Harnessing Agentic AI for Cohort Identification: A Case Study in Patients With CLL/SLL Treated With BTK Inhibitors at a US Academic Health System

SA28

Sandeep Kumar,[1] Gnaanikko Patchiappan Pa,[1] Melissa Hagan,[2] Allison Smither[3]

[1]nference, Bangalore, India; [2]BeOne Medicines, Ltd., San Carlos, CA, USA; [3]nference, Cambridge, MA, USA

CONCLUSIONS

- The findings underscore the potential value of AI for streamlining data abstraction processes when using RWE and EHR data
- These strategies may enhance our understanding of the patient journey by improving patient cohort identification in future healthcare research initiatives

INTRODUCTION

- Unstructured data within electronic health records (EHR) are rich with insights relevant to real-world evidence (RWE) studies, yet extracting and utilizing this data at scale remains a complex and time-intensive process[1]
- Information extracted from clinical notes may provide a more complete view of the patient journey than structured medication orders and diagnosis codes alone, enabling deeper characterization of a patient's clinical journey relative to a diagnosis[2]
- Recent advancements in artificial intelligence (AI), especially in natural language processing (NLP) and large language models (LLMs), have greatly enhanced the efficiency of information extraction from clinical notes in medical records[3-4]
- Previous studies have demonstrated the feasibility and scalability of using AI methods to characterize clinical characteristics, diagnostic testing, and treatment patterns in patients with newly diagnosed chronic lymphocytic leukemia (CLL)[5]

OBJECTIVE

- To examine the utility and value of AI technologies for identifying patients diagnosed with chronic lymphocytic leukemia or small lymphocytic lymphoma (CLL/SLL) and treated with a Bruton tyrosine kinase inhibitor (BTKi) in the first-line setting using EHR data

METHODS

Patient Identification

- Patients were identified using EHR data from a large US academic healthcare system
- Patients were selected if they were diagnosed ≥2 times with CLL/SLL during CLL/SLL diagnosis period (January 1, 2005 – December 31, 2024); had initial use of BTKi between January 1, 2020, and December 31, 2024; and had first CLL diagnosis date on or prior to initial use of BTKi
- Patients were excluded if they participated in a clinical trial during the study treatment period; had <10 days of any single BTKi, or had prior BTKi use

Cohort Identification

- Once selected, patients were categorized into three unique cohorts using different but distinct AI identification strategies:
  - **International Classification of Diseases (ICD)+NLP cohort**: Applied the ICD+NLP structured ICD codes in concert with unstructured clinical notes transformed into computable variables using nference's proprietary NLP algorithms
  - **LLM cohort**: Applied an LLM-assisted analysis of unstructured data
  - **ICD+NLP+LLM cohort**: Applied a combined approach using all three methods
- The DeepSeek-R1-Distill-Qwen-14B[6] LLM model was used to identify presence of CLL/SLL, BTKi use, and clinical trial participation

Descriptive and Sensitivity Analysis

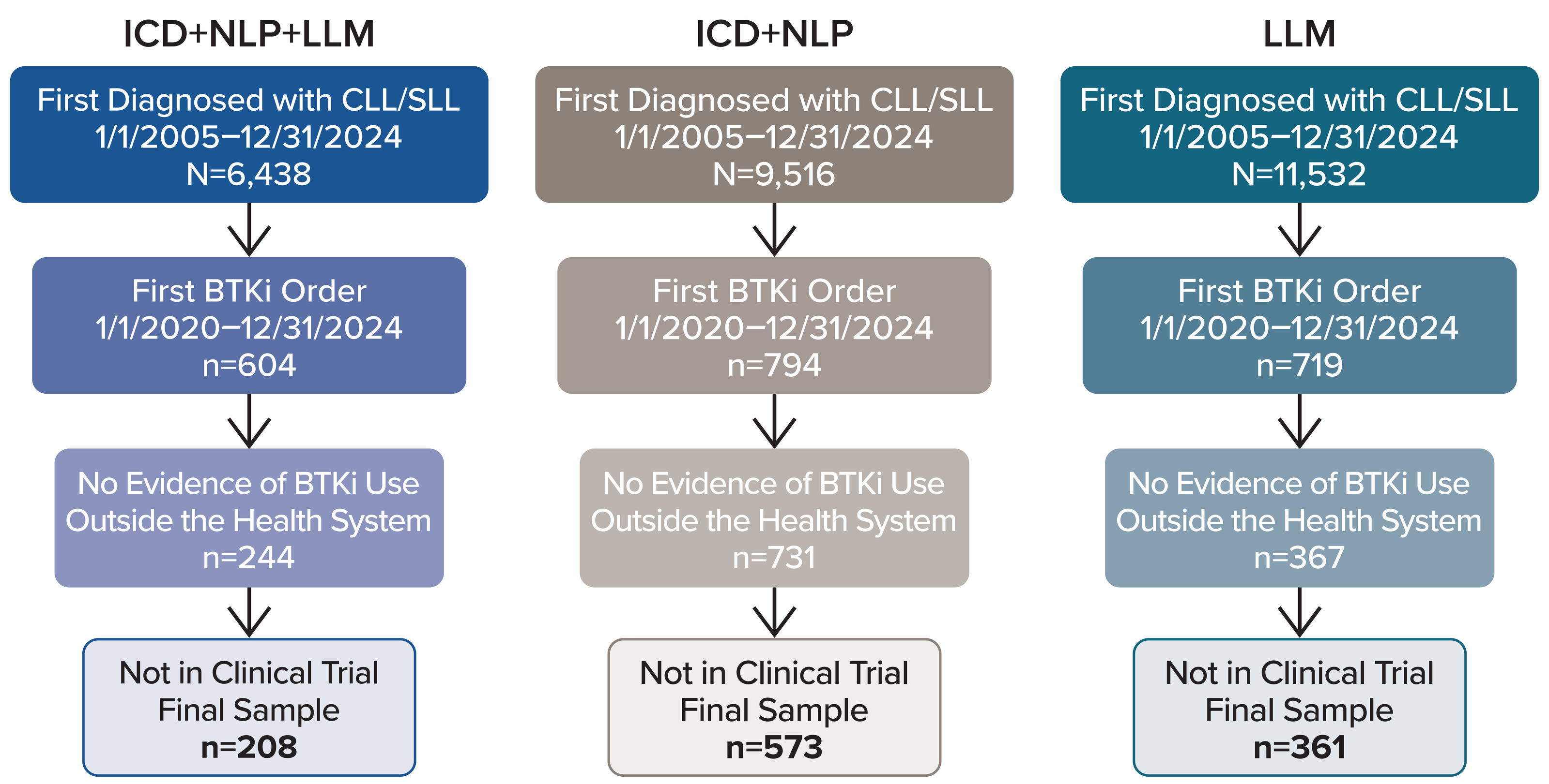
- Once patients were assigned to cohorts, demographic variables (age, race, and gender) were analyzed descriptively
- For the sensitivity analysis, 20% of patients included in the cohort and an equal number of excluded patients were chosen randomly by human oversight
- Positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, and accuracy were calculated from values of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN)
- Statistical differences were evaluated using the Pearson's Chi-Square test for independence with the Benjamini-Hochberg correction for multiple comparisons

RESULTS

Cohort Identification and Demographic Characteristics

- A total of 573 patients were identified in the ICD+NLP cohort, 361 in the LLM cohort, and 208 in the ICD+NLP+LLM cohort (**Figure 1**)
- The mean age at BTKi initiation ranged from 65.5 to 66.3 years across the three cohorts (**Table 1**)
- Demographic characteristics for each cohort demonstrated similar distributions: 65.9%-67.2% were male, 93.8%-95.0% identified as White, and 95.1%-95.8% as non-Hispanic

Figure 1. Cohort Identification by AI Strategy



AI, artificial intelligence; BTKi, Bruton tyrosine kinase inhibitor; CLL/SLL, chronic lymphocytic leukemia/small lymphocytic lymphoma; ICD, International Classification of Diseases; LLM, large language model; NLP, natural language processing.

Table 1. Demographic Characteristics by Cohort

Measure	ICD+NLP+LLM (n=208)	ICD+NLP (n=573)	LLM (n=361)
Mean Age at BTKi Initiation (SD)	66.3 (10.4)	65.9 (10.6)	65.5 (10.6)
Sex, n (%)			
Male	137 (65.9%)	365 (67.2%)	238 (65.9%)
Female	68 (34.1%)	188 (32.8%)	123 (34.1%)
Race, n (%)			
White	195 (93.8%)	538 (93.9%)	343 (95.0%)
Other[a]	13 (6.2%)	35 (6.1%)	18 (5.0%)

[a]Other includes "Black or African American", "Asian or Asian-American", "Native American or Pacific Islander", and "Unknown". Patient counts <10 were consolidated to prevent patient re-identification.

BTKi, Bruton tyrosine kinase inhibitor; ICD, International Classification of Diseases; LLM, large language model; NLP, natural language processing; SD, standard deviation.

Sensitivity Analysis

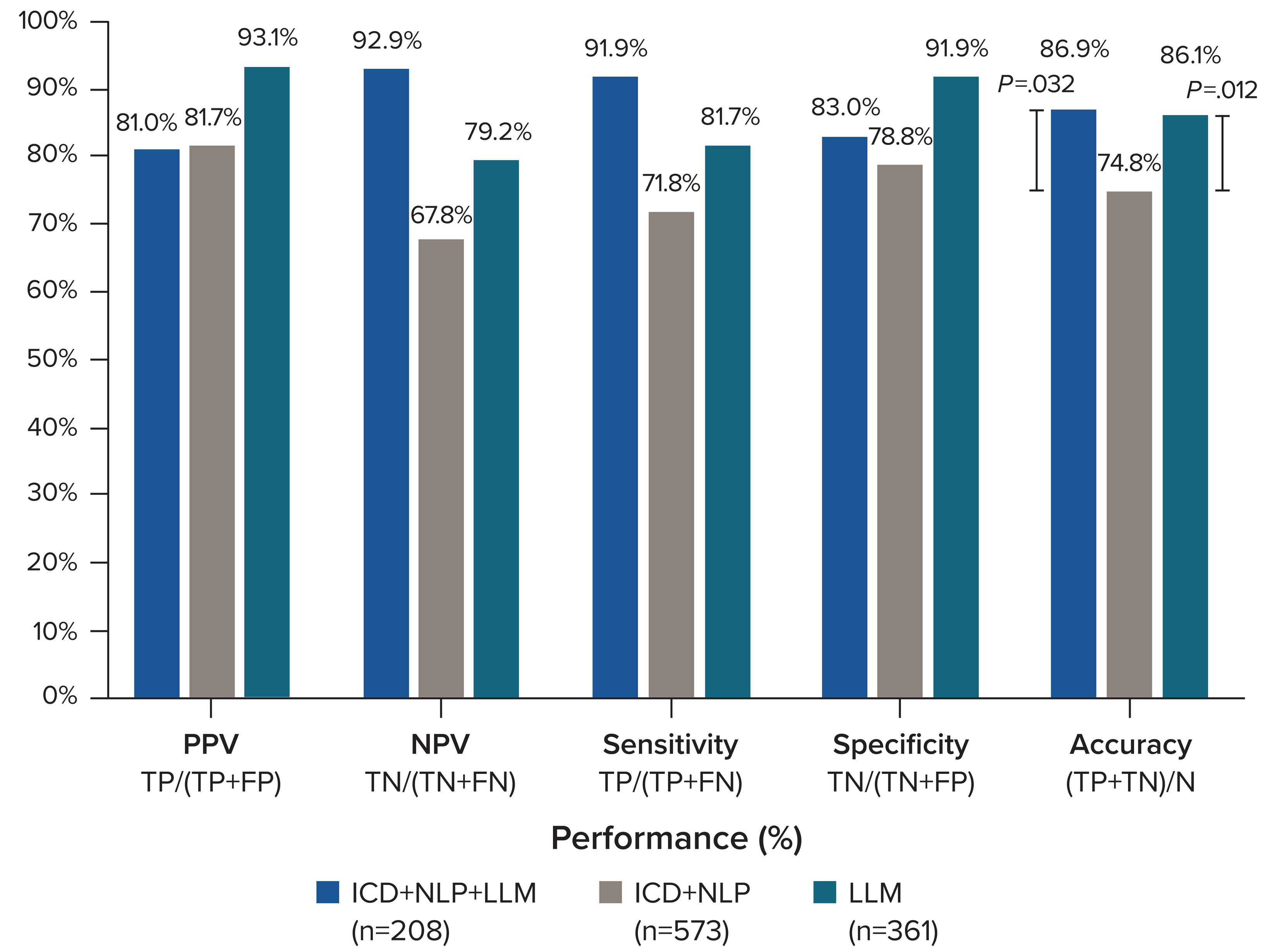
- Evaluated metrics and results of the sensitivity analysis are shown in **Table 2** and results in **Figure 2**
- The ICD+NLP+LLM cohort had the best NPV (92.9%) and sensitivity (91.9%)
- The LLM cohort exhibited the best PPV (93.1%) and specificity (91.9%)
- Both the LLM (86.1%) and ICD+NLP+LLM (86.9%) achieved significantly higher accuracy for cohort inclusion compared to the ICD+NLP cohort (74.8%) (P=.012 and P=.032, respectively)

Table 2. Sensitivity Analysis Evaluated Metrics

Measure	ICD+NLP+LLM (n=208)	ICD+NLP (n=573)	LLM (n=361)
Cases Reviewed	84	230	144
TP	34	94	67
TN	39	78	57
FP	8	21	5
FN	3	37	15

FN, false negatives; FP, false positives; ICD, International Classification of Diseases; LLM, large language model; NLP, natural language processing; TN, true negatives; TP, true positives.

Figure 2. Sensitivity Analysis Results by Cohort



FN, false negatives; FP, false positives; ICD, International Classification of Diseases; LLM, large language model; NLP, natural language processing; NPV, negative predictive value; PPV, positive predictive value; TN, true negatives; TP, true positives.

DISCUSSION

- The measures of sensitivity, specificity, and accuracy differed as a function of the cohort identification strategy employed
- The LLM cohort provided the best combination of patient counts and accuracy
- The ICD+NLP cohort had the largest size but the lowest accuracy
- The ICD+NLP+LLM cohort had the smallest size but was the most accurate with high NPV and sensitivity
- Despite the limitations of the study, these findings underscore the potential of AI in streamlining data abstraction processes, facilitating deeper insights, and improving patient cohort identification in future healthcare research initiatives

LIMITATIONS

- The performance of LLMs compared to ICD and NLP approaches may vary between health systems, with performance depending on the completeness and quality of clinical documentation within the EHR
- The analysis was limited to cohort identification and does not establish equivalent performance for other data abstraction tasks

REFERENCES

1. Kong HJ. *Healthc Inform Res*. 2019;25(1):1-2.
2. Bettini R, et al. *Am J Lifestyle Med*. 2026:15598276261423086.
3. Li I, et al. *Comp Sci Rev*. 2022;46:100511.
4. Hu Y, et al. *AMIA Annu Symp Proc*. 2025;2024:500-505.
5. Al-Agil M, et al. *Health Informatics J*. 2025;31(3):14604582251342178.
6. DeepSeek-AI, et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. Accessed March 4, 2026. https://arxiv.org/abs/2501.12948

DISCLOSURES

AS, **GPP**, and **SK**: Support for current work from nference (sourced data and medical writing assistance); Employment with nference. **MH**: Employment with BeOne Medicine.

ACKNOWLEDGMENTS

This study was funded by BeOne Medicines, Ltd.

CORRESPONDENCE: Allison Smither (asmither244@gmail.com)

Presented at ISPOR 2026; May 17-20, 2026, Philadelphia, PA, USA