



**Deep Learning Program
Triangle Machine Learning Day
September 20, 2019**

SPEAKER TITLES/ABSTRACT

David Page

Duke University

“Machine Learning from De-Identified Coded Electronic Health Records (EHRs)”

This talk begins by showing how accurately 4000 different diagnoses can be predicted in advance for any patient, from one month to twenty years before first occurrence in the patient, using high-throughput machine learning. Shortcomings of this approach motivate ways to turn prediction methods into algorithms for finding causal associations; the resulting algorithms attain high accuracy in tasks of drug repurposing and discovery of adverse drug events, but they do not come with provable guarantees of making correct causal inferences. We then introduce variants of probably-approximately correct (PAC) learning for finding causal associations, that can provide weaker but useful guarantees for such algorithms as these motivated by our experiences with EHR data.