School of Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY

Data-Adaptive Super Learning to Predict Viral Rebound based on Electronic Adherence Monitoring

An Analysis of the MACH-14 Cohort Consortium

Maya Petersen, MD, PhD
Div. of Biostatistics, School of Public Health,
University of California, Berkeley

on behalf of

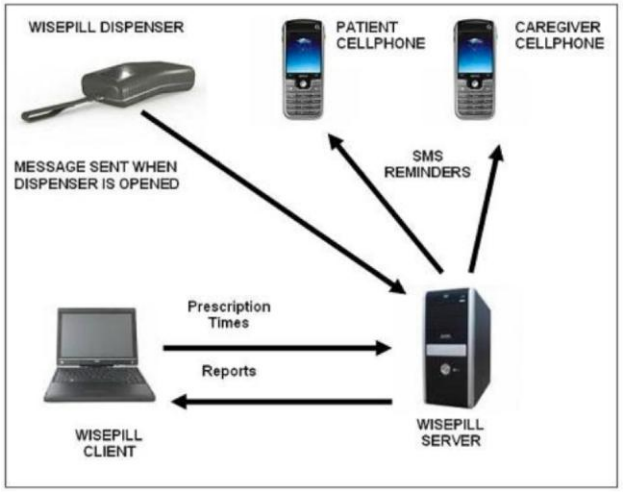
Varada Sarovar, Anna Decker, Erin LeDell, Joshua Schwab, Robert Gross, Ira Wilson, Carol Golin, Nancy Reynolds, Robert Remien, Kathleen Goggin, Jane Simoni, Marc Rosen, Mark van der Laan, Honghu Liu, David Bangsberg

HIV Treatment and Prevention Adherence, Miami Beach FL, June 3-5, 2012

I have no conflicts of interest to declare

Remote Electronic Adherence Monitoring

- Mobile technologies can
  - Monitor adherence remotely
    - Medication Event Monitoring System (MEMS)
  - Transmit data in real time over the cellular network
  - Real time adherence data from Uganda available
    - Haberer Abs #80027, Session 10



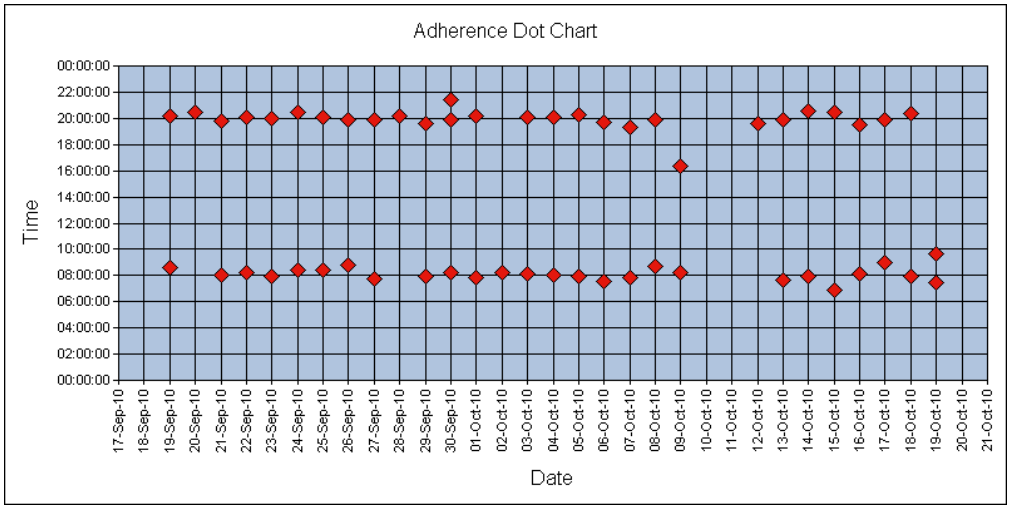
Haberer et al, AIDS Behav. 2010 Dec;14(6):1340-6

Real Time Adherence Monitoring and Personalized Medicine

- Targeted viral load testing?
  - Viral loads every 3 months: Expensive and most tests are negative
  - Target testing to those patients with non-negligible probability of failure?
- Targeted clinic visits?
  - Visits every 3 months: Does not correspond with when people are having trouble
  - Clinic visits and adherence interventions triggered by worrisome adherence patterns?
- Ability to detect relapse in adherence problems?
  - Adherence interventions are short-lived

Can Electronic Adherence Data Accurately Predict Viral Rebound?

- Objective: Predict viral rebound using adherence (MEMS) data and other patient characteristics
- Challenge: MEMS data highly multi-dimensional
  - How to summarize?
    - Average adherence?
    - Interruptions?
    - Nadir adherence?
    - Variance?
  - Over what period?
    - Since previous VL?
    - Shorter? Longer?
  - Interactions?
    - With regimen?
    - With each other?



Building an Optimal Predictor

- Most previous analyses of adherence data to predict viral failure or rebound have used:
  - Single variable predictors
    - Example: Average adherence since previous viral load
  - A *priori* specified models
    - Example: Logistic regression with average adherence and interruptions since previous viral load as main terms
- Machine learning (automated algorithms for signal detection from complex data) may improve the accuracy with which viral rebound can be predicted

Data: MACH-14 Consortium

- Multi-site collaborative study

  - 16 studies at 14 sites
  - 2835 subjects followed between 1997 and 2009
  - Longitudinal MEMS, virologic and clinical data
- We predicted viral rebound (>400 copies/ml) among subjects with
  - Previous viral load <= 400 copies/ml
  - >=1 day of MEMS monitoring in preceding 7 days
  - Basic clinical data (prior CD4, regimen) observed

➢ 1768 viral loads in 754 subjects

  - 147 viral rebounds observed in 134 (18%) subjects

Candidate Predictor Variables

- Adherence summaries
  - Average adherence
  - # of Interruptions of at least 2-14 days duration
  - Nadir of moving average (over 2-28 day windows)
  - Variance of average daily adherence
  - Above summaries using only weekdays and only weekends
- Each summarized over
  - Days since previous viral load
  - 7-168 days preceding current viral load date

- Additional Predictors
  - Days monitored with MEMS
    - For each period over which adherence summarized
  - Site
  - Regimen
  - Drug class monitored
  - Most recent CD4
  - Time on study
  - Time since viral load
  - Time since CD4

- 809 *a priori* specified candidate predictor variables

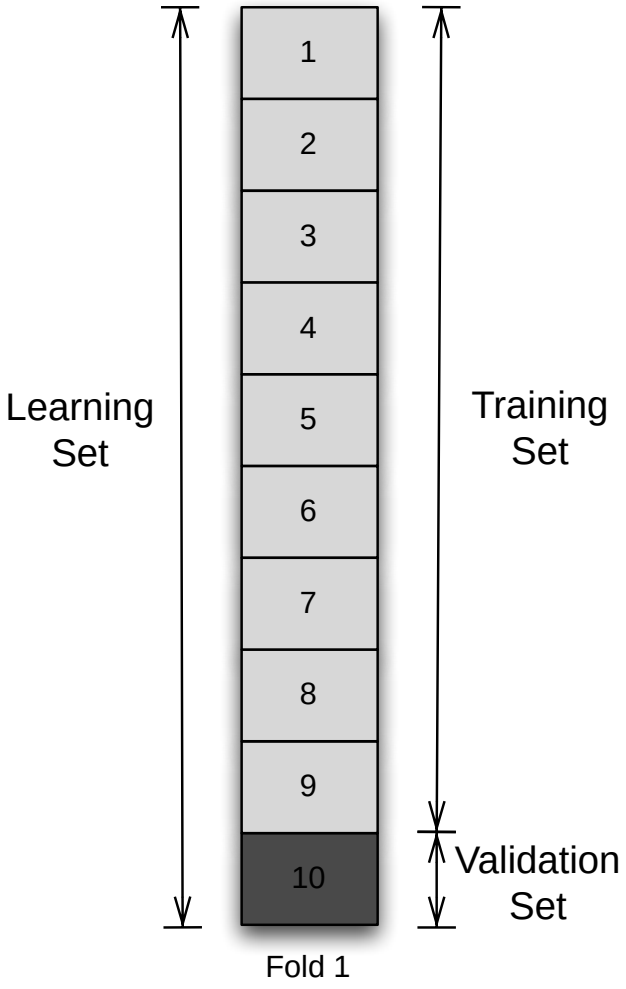
Super Learning

- The user supplies a library of competing prediction algorithms
  - Generalized additive models, Lasso regularized generalized linear models, generalized boosted regression models, multivariate adaptive polynomial spline regression
  - Also consider combinations of all of the above
- Data split into 10 parts and 10-fold cross-validation is used to choose between algorithms
  - Predictive performance is evaluated on data that are not used to fit the prediction model

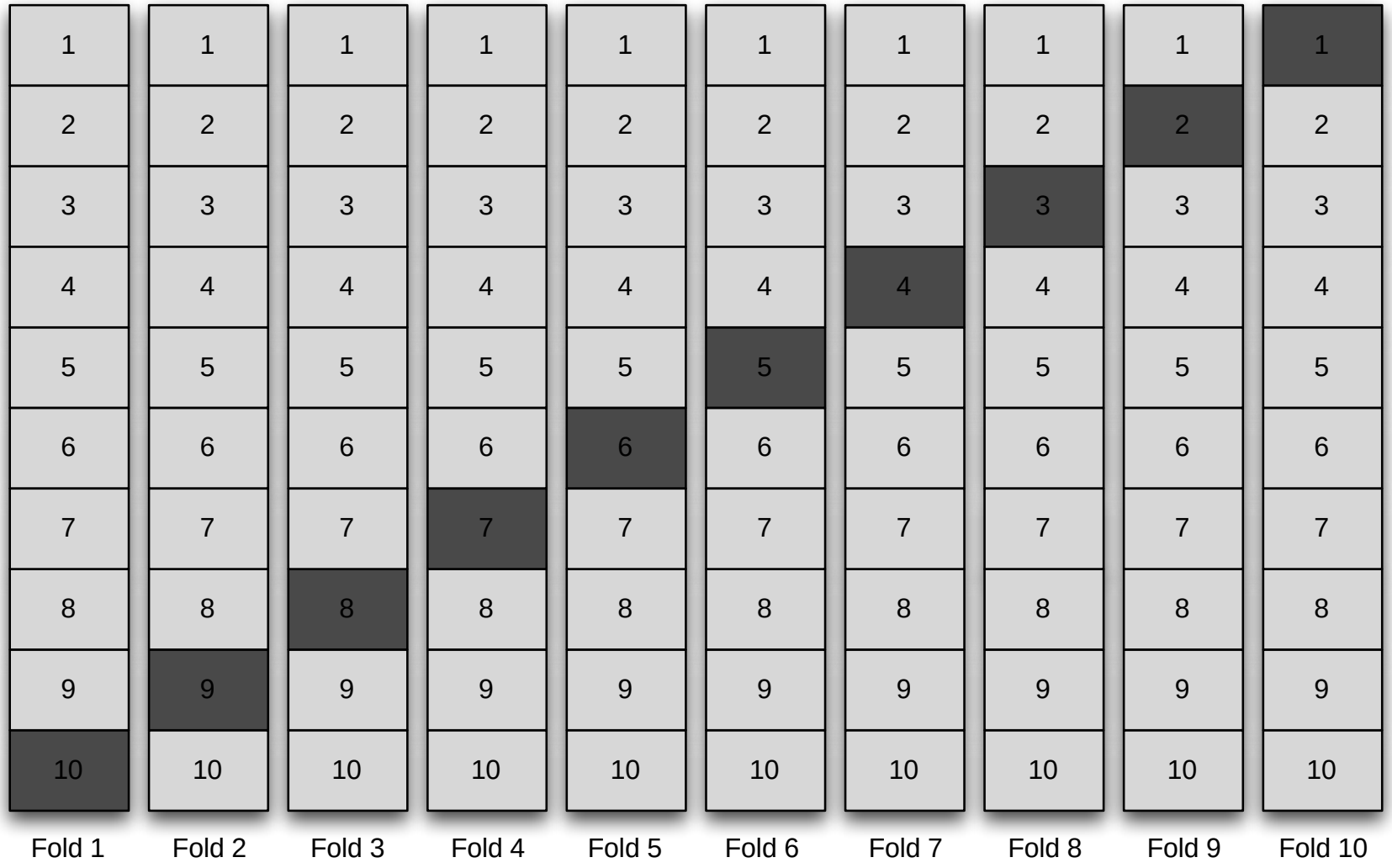
Van der Laan, Polley and Hubbard (2007) Super Learner. *SAGMB*; 6(25)

Example: 10-fold Cross-Validation

1. Split the data into 10 parts
2. Run the competing algorithms on 9/10ths (the training set)
3. Evaluate performance on the remaining 1/10th (the validation set)
4. Repeat for each of 10 training/validation sets
5. Average estimates of performance



Example: 10-fold Cross-Validation



Measuring Performance: ROC Curves

- Cross-validated Area Under the Receiver Operating Characteristic curves [cvAUC] calculated based on data not used in model fitting
  - Super Learner run separately on each of 10 training sets
  - Resulting predictor used to generate Area Under the ROC Curve for the corresponding validation set
  - cvAUC calculated as the average across validation sets

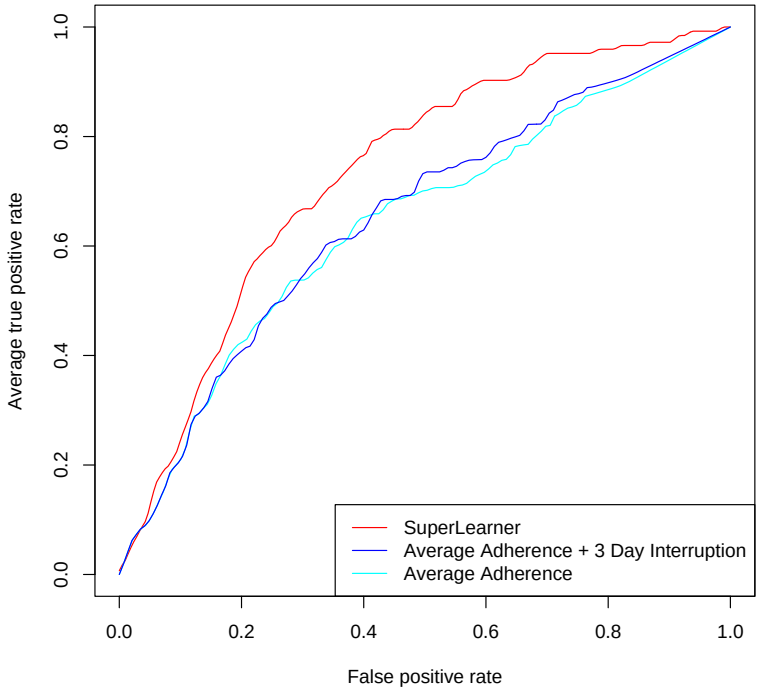
- 95% CIs for cvAUC calculated using influence curve based approach
  - Accounting for repeated measures on a subject
  - New theoretical results (van der Laan)
  - R package to be released soon (`cvAUC`, LeDell)

Results: Sample Characteristics

Candidate Predictor Variable *(Summarized over VLs meeting inclusion criteria)*	**Median (IQR)**
Average Adherence since previous viral load	87% (55%, 99%)
Number of interruptions >24 hours since previous viral load	2 (1, 7)
Days since previous viral load	32 (28, 86)
Days monitored using MEMS since previous viral load	28 (23, 56)
Most recent CD4 T cell count (cells/µl)	384 (218, 570)
	Proportion
At least one interruption >72 hours since prior viral load	38%
NNRTI-based regimen	23%
PI- based regimen	34%
Boosted PI- based regimen	12%
Other ART regimen	31%

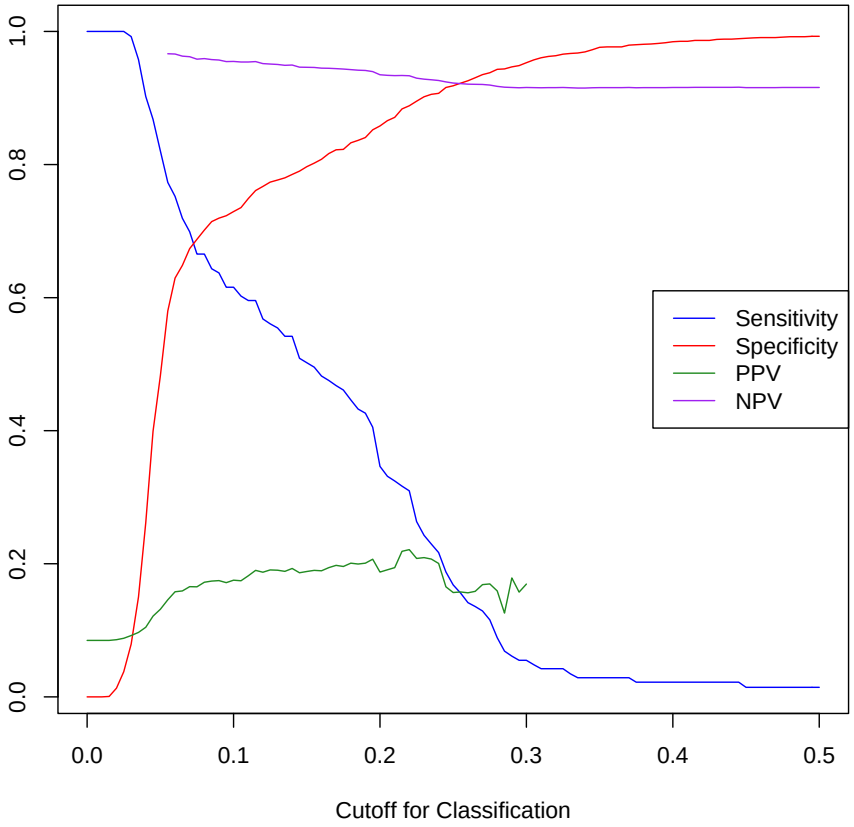
CV Area Under the ROC Curve

- Super Learning resulted in improved classification of viral rebound compared to simple *a priori* specified predictors (p<0.001)



Prediction Model	cvAUC (95% CI)
Average Adherence	0.64 (0.59, 0.70)
Average Adherence + 3 day interruption (main term logistic regression)	0.65 (0.60, 0.70)
Super Learner	0.73 (0.69, 0.77)

Classification of Viral Rebound using Super Learner: Cross Validated Performance Measures



Sensitivity	Specificity	PPV	NPV
0.90	0.26	0.10	NaN
0.75	0.63	0.16	0.97
0.67	0.70	0.17	0.96
0.62	0.73	0.18	0.95
0.57	0.77	0.19	0.95
0.54	0.79	0.19	0.95
0.48	0.81	0.19	0.94
0.45	0.83	0.20	0.94
0.35	0.86	0.19	0.93
0.31	0.89	0.22	0.93
0.22	0.91	0.20	0.93
0.14	0.93	0.16	0.92
0.09	0.94	0.16	0.92

Conclusions/Future Work

- Super Learner analysis of electronic adherence data classified viral failure with reasonable accuracy in a highly heterogeneous population of HIV infected individuals
- This method could potentially be combined with real time monitoring to target clinical visits, viral load testing, and referral to adherence intervention to individuals at risk of failure
- Ongoing work: comparison of targeted monitoring strategies as compared to standard of care (costs and delay in rebound detection)

Acknowledgements

UC Berkeley Biostatistics

- Mark van der Laan
- Anna Decker
- Erin LeDell
- Varada Sarovar
- Joshua Schwab

School of Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY

MACH 14

- Robert Gross
- Ira Wilson
- Carol Golin
- Nancy Reynolds
- Robert Remien
- Kathleen Goggin
- Jane Simoni
- Marc Rosen
- Honghu Liu
- David Bangsberg

The Doris Duke Charitable Foundation

DORIS DUKE CHARITABLE FOUNDATION

- Clinical Scientist Development Award (mentor David Bangsberg)

Thank You!

School of Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY