Submitter Name: Roy Perlis Submitted Email:rperlis@mgh.harvard.edu

Uses and misuses of machine learning for electronic health records

Roy H. Perlis, MD MSc

Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA

Background: Large clinical data sets have been enthusiastically embraced across medicine for their capacity to drive discovery as well as test hypotheses entirely in silico.
Significance: Applying real-world evidence for clinical and translational investigation offers tremendous advantages over other methods in terms of feasibility as well as face validity.
Hypothesis: Established as well as emerging machine learning methods allow novel phenotypes, dimensional traits, as well as treatment outcomes to be extracted from coded and uncoded clinical data.

Results: These strategies incur costs in terms of risk for bias and confounding as well as type 1 error, but with cautious application can facilitate genomic, translational, and clinical investigation.
Discussion: This presentation will present multiple vignettes illustrating the application of machine learning to electronic health records, and propose ways of addressing their limitations.