Kimberly Shoenbill, PhD, MD, MS
Assistant Professor; UNC Tobacco Treatment Program Director
Family Medicine

Key Members:
Louise Henderson, PhD, MSPH, Co-I (LCS, registry expert)
Adam Goldstein, MD, MPH, Co-I (Director UNC Tobacco Intervention Programs)
Danielle Durham, PhD & Pasangi Perera, MS - data analytics
Derek Lam, BS - EHR data abstractor; TBN - NCTraCS Analyst for CDW data retrieval
TBN - Computer programmer for NLP and ML tool development

Project Title: Using NLP Algorithms to Improve Identification and Use of EHR Tobacco Information

Describe your key question, in particular, what clinical decisions this work intends to inform. (Max 350 characters)

How can researchers more accurately identify tobacco use documentation in the EHR? This work will develop methods to improve tobacco use data identification to increase equitable, evidence-based lung cancer screening and tobacco use treatment (TUT) now, with future use in other tobacco-related illness risk assessment, outreach, and interventions.

Describe the significance and innovation of this work. (Max 350 characters)
Provide a statement of the problem. What is currently the standard? What is the proposed intervention? What is the alternative to the intervention? What is the clinical decision this would inform? How would this potentially change clinical care? What is the anticipated scale of impact ($/people/procedures)?

136,000 patients die of lung cancer yearly. EHR structured tobacco data is used to inform lung cancer screening (LCS) but is inaccurate, identifying only 35% of eligible patients. NLP and ML could extract and use tobacco data from clinical notes to improve evidence-based LCS and TUT referrals to decrease tobacco-related morbidity and mortality.

Describe what additional assistance is needed to move your project forward. (Max 600 characters)
What assistance from PPMH would be most helpful? Describe what stage in the translational pipeline you view your project to be at, and consider different domains of support that could be useful in your work: In addition, please indicate whether there are existing resources that could help your team accomplish your goals; if not, how are your existing resources inadequate?

PPMH could assist this work in providing methodological insight and NLP and machine learning programming assistance to develop and implement NLP and ML tools for tobacco use data extraction and predictive model development for LCS and TUT referrals. If successful, these algorithms could inform an EHR CDS tool for LCS and TUT referrals. We do not have adequate computer programming expertise for NLP and ML development or expertise in CDS tool development and implementation that PPMH could potentially provide.

What is the next step for this project after completion? (Max 350 characters)

If NLP and ML tools are successful in tobacco data capture and prediction of appropriate referrals to LCS and TUT, funding for a pilot study of CDS tool development for LCS and TUT referrals will be pursued. Data from the pilot study would then support an RO1 application for CDS tool refinement, implementation and outcome analysis.