Acceptance code MSR57

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

Evelyn Rizzo [1], Michael Buck [2], Kevin Kallmas [3], David Thompson [4], Aaron W. C. Kamauu [2]

[1]Mobility HEOR, Akron, Ohio, USA; [2]Navidence Inc., Cottonwood Heights, Utah, USA; [3]Nested Knowledge St. Paul, Minnesota, USA; [4]Rubidoux Research, Boston, Massachusetts, USA

Scan here or follow link for e-poster*

Objective:
Computable operational definitions (CODefs) are essential for identifying patient cohorts in real-world evidence (RWE) studies. However, manual development of these phenotypes is time-consuming and often lacks standardization or validation.

The objective is to evaluate the feasibility and effectiveness of using artificial intelligence (AI) software to identify algorithms used in CODefs for population identification in RWE research libraries.

*Online Poster

How did we perform the AI-driven search?

AI-driven Lit Review

We developed a search strategy to identify algorithms for Lung Cancer (LC) (**Figure 1**).

We executed a 'living' search in an AI-driven software platform that utilizes natural language processing and machine learning algorithms to analyze literature from PubMed and ClinicalTrials.gov.

Screening

The articles were screened for relevance to LC and for the presence of CODef-related terminology or validation statistics (**Figure 2**).

Tagging

A tagging hierarchy identified therapeutic and coding definitions (e.g., ICD-10-CM, CPT, SNOMED), of LC concepts reported in the literature using AI tagging recommendations which were highlighted in the text for the reviewer tagging the articles (**Figure 3**).

CODef Creation

Information from the literature review was used to create the CODefs for each relevant concept, covering different data variable types and value set lists (**Figure 4**)

- Diagnoses
- Procedures
- 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 AI-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: AI-driven Lit Review

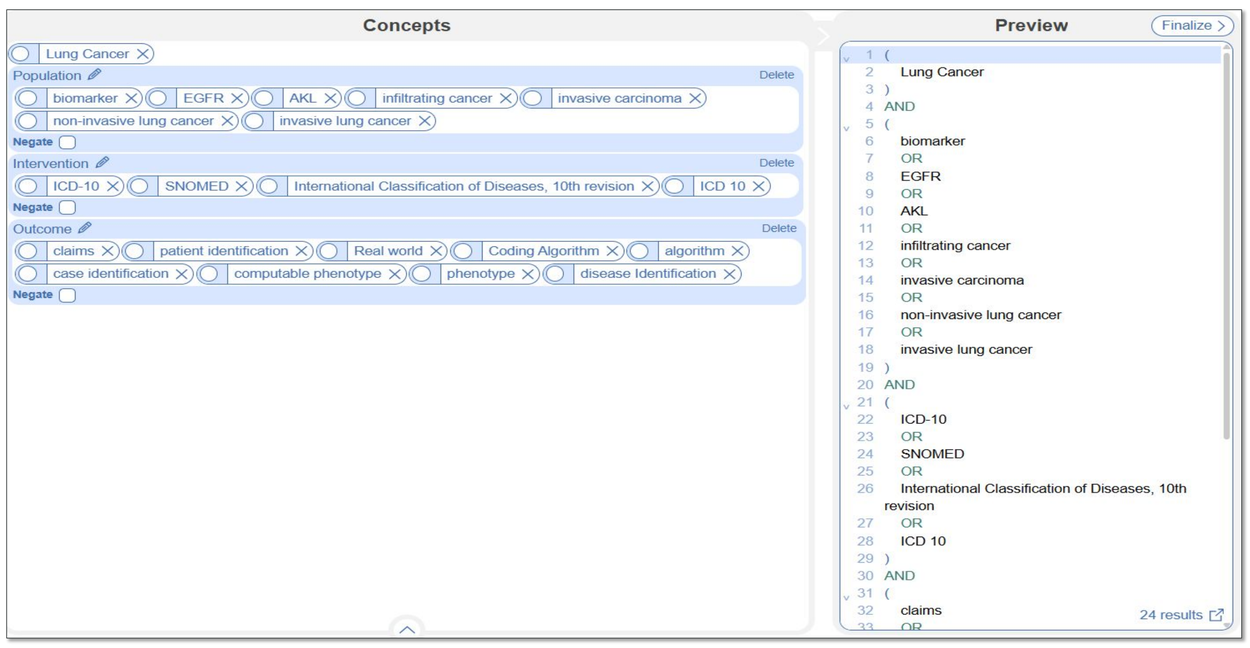


Figure 2: Screening

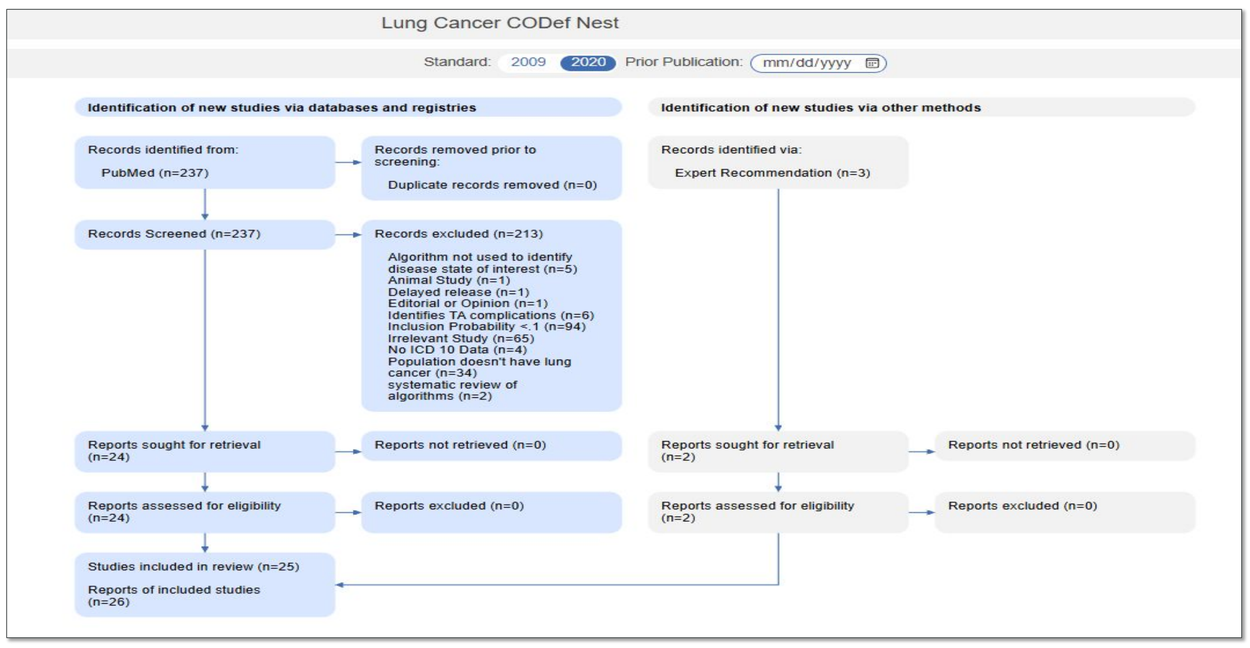


Figure 3: Tagging

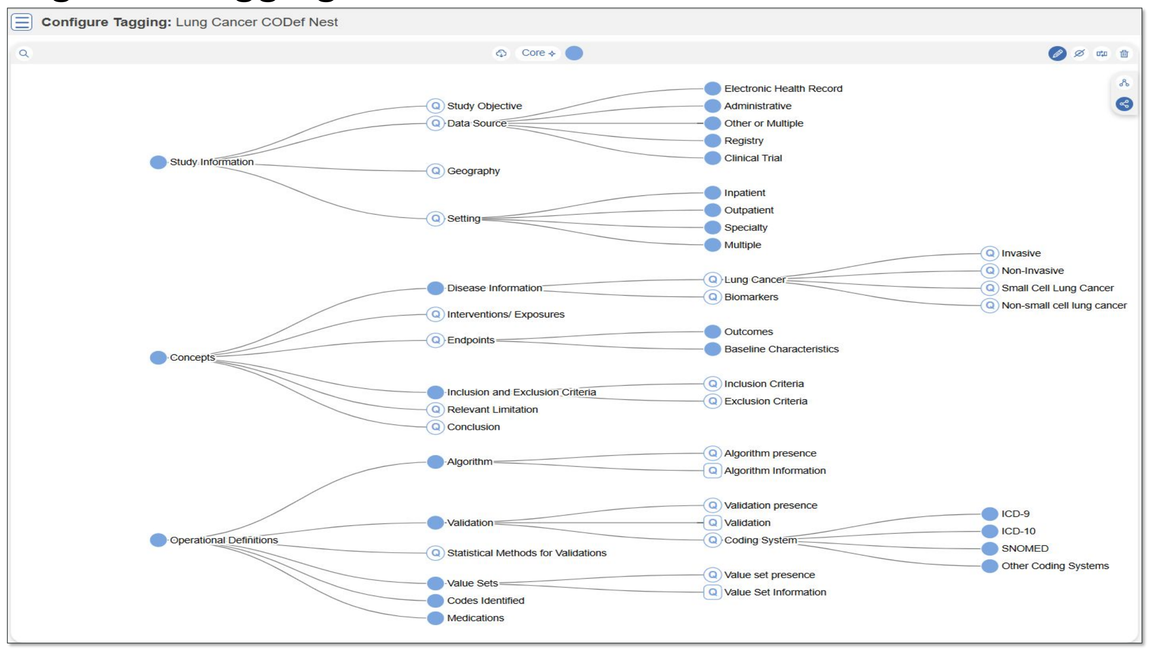
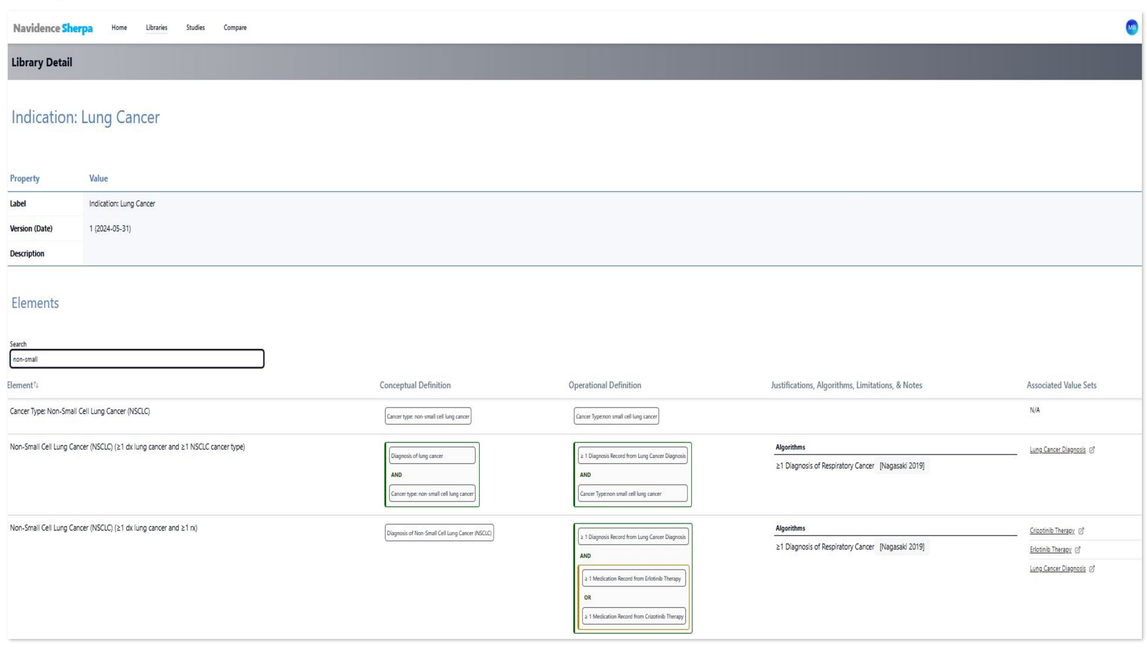


Figure 4: CODef Creation



Conclusion: AI-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