Leveraging Artificial Intelligence for the Generation of Computable Operational Definitions in Electronic Health Records: Facilitating Real-World Evidence Research

Evelyn Rizzo¹, Michael Buck², Kevin Kallmas³, David Thompson⁴, Aaron W. C. Kamauu² Scan her or follow lin for e-post ¹Mobility HEOR, Akron, Ohio, USA; ²Navidence Inc., Cottonwood Heights, Utah, USA; ³Nested Knowledge St. Paul, Minnesota, USA; ⁴Rubidoux Research, Boston, Massachusetts, USA How did we perform the Al-driven search? **Objective:** Computable operational definitions (CODefs) are essential for identifying patient cohorts in **Al-driven Lit Review CODef Creation** real-world evidence (RWE) studies. However, Screening Tagging manual development of these phenotypes is time-consuming and often lacks We developed a search A tagging hierarchy Information from the literature The articles were screened standardization or validation. strategy to identify algorithms for relevance to LC and for review was used to create the identified therapeutic and for Lung Cancer (LC) the presence of CODefs for each relevant coding definitions (e.g., The objective is to evaluate the feasibility and **CODef-related terminology** (Figure 1). concept, covering different data effectiveness of using artificial intelligence (AI) ICD-10-CM, CPT, variable types and value set or validation statistics software to identify algorithms used in CODefs We executed a 'living' search SNOMED), of LC concepts lists (Figure 4) (Figure 2). in an Al-driven software for population identification in RWE research reported in the literature platform that utilizes natural Diagnoses libraries. using AI tagging language processing and • Procedures



Copies of this poster obtained through the QR code or link are for personal use only and may not be reproduced without permission.

machine learning algorithms to analyze literature from PubMed and ClinicalTrials.gov.

recommendations which were highlighted in the text for the reviewer tagging the articles (Figure 3).

Medications

• Labs

Encounters

Results: The AI-supported search returned 240 studies for screening, of which 94 were excluded for having a <.1 probability of inclusion by the AI model. Twenty-three studies were included and underwent full-text tagging with Al-driven smart tagging recommendations reviewed and applied by team members. The tagging process yielded 31 algorithms for identifying patients with LC which included three algorithms for distinguishing small cell LC and 10 for identifying non-small cell LC within the data sets with varying algorithmic accuracy. The software allowed algorithms to be downloaded to an excel sheet so CODef performance could be compared and referenced for future RWE research.

Figure 1: Al-driven Lit Review

Concepts	Preview (Finalize >)	
Lung Cancer X	v 1 (9
Population Ø	2 Lung Cancer	
O biomarker X O EGFR X O AKL X O infiltrating cancer X O invasive carcinoma X	3) 4 AND	
non-invasive lung cancer X invasive lung cancer X	v 5 (
Negate	6 biomarker	
Intervention Ø	7 OR	
○ ICD-10 × ○ SNOMED × ○ International Classification of Diseases, 10th revision × ○ ICD 10 × ○	8 EGFR 9 OR	
Negate	10 AKL	

Figure 3: Tagging

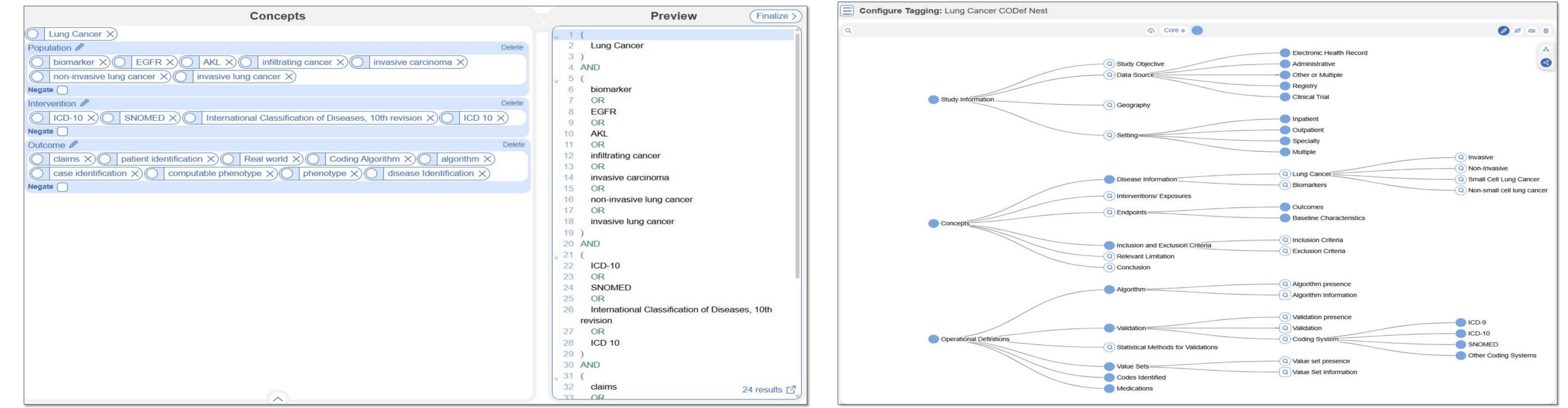


Figure 2: Screening

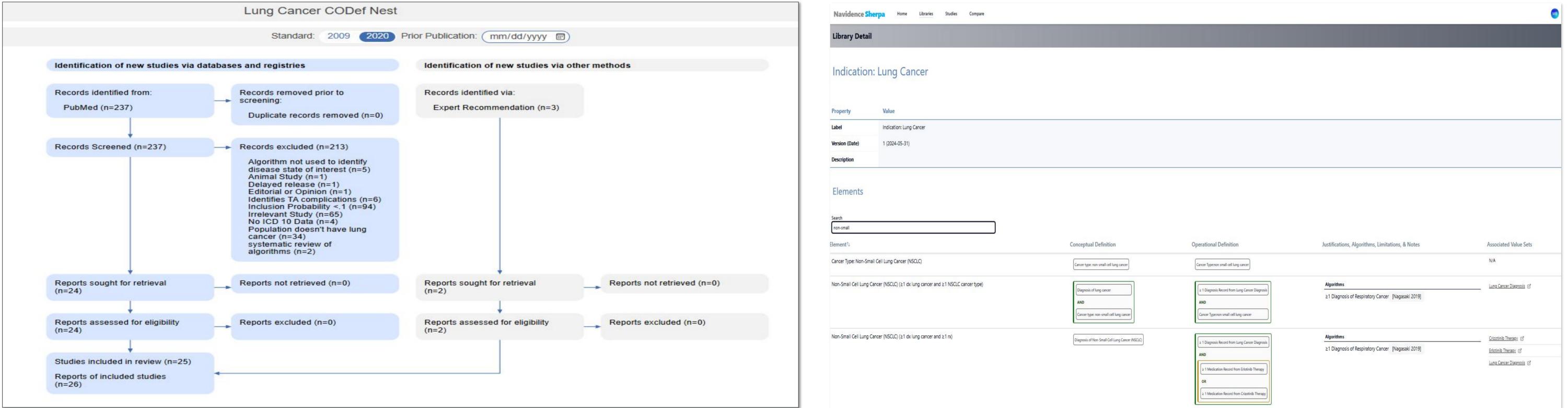


Figure 4: CODef Creation

Navidence S	herpa Home Libraries Studies Compare				MB		
Library Detail							
Indicatio	n: Lung Cancer						
Property	Value						
Label	Indication: Lung Cancer						
Version (Date)	1 (2024-05-31)						
Description							
Elements							
Search non-small							
Element↑↓		Conceptual Definition	Operational Definition	Justifications, Algorithms, Limitations, & Notes	Associated Value Sets		
Cancer Type: Non-Sr	mall Cell Lung Cancer (NSCLC)	Cancer type: non-small cell lung cancer	Cancer Typemon small cell lung cancer		N/A		
Non-Small Cell Lung	Cancer (NSCLC) (≥1 dx lung cancer and ≥1 NSCLC cancer type)			Algorithms	Lung Cancer Diagnosis 12		

Conclusion: Al-assisted identification of algorithms for CODefs is feasible and faster than reviewing articles manually. This approach has the potential to accelerate research timelines and improve reproducibility, as coding methods continue to evolve.





Presented at ISPOR Europe 2024 | 17–20 November 2024 | Barcelona, Spain **Contact:** evelyn@mobilityheor.com