RISK PREDICTION OF A MULTIPLE SCLEROSIS DIAGNOSIS

Joyce C. Ho, Joydeep Ghosh, K.P. Unnikrishnan

Multiple Sclerosis Introduction

- Chronic, progressive, and incurable immune disease
- Nerve damage causes signal disruption in brain and spinal cord
- 2.5 million people worldwide suffer from MS
- Early diagnosis can slow progression and prolong a “normal” lifestyle

Related Work

- Epidemiology studies
  - Kahana, E. 2000. Epidemiologic studies of multiple sclerosis: a review. *Biomedicine & pharmacotherapy*. 54, 2 (2000), 100–102.
  - Ascherio, A.A. and Munger, K.L.K. 2007. Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Annals of neurology*. 61, 4 (Apr. 2007), 288–299.
  - Ascherio, A.A. and Munger, K.L.K. 2007. Environmental risk factors for multiple sclerosis. Part II: Noninfectious factors. *Annals of neurology*. 61, 6 (Jun. 2007), 504–513.
  - Ramagopalan, S.V. and Sadovnick, A.D. 2011. Epidemiology of multiple sclerosis. *Neurologic Clinics*. 29, 2 (May. 2011), 207–217.
- Predictive models
  - Bergamaschi, R. et al. 2007. Early prediction of the long term evolution of multiple sclerosis: the Bayesian Risk Estimate for Multiple Sclerosis (BREMS) score. *Journal of Neurology, Neurosurgery & Psychiatry*. 78, 7 (Jul. 2007), 757–759.
  - Bakshi, R. et al. 2008. Predicting clinical progression in multiple sclerosis with the magnetic resonance disease severity scale. *Archives of neurology*. 65, 11 (Nov. 2008), 1449–1453.
  - De Jager, P.L. et al. 2009. Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *The Lancet Neurology*. 8, 12 (2009), 1111–1119.

Objective

Can EMR-based risk prediction model help medical professionals diagnosis MS?

MATERIALS & METHODOLOGY

NorthShore EMR Data

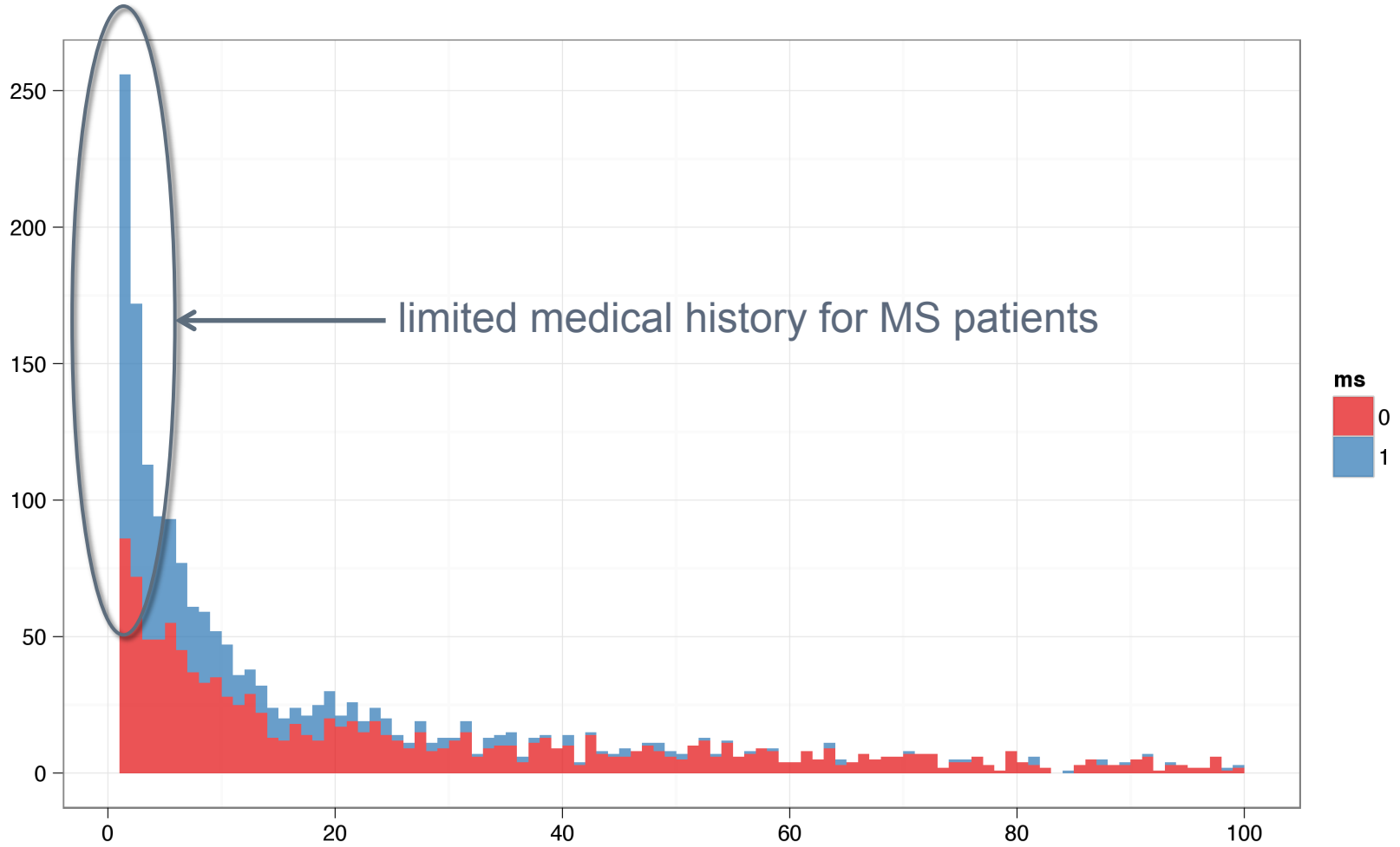
- De-identified patient data from NorthShore Enterprise Data Warehouse between Jan 2006 – July 2012
  - Demographics
  - Family medical history
  - Medical diagnosis
  - Vaccination history
  - Blood test results

Derived from ICD-9

NorthShore MS Data

- Case patients identified using MS ICD-9 codes during neurology visit
- Age and gender matched case-control
- 3,685 patients – 20% prevalence
- Encounters after diagnosis date excluded
  - Control: random encounter date
  - Case: initial MS diagnosis

Encounters Prior to Diagnosis



Feature Construction

- Complete demographic info required
- Family history, medical diagnoses, vaccination history are binarized
- Blood test results encoded as observed-abnormal, observed-normal, unobserved
- Total 56 features

EMR Features

Set	Features
1	Demographic + Family medical history
2	Feature Set 1 + Autoimmune diagnoses
3	Feature Set 2 + Microbial diagnoses
4	Feature Set 3 + Mental illness diagnoses
5	Feature Set 4 + Cancer diagnoses
6	Feature Set 5 + Vaccination history
7	Feature Set 6 + Reproductive related codes
8	Feature Set 7 + MRI scans + obesity
9	Feature Set 8 + Blood tests results

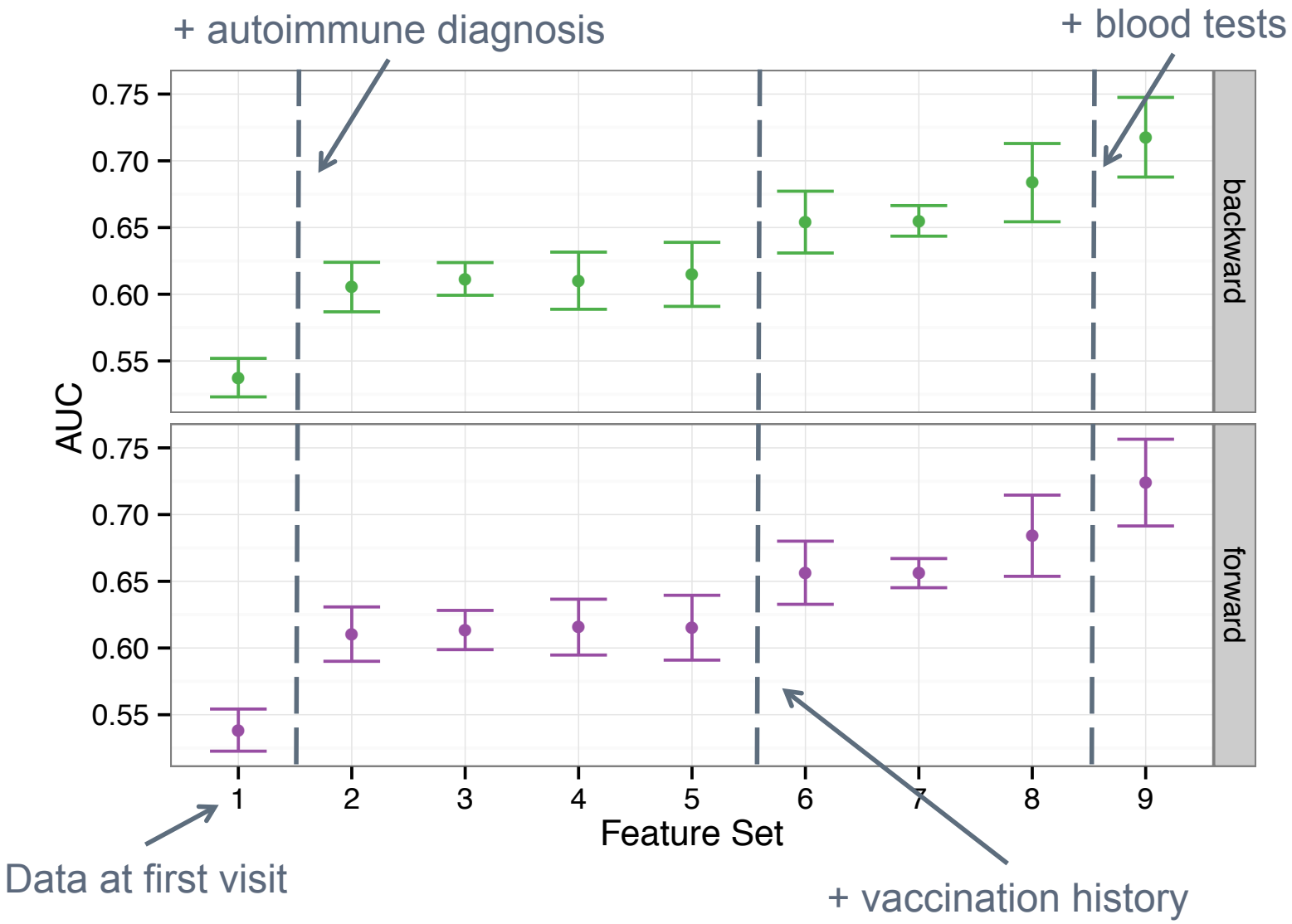
Experimental Setup

- 10-fold cross validation
- Feature selection with forward and backward stepwise regression using Akaike information criterion
- Multivariate logistic regression for classification

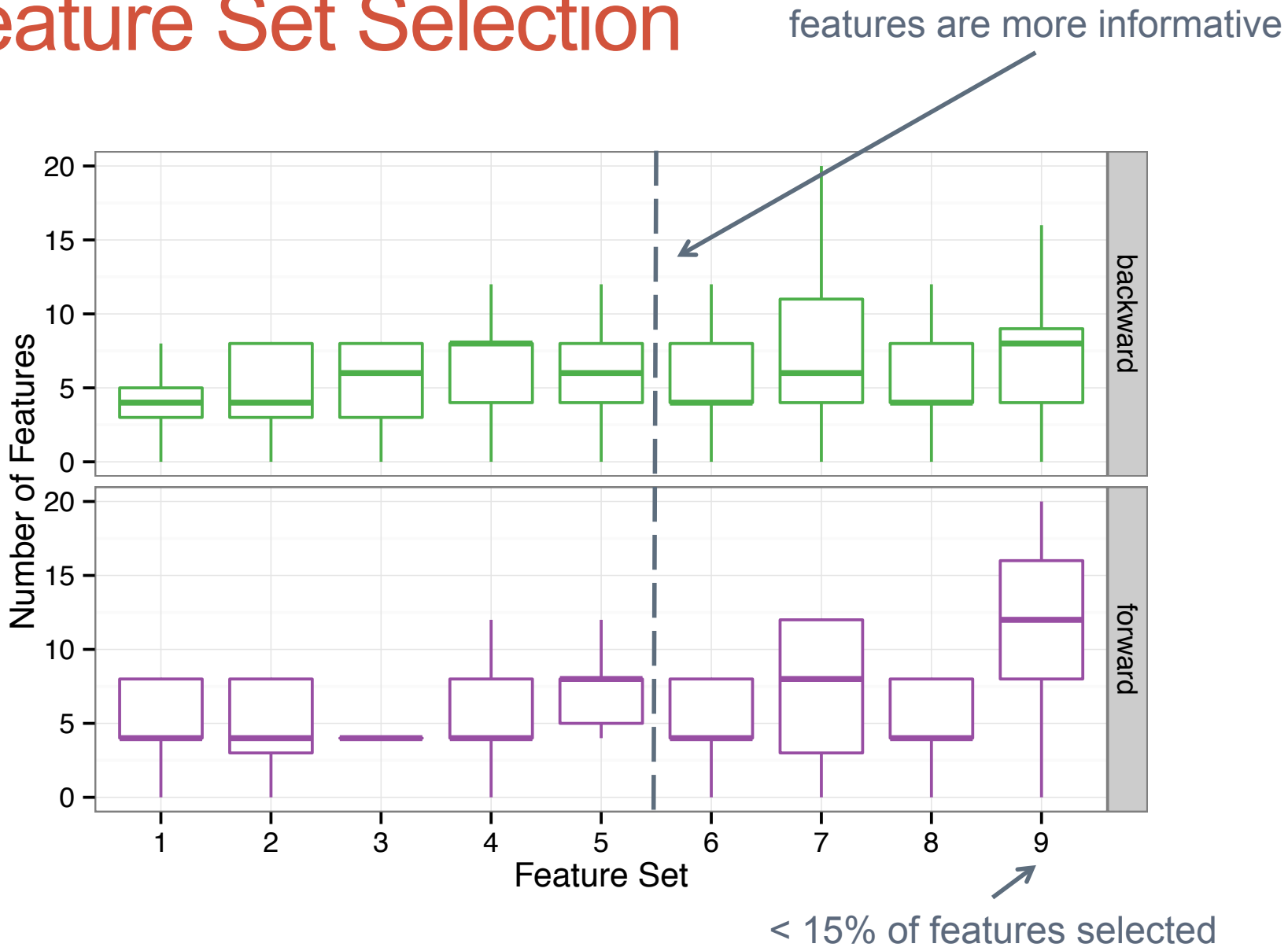
RESULTS

Comparison of EMR Features

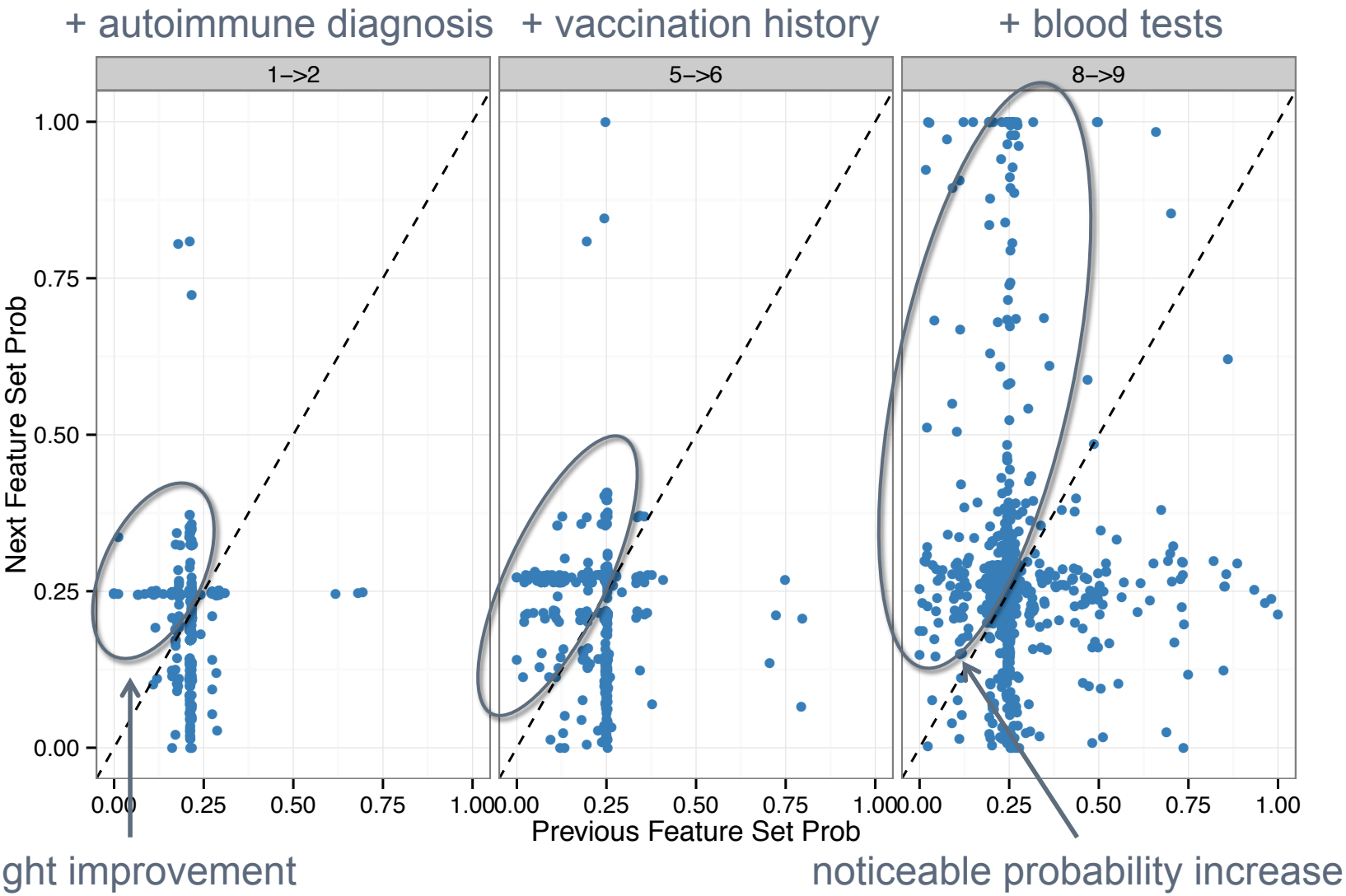
Feature Set Performance



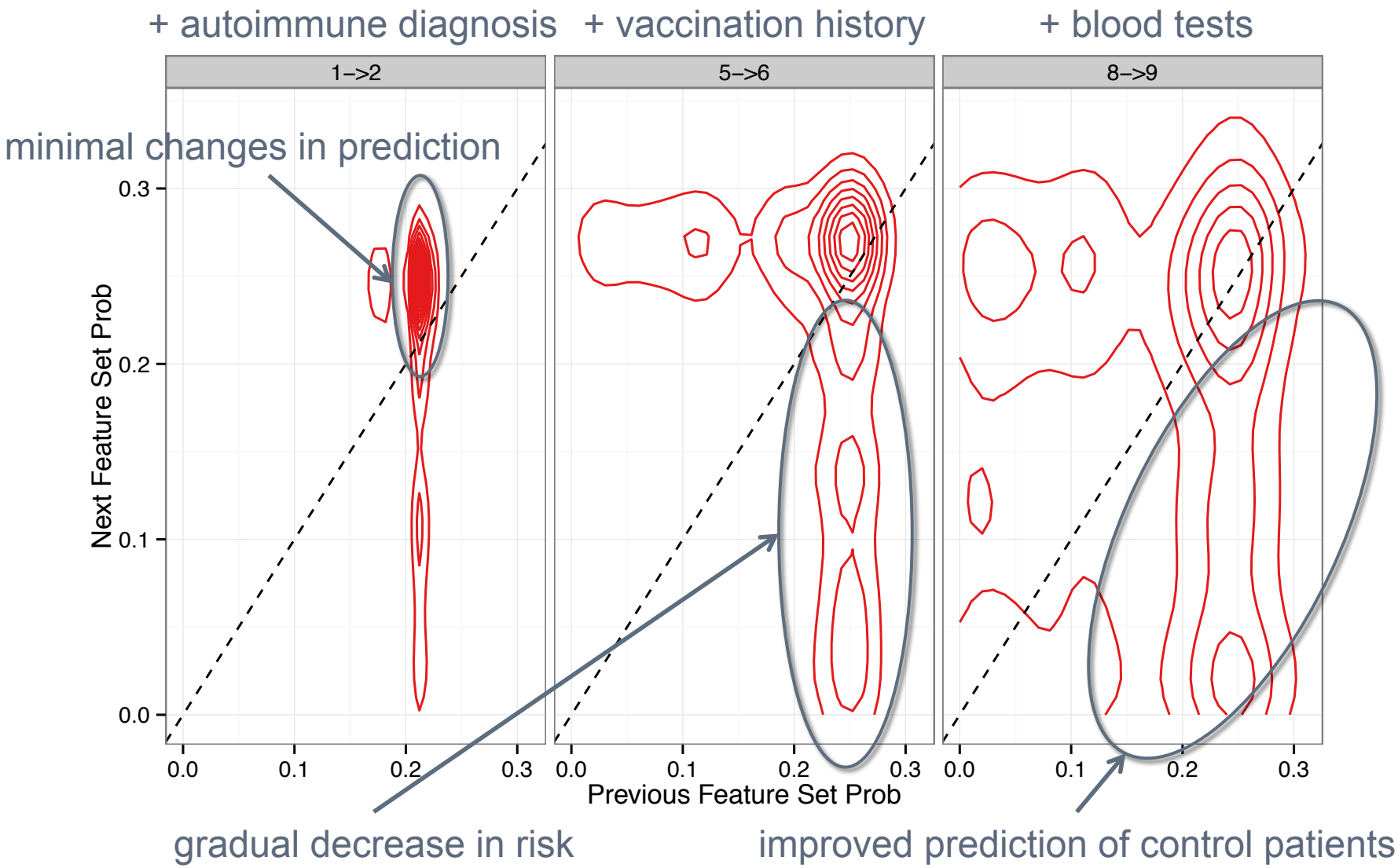
Feature Set Selection



Case Transition Probabilities



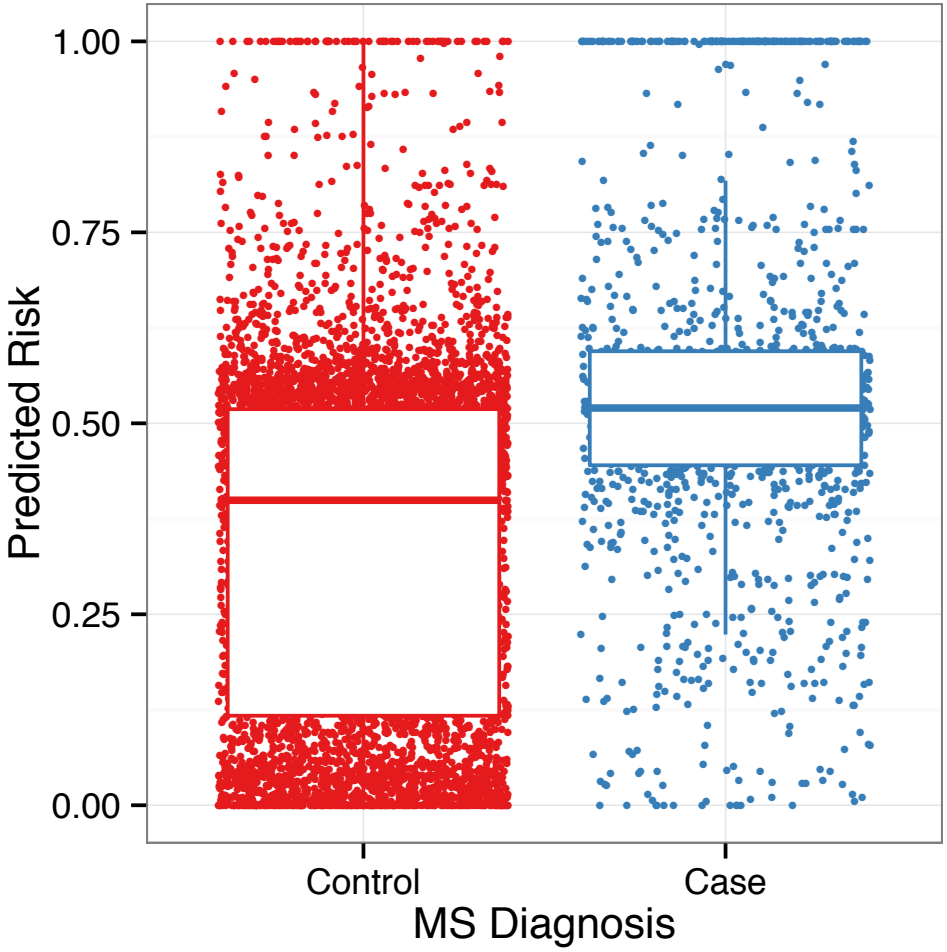
Control Transition Probabilities



RESULTS

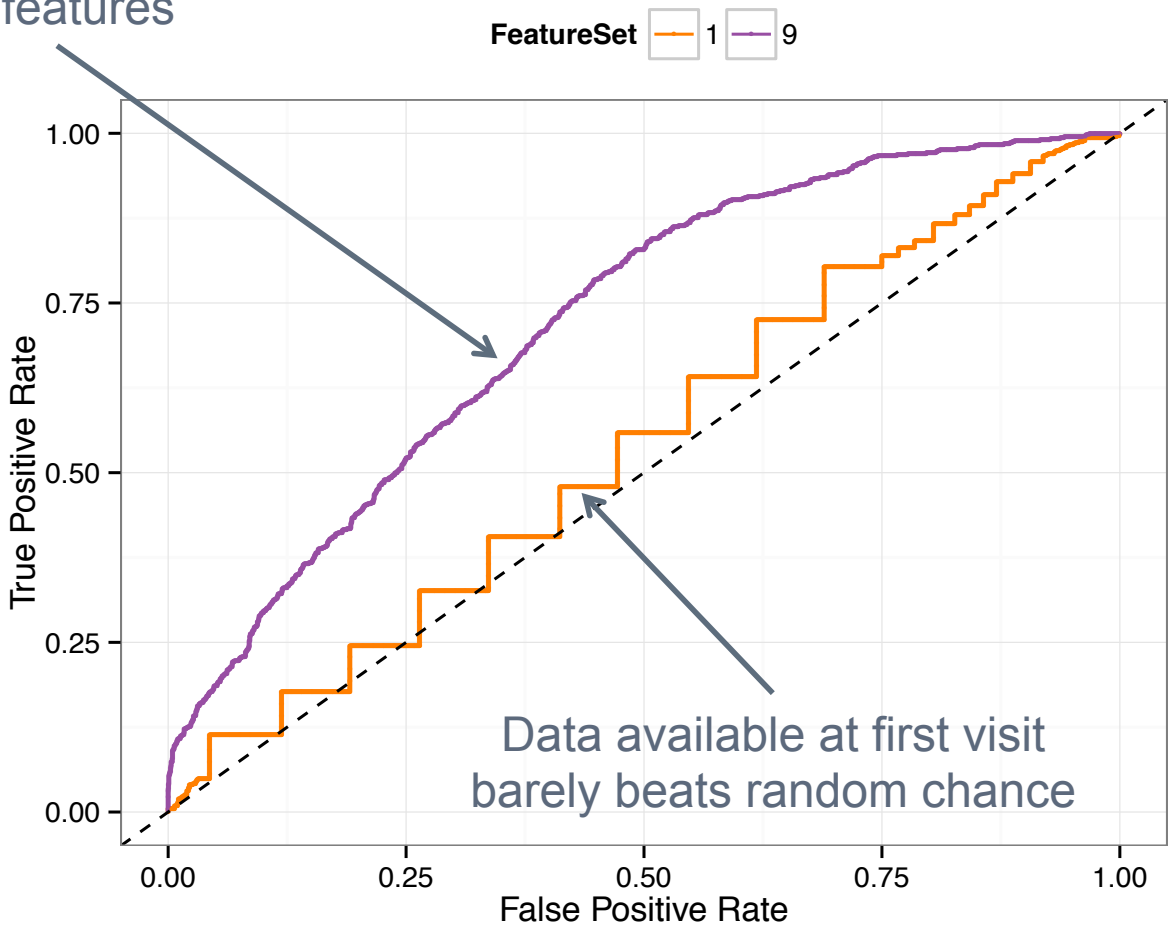
Using All Features with Forward Selection Model

Predicted Risk Probabilities



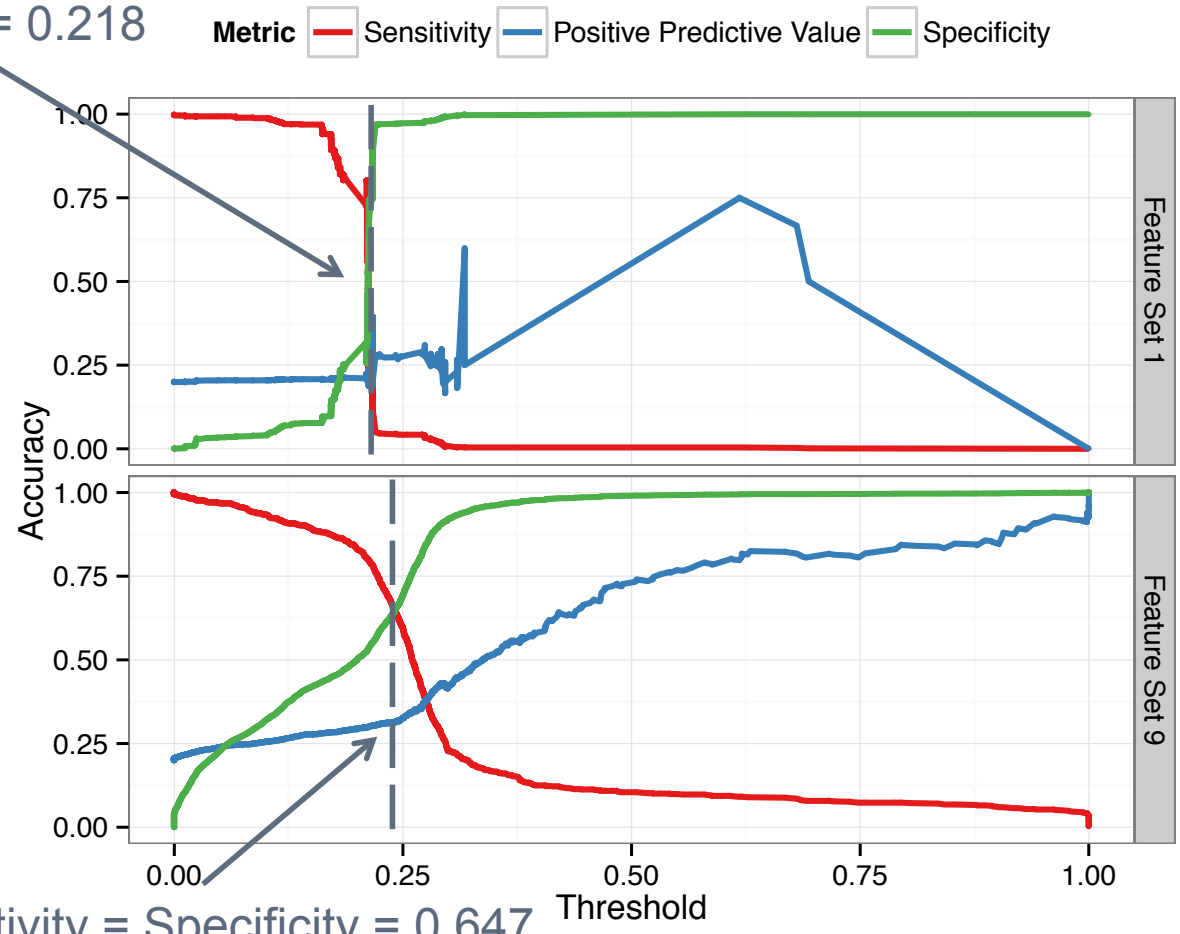
ROC Performance

Significant improvement using all features



Sensitivity, Specificity, PPV Tradeoffs

Sensitivity = Specificity = 0.528
PPV = 0.218



Sensitivity = Specificity = 0.647
PPV = 0.314

Logistic Regression Coefficients

Feature	Beta	P-value
CSF oligoclonal bands (present)	16.255±0.545	0.000
Mental illness (FH)	6.298±3.101	0.033
Epstein-Barr Virus	3.974±3.924	0.093
Abnormal brain MRI	2.877±0.313	0.000
B12 blood test (unobserved)	2.527±0.149	0.047
HPV Vaccine	-15.728±2.188	0.000
Schizophrenia	-15.763±0.369	0.235
Estrogen replacement	-15.823±0.209	0.037
Inflammatory bowel disease (FH)	-17.236±0.420	0.885
HIB vaccine	-18.156±3.101	1.000

CONCLUSIONS

Discussion

- Usage of ICD-9 codes for patient's history
  - Billing data has poor sensitivity and specificity
- Limited patient history for case patients
- Clinical notes review may produce additional, more informative features

Summary

- Model to identify patients at high-risk of developing MS
- Sparse set of features obtains reasonable predictive performance
- Demonstration of leveraging EMRs to aid medical professionals with early disease diagnosis