An AI-Powered Tool to Support Older Adults in Understanding Lab Test Results (LabGenie)

Enhancing older adults' understanding and engagement with their lab test results through an AI-powered web-based tool.

Study Overview

Problem: Patient portals are widely used in primary care settings to give people direct access to their health information, including doctor's notes, medication lists, and laboratory (lab) test results. Viewing lab test results is consistently rated as one of the most useful features of portals. However, many patients—particularly older adults—often struggle to locate, view, and interpret the lab results they receive through these systems. A previous survey conducted by the research team found that, compared to younger patients, older adults were more likely to use patient portals and scored higher on lab result comprehension tests. [1] However, older patients still reported difficulty interpreting lab values and preferred to discuss results directly with their doctors. [1] This finding suggests that current portal systems may not provide sufficient interpretive support to help patients understand their results without clinical guidance.

Main Objective: To develop and evaluate a web-based patient engagement tool (LabGenie) that will help older adults understand, manage, and meaningfully use the lab test results they receive through patient portals.

Approach: Using an iterative and user-centered approach, the research team will develop and test paper prototypes of LabGenie to determine how to present lab test results in ways that are easily understood by patients. Building on this, the team will develop and implement a functional prototype of LabGenie that includes an AI-powered module capable of extracting relevant information from the patient's electronic health record. This data will be used to generate contextualized explanations and tailored question prompts that can help patients prepare for follow-up consultations with their clinicians. The research team will then evaluate the impact of the LabGenie tool based on how patients perceive the ease of use, accessibility, and usefulness of the tool; patient engagement; and patient self-efficacy in shared decision-making. They also will conduct clinician interviews to assess LabGenie's potential effects on clinical workflow and burden in primary care settings.

Results: To date, the research team has conducted four design workshops with older adult participants to explore their preferences for lab test visualization styles. Participants preferred clearly labeled number line formats with solid color blocks indicating levels of normality. These preferences were confirmed through A/B testing, providing guidance for developing the LabGenie user interface. The team is planning to conduct a randomized controlled trial to compare the effectiveness of adding contextualized explanations versus generic explanations to lab reports in enhancing patient comprehension. The findings will help determine whether LabGenie's ability to generate personalized, patient-specific information can improve comprehension and support patient activation.

To support the development of question prompts and responses to frequent questions about lab test results, the research team compared large language models (LLMs)—GPT-4, GPT-3.5, Llama 2, ORCA_mini, and MedAlpaca—as well as human responses from lay persons responding to Q&As on Yahoo Answers.[2] LLMs are a type of AI that are pre-trained on vast amounts of data and capable of generating natural language responses. When the team had medical experts evaluate the various responses, they found that GPT-4 responses outperformed the other LLMs and lay responses across all four measured domains: relevance, correctness, helpfulness, and safety.[2] The team identified ways to improve the quality of responses to address cases where GPT-4 responses were inaccurate and not individualized. These improvements included prompt engineering, prompt augmentation, retrieval-augmented generation, and structured response evaluation strategies.

The research team also compared five LLMs—GPT-4, GPT-3.5, Llama-2-70B, Claude-2, and Mixtral-8x7B—for generating differential diagnoses from selected case reports available in PubMed Central.[3] They found that GPT-4 again performed better than other models. The team also found that incorporating lab results data enhanced accuracy by up to 30% across models, underscoring the critical role of lab test results in accurate medical diagnosis.[3] Lab data, such as liver function tests, toxicology/metabolic panels, and serology/immune tests, were generally interpreted correctly, improving the models' ability to generate relevant diagnoses.[3]

Additional findings for this study are forthcoming and will be shared in future publications, which will be posted here.

Primary Care Relevance

LabGenie has the potential to transform how lab test results are delivered to patients and thereby enhance patient comprehension, engagement, and activation—ultimately supporting shared decision-making and improving health outcomes. In addition, LabGenie has the potential to improve clinical workflow and reduce clinician burden in primary care settings.

_____

1. Lustria MLA, Aliche O, Killian MO, He Z. Enhancing patient engagement and understanding: is providing direct access to laboratory results through patient portals adequate? JAMIA Open. 2025 Mar 24;8(2):ooaf009. doi: 10.1093/jamiaopen/ooaf009. PMID: 40130170

2. He Z, Bhasuran B, Jin Q, Tian S, Hanna K, Shavor C, Arguello LG, Murray P, Lu Z. Quality of answers of generative large language models versus peer users for interpreting laboratory test results for lay patients: evaluation study. J Med Internet Res. 2024 Apr 17;26:e56655. doi: 10.2196/56655. PMID: 38630520; PMCID: PMC11063893.

3. Bhasuran B, Jin Q, Xie Y, Yang C, Hanna K, Costa J, Shavor C, Han W, Lu Z, He Z. Preliminary analysis of the impact of lab results on large language model generated differential diagnoses. NPJ Digit Med. 2025 Mar 18;8(1):166. doi: 10.1038/s41746-025-01556-8. PMID: 40102561; PMCID: PMC11920259.