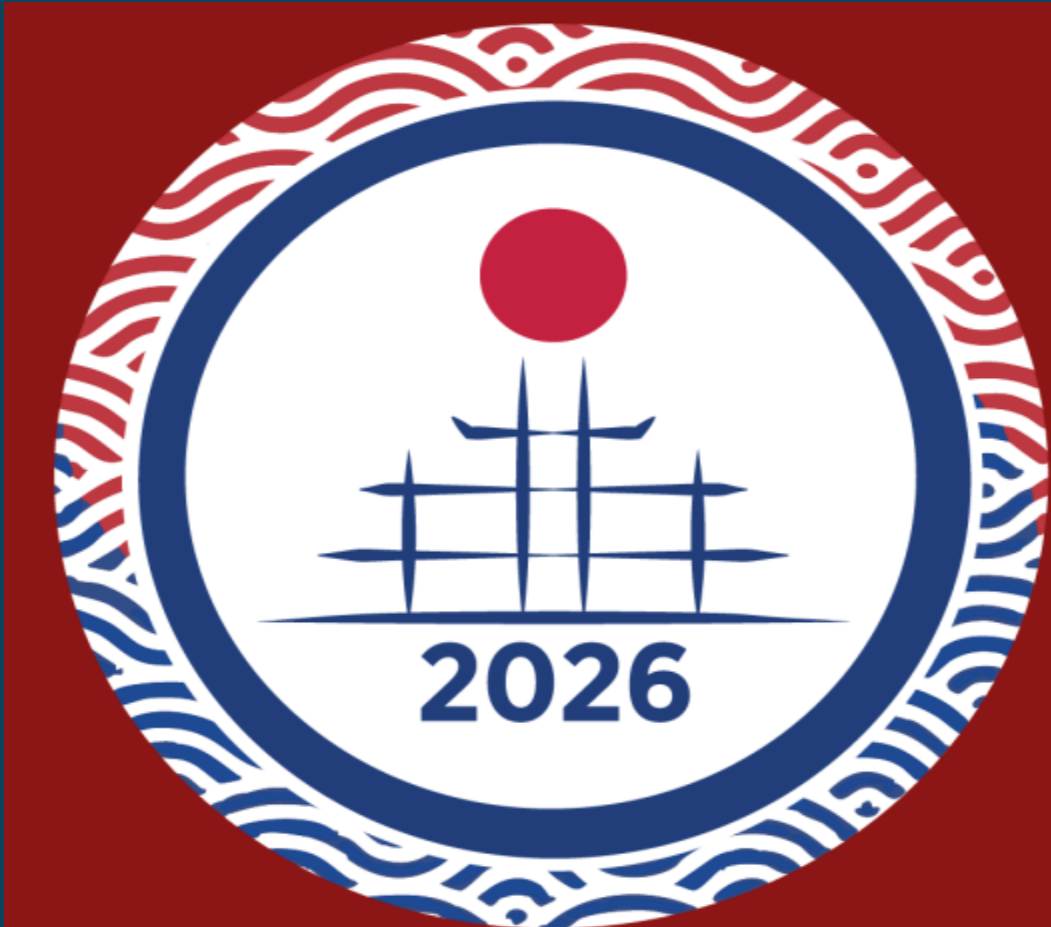




Leveraging Generative AI to Identify CT Imaging Indications for Post-Treatment Surveillance in Lung Cancer

*Annabel X. Tan, PhD, MPH, *Aparajita Khan, PhD, *Ingrid Luo, MS; Grant M. Nieda, BS; Tony Chen, BS; Emily Rodriguez, BS; Mina Satoyoshi, MS; Archana Bhat, MS; Maggie Shaw, MS; Solomon Henry, MS; Meghan Martinez, MPH; Victoria Ding, MS; Tae Yoon Lee, PhD; Kevin ten Haaf, PhD; Rafael Meza, PhD; Curtis Langlotz, MD, PhD; Joel Neal, MD; Allison Kurian, MD, MSc; Heather Wakelee, MD; Ann N. Leung, MD; Su-Ying Liang, PhD; Julie T. Wu, MD, PhD; Summer Han, PhD



BACKGROUND

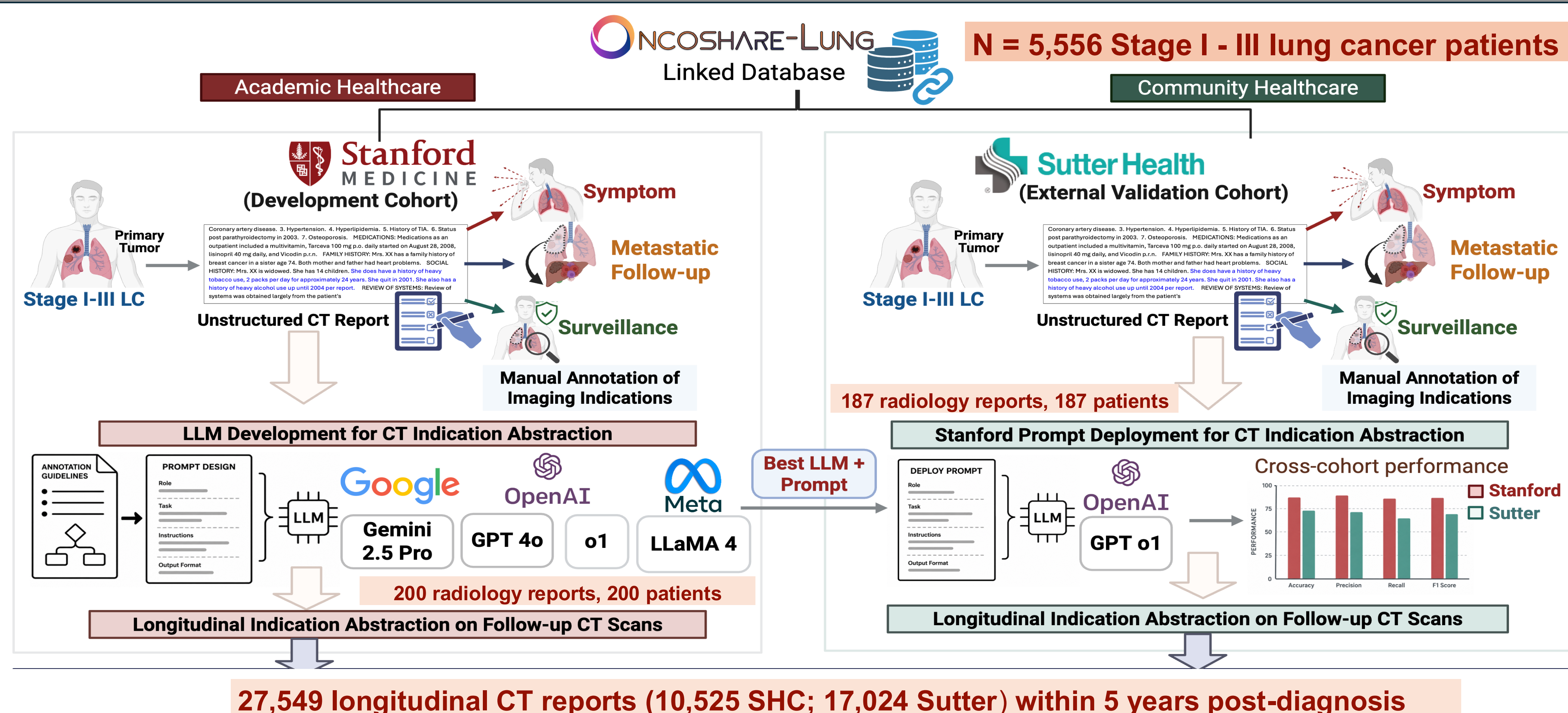
- Current clinical guidelines recommend routine chest CT surveillance every 6–12 months, yet these remain consensus-based
- A fundamental barrier is that health records rarely distinguish whether CTs are performed for routine surveillance versus symptom evaluation
- Conflating these clinically opposite indications introduces confounding by indication, obscuring true associations.

OBJECTIVES

- We leveraged large language models (LLMs) to extract imaging intent from unstructured radiology reports, enabling evaluation of surveillance CT survival benefits across different healthcare systems.

Hypothesis: Surveillance CT is associated with better survival, and this may be through the early detection of recurrence or SPLC

METHODS



- Associations with LC-specific survival were estimated using time-varying Cox models, treating LLM-predicted surveillance as a time-varying exposure to mitigate immortal time bias and adjusting for clinicodemographic variables.
- Secondary analyses evaluated recurrence and SPLC.

Contact: Summer Han, PhD; summer.han@stanford.edu
Associate Professor of Neurosurgery, Medicine and Epidemiology & Population Health
Faculty Lead, Cancer Data Science Core, Stanford Cancer



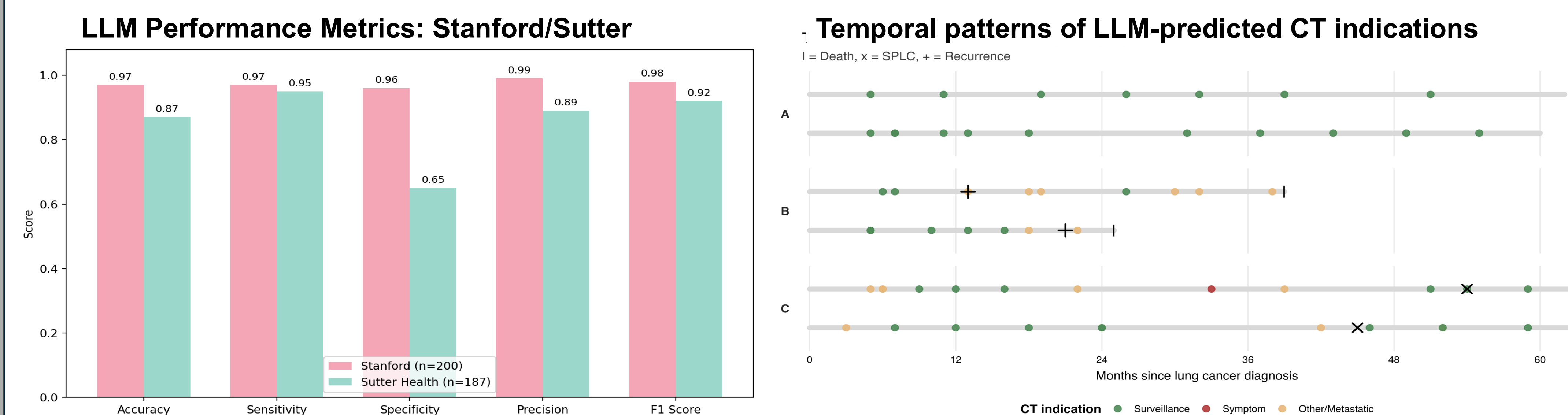
Cancer Data Science Core

Scan to learn more about the Cancer Data Science core!

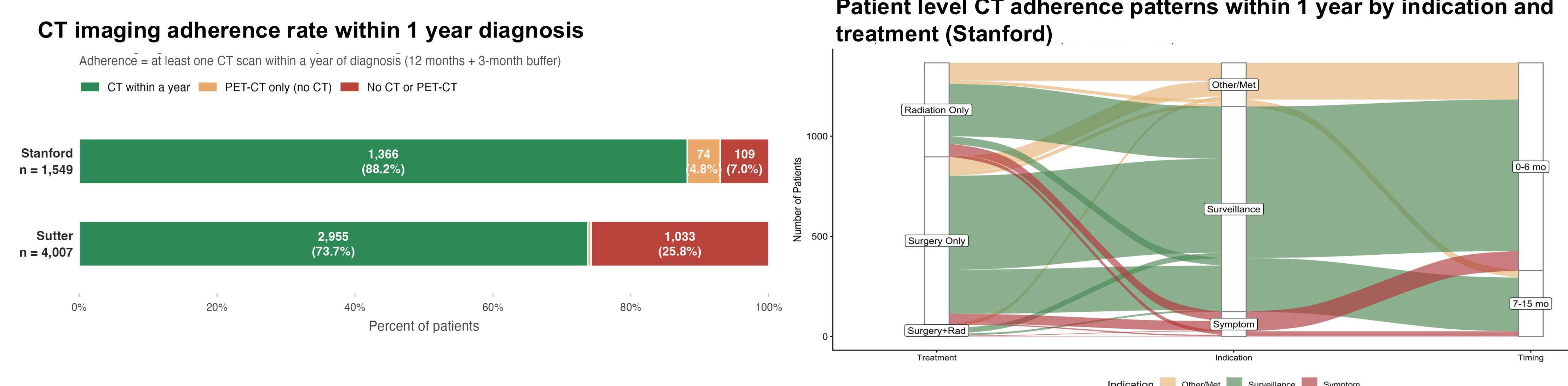


RESULTS

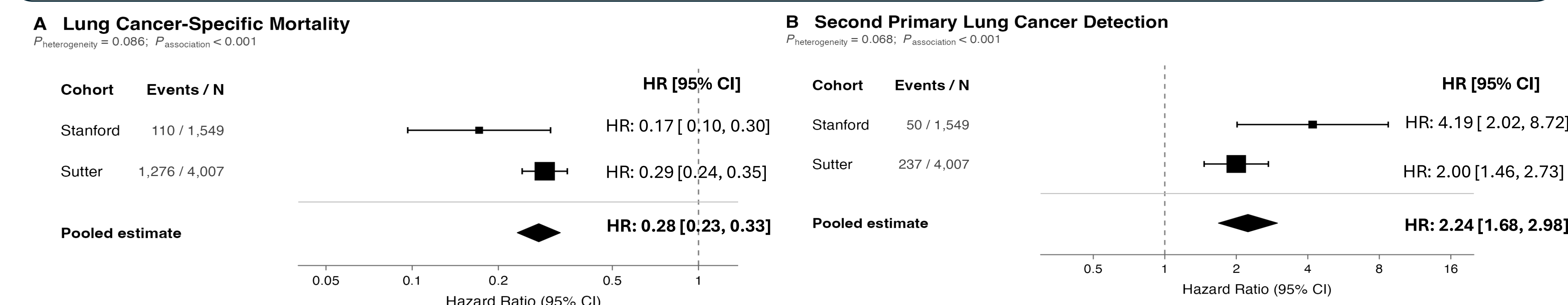
Result 1: In classifying surveillance CTs vs others, the most accurate LLM (OpenAI o1) achieved 97% accuracy (F1: 98%) on SHC annotations and 87% accuracy (F1 92%) on external validation at Sutter



Result 2: First-year CT adherence (≥ 1 any CT within 12-months ± 3 -months buffer) was 88.1% at SHC and 73.3% at Sutter. At both, surveillance was the most common indication



Result 3: Time-varying “surveillance” CT was associated with improved LC-specific survival. Surveillance CT was also associated with higher SPLC detection (SHC HR=4.19; Sutter HR=1.98; both $p < 0.001$) but not recurrence ($p > 0.05$)



CONCLUSIONS

- This study provides rigorous real-world evidence associating guideline-recommended CT surveillance with improved LC-specific survival across academic and community settings.
- LLM-extracted imaging intent offers a scalable, multi-system framework for surveillance effectiveness research that addresses confounding.