

## Leveraging Artificial Intelligence Based and EHR-integrated Tool to Optimize Real-World Clinical Trial Matching in Oncology at an Academic Medical Center

C. Passaglia<sup>1</sup>, J. Lorch<sup>1</sup>, K. Opfer<sup>1</sup>, M.J. Gurley<sup>1</sup>, A. Zafirovski<sup>1</sup>, F. Lin<sup>1</sup>, A. Brikha<sup>1</sup>, J. Sugay<sup>1</sup>, T. Kumar<sup>2</sup>, S. Mahatma<sup>2</sup>, C. Nienhaus<sup>2</sup>, A. Folkers<sup>2</sup>, M. Hagen<sup>2</sup>, D. Mahalingam<sup>1</sup>, W. Gradishar<sup>1</sup>, S. Bagchi<sup>2</sup>, L. Bezler<sup>1</sup>, L. Schulte<sup>1</sup>, C. Yarbrough<sup>1</sup>, A. Gogunskaya<sup>1</sup>, P. Kartcheske<sup>1</sup>

<sup>1</sup>Robert H. Lurie Comprehensive Cancer Center of Northwestern University; <sup>2</sup>Vizlitics

### 1. Background

Recent advances in AI and electronic health record (EHR) integration have the potential for clinical decision support systems to increase clinical trial enrollment with significantly reduced manual effort. The Lurie Cancer Center (LCC) of Northwestern University's Clinical Trial Office alongside Vizlitics, a healthcare AI company, developed and piloted OnTrial, an EHR-integrated and AI clinical trial screening tool that automates the end-to-end matching workflow — from extracting eligibility criteria from trial documents to reasoning over a patient's full medical record and generating cited explanations, enabling the study team to audit every decision. OnTrial uses HIPAA-compliant infrastructure with large language models (LLMs) designed to scale with the high throughput of patient appointments at large cancer centers.

### 2. Goals

To prospectively validate OnTrial's criterion-level eligibility assessments against experienced oncology nurse reviews in a blinded real-world pilot, and to characterize the sources of disagreements with hopes of increasing the volume of patients considered for clinical trials by using OnTrial when deployed at LCC. Patients with upcoming appointments are screened using EHR information, matching the inclusion and exclusion criteria of the trials. The screening results are presented with logical reasoning and citation of evidence from the EHR.

### 3. Solutions and Methods

We used 10 open clinical trials for Head and Neck carcinoma (454 criteria) and 12 patients (18010 indexed records) actively being treated at LCC and reviewed at head and neck tumor boards. Oncology nurses with experience of reviewing patients for clinical trials used OnTrial to review auto-extracted eligibility criteria for correctness and compare against the patients' EHR using search and browsing capabilities. Nurses made criterion-level "yes", "no", or "unknown" eligibility assessments and provided snippets of evidence (for or against) from relevant patient documents. This produced over 2000 patient-criteria assessments for 65 patient-trial combinations.

These patient-trial combinations were then run with OnTrial using language-model based search, evidence generation and "yes", "no", or "unknown" assessments for each eligibility criterion. A validation set of five trials screened for one patient was used to develop and tune the system which was then run on the test set of the remaining 60 patient-trials which were held blind to the developers. A random 10% of disagreements between OnTrial and original nurses were reviewed and adjudicated by two Head and Neck clinical trial nurses.

#### **4. Outcomes**

OnTrial's criteria-level eligibility assessments matched the original nurse assessments for 49 percent of the criteria using an exact concordance (yes/no/unknown assessments) rising to 73 percent when system "unknown" assessments are acceptable (as it was not allowed to infer clinical judgements when lacking explicit information). A random 10 percent of the criteria with initial disagreements were adjudicated by Head and Neck clinical trial nurses to be 64 percent in OnTrial's favor. This points to the inadequacies of the existing capabilities available for the original assessments.

#### **5. Lessons Learned and Future Directions**

The preliminary evaluation outlined above indicate the difficulty in obtaining consistent ground truth assessments at the criteria level especially for those with: (1) nested sub-criteria with complex logic, (2) ambiguous criteria language, and (3) unknown, missing, or old information in the EHR. Ongoing work focuses on prospective nurse-supervised deployment to measure increases in patient screening volume, time-to-eligibility determination, and downstream trial enrollment.