age category						
All	3.37 (2.54-4.19)	4.53 (3.60-5.47)	2.54 (1.73-3.34)	49.59 (47.40-51.77)	65.02 (61.78-68.25)	37.76 (36.20-39.32)
20-49 years	0.18 (0.09-0.27)	0.20 (0.10-0.29)	0.17 (0.08-0.25)	1.88 (1.81-1.96)	2.54 (2.38-2.71)	1.25 (1.21-1.29)
50-69 years	1.14 (0.82-1.45)	1.51 (1.19-1.82)	0.79 (0.46-1.13)	14.31 (13.32-15.31)	18.05 (17.06-19.05)	10.90 (9.87-11.94)
70-79 years	0.97 (0.60-1.33)	1.13 (0.78-1.49)	0.83 (0.42-1.23)	16.36 (15.45-17.26)	21.63 (20.54-22.71)	12.00 (11.01-12.98)
80-89 years	0.91 (0.69-1.13)	1.28 (0.88-1.68)	0.67 (0.42-0.93)	14.11 (12.76-15.47)	18.52 (16.20-20.83)	11.28 (10.46-12.10)
≥90 years	0.18 (0.14-0.22)	0.42 (0.23-0.60)	0.08 (0-0.15)	2.92 (2.00-3.83)	4.28 (2.84-5.72)	2.33 (1.61-3.05)

CONCLUSIONS

- This study presents comprehensive information on patients with CLL in Spain, obtained through NLP analysis of EHRs
- The estimated incidence of CLL highly varies across age groups and sex, as previously reported in the literature[3,4]
- EHRs do not systematically capture some variables of interest across all patients, which reduces the sample size and may be a source of selection bias
- Artificial intelligence and NLP may be reliable methods for studying some aspects of epidemiology, but results should be interpreted with caution owing to potential limitations in NLP or in the source of information[9]

- The mean age was 72.10 years (SD, 12.60 years), and 299 patients (42.90%) were female (**Table 2**)

Table 2. Demographics at Index Date

	Patients With CLL (N=697)
Age, years	
Mean (SD)	72.10 (12.60)
Median (Q1, Q3)	73 (65, 82)
Sex, n (%)	
Male	398 (57.10)
Female	299 (42.90)

Q, quartile.

- The most common comorbidities were arterial hypertension (64.56%), diabetes (33.00%), chronic obstructive pulmonary disease (22.67%), obesity/overweight (20.95%), and renal insufficiency (20.66%) (**Table 3**)

Table 3. Clinical Comorbidities at Index Date

Comorbidity, n (%)	Patients With CLL (N=697)
Arterial hypertension	450 (64.56)
Diabetes	230 (33.00)
Chronic obstructive pulmonary disease	158 (22.67)
Obesity/overweight	146 (20.95)
Renal insufficiency (acute or chronic)	144 (20.66)
Hypercholesterolemia	126 (18.08)
Ischemic heart disease	116 (16.64)
Congestive heart failure	91 (13.06)
Bronchitis	76 (10.90)
Cerebrovascular disease	61 (8.75)
Asthma	50 (7.17)
Solid tumors	49 (7.03)
Deep vein thrombosis	39 (5.60)
Metastatic solid tumors	37 (5.31)
Emphysema	24 (3.44)
Pulmonary embolism	10 (1.43)
Peripheral vascular disease	3 (0.43)

- The most common clinical complications observed were lymphocytosis (65.85%), lymphadenopathy (35.44%), anemia (26.26%), infections (18.36%), thrombocytopenia (14.20%), and splenomegaly (13.92%) (**Table 4**)

Table 4. Clinical Complications at Index Date

Complication, n (%)	Patients With CLL (N=697)
Lymphocytosis	459 (65.85)
Lymphadenopathy	247 (35.44)
Anemia	183 (26.26)
Infections	128 (18.36)
Thrombocytopenia	99 (14.20)
Splenomegaly	97 (13.92)
Weight loss	58 (8.32)
Fatigue	40 (5.74)
Hepatomegaly	24 (3.44)
Fever of unknown origin	5 (0.72)

- Eastern Cooperative Oncology Group (ECOG) performance status was reported in 13.5% of patients; an ECOG performance status of 0 was most common (43.6%)
- Of the 697 patients identified, 406 (58.2%) underwent a watch-and-wait approach and 291 (41.8%) received treatment
- Cytogenetic alterations (del 11, 13, 17; trisomy 12; p53 or IGHV mutation) were reported in 27.5% of patients

REFERENCES

1. Eichhorst B, et al. *Ann Oncol.* 2021;32(1):23-33.
2. Canales L, et al. *JMIR Med Inform.* 2021;9(7):e20492.
3. Kipps TJ, et al. *Nat Rev Dis Primers.* 2017;3:16096.
4. Jeyakumaran D, et al. *Value Health.* 2016;19(7):PA574-A575.
5. Burn E, Català M. DARWIN EU C1-001: Prevalence of rare blood cancers in Europe. 28/03/2023, v3.2.
6. Orphanet Report Series. Prevalence and incidence of rare diseases: bibliographic data. Number 1, January 2022. Accessed October 15, 2023. https://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf
7. Mattsson M, et al. *Am J Hematol.* 2020;95(2):E36-E38.
8. Loscertales J, et al. *Cancers (Basel).* 2023;15(16):4047.
9. Del Rio-Bermudez C, et al. *J Pharm Policy Pract.* 2020;13(1):75.

DISCLOSURES

NA, CB, AP are BeiGene employees. **JAHR, MO, AG** have provided consulting services to the pharmaceutical industry.
EC is an employee of Savana Research, which was commissioned by BeiGene to apply NLP to extract the data and analyze it.

ACKNOWLEDGMENTS

This study was sponsored by BeiGene, Ltd. Editorial assistance was provided by Nucleus Global, an Inizio Company, and supported by BeiGene.

CORRESPONDENCE: José Ángel Hernández Rivas, jahr_jahr2006@yahoo.es

Presented at ISPOR Europe 2023; November 12-15, 2023; Copenhagen, Denmark