



Published in final edited form as:

J Acquir Immune Defic Syndr. 2015 May 1; 69(1): 109–118. doi:10.1097/QAI.0000000000000548.

Super learner analysis of electronic adherence data improves viral prediction and may provide strategies for selective HIV RNA monitoring

Maya L. Petersen, MD, PhD¹, Erin LeDell, MA¹, Joshua Schwab, MA¹, Varada Sarovar, MA¹, Robert Gross, MD², Nancy Reynolds, RN, PhD³, Jessica E. Haberer, MD⁴, Kathy Goggin, PhD⁵, Carol Golin, MD⁶, Julia Arnsten, MD, MPH⁷, Marc Rosen, MD⁸, Robert Remien, PhD⁹, David Etoori, MPH¹, Ira Wilson, MD¹⁰, Jane M. Simoni, PhD¹¹, Judith A. Erlen, RN, PhD¹², Mark J. van der Laan, PhD¹, Honghu Liu¹³, and David R Bangsberg, MD, MPH¹⁴

¹School of Public Health, University of California, Berkeley, Division of Biostatistics, Berkeley, CA

²Perelman School of Medicine, University of Pennsylvania, Departments of Medicine (Infectious Disease) and Epidemiology, Philadelphia, PA

³School of Nursing, Yale University, New Haven, CT

⁴Massachusetts General Hospital, Center for Global Health; Harvard Medical School, Boston, MA

⁵Health Services and Outcomes Research, Children's Mercy Hospitals and Clinics, University of Missouri - Kansas City Schools of Medicine and Pharmacy, Kansas City, MO

⁶Departments of Health Behavior and Medicine, School of Medicine and Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC

⁷Department of Medicine, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY

⁸School of Medicine, Yale University, Dept. of Psychiatry, New Haven, CT

⁹HIV Center for Clinical and Behavioral Studies; NY State Psychiatric Institute and Columbia University, New York, NY

¹⁰Department of Health Services, Policy & Practice, Brown University School of Public Health, Providence, RI

¹¹University of Washington, Department of Psychology, Seattle, WA

¹²University of Pittsburgh, School of Nursing, Department of Health and Community Systems, Pittsburgh, PA

Correspondence: Maya Petersen, Assistant Professor of Biostatistics and Epidemiology, University of California, Berkeley School of Public Health, 101 Haviland Hall, Berkeley, CA 94720-1358, mayaliv@berkeley.edu.

Conflict of Interest: The authors report no conflicts of interest.

Author contributions:

MP and DB conceived the study, planned and supervised the analysis, interpreted results, and wrote the manuscript. EL, JS, VS, and DE implemented the analysis and contributed to interpretation and writing. MP, MVL, and EL developed, validated, and implemented analytic methods used, and contributed to interpretation and writing. RG, NR, JH, KG, CG, JA, MR, RR, IW, JS, JE, and HL contributed to data collection, design and interpretation of analyses, and writing.

¹³School of Dentistry, University of California, Los Angeles, Los Angeles, CA

¹⁴Massachusetts General Hospital, Center for Global Health; Harvard School of Public Health, Department of Global Health and Population, Boston, MA

Abstract

Objective—Regular HIV RNA testing for all HIV positive patients on antiretroviral therapy (ART) is expensive and has low yield since most tests are undetectable. Selective testing of those at higher risk of failure may improve efficiency. We investigated whether a novel analysis of adherence data could correctly classify virological failure and potentially inform a selective testing strategy.

Design—Multisite prospective cohort consortium.

Methods—We evaluated longitudinal data on 1478 adult patients treated with ART and monitored using the Medication Event Monitoring System (MEMS) in 16 United States cohorts contributing to the MACH14 consortium. Since the relationship between adherence and virological failure is complex and heterogeneous, we applied a machine-learning algorithm (Super Learner) to build a model for classifying failure and evaluated its performance using cross-validation.

Results—Application of the Super Learner algorithm to MEMS data, combined with data on CD4+ T cell counts and ART regimen, significantly improved classification of virological failure over a single MEMS adherence measure. Area under the ROC curve, evaluated on data not used in model fitting, was 0.78 (95% CI: 0.75, 0.80) and 0.79 (95% CI: 0.76, 0.81) for failure defined as single HIV RNA level >1000 copies/ml or >400 copies/ml, respectively. Our results suggest 25–31% of viral load tests could be avoided while maintaining sensitivity for failure detection at or above 95%, for a cost savings of \$16–\$29 per person-month.

Conclusions—Our findings provide initial proof-of-concept for the potential use of electronic medication adherence data to reduce costs through behavior-driven HIV RNA testing.

Keywords

HIV; Adherence; Antiretroviral Therapy; Virological failure; HIV RNA monitoring; Medication Event Monitoring System; Super Learner

Introduction

Testing HIV patients' plasma HIV RNA level every three months is the standard of care in resource-rich settings and is used to alert providers to the potential need for enhanced adherence interventions and/or the need to change a failing antiretroviral therapy (ART) regimen before accumulation of resistance mutations, disease progression, and death.^{1–3} Routine serial HIV RNA testing for all patients on ART, however, is expensive and the majority of tests in patients on stable regimens yield undetectable HIV RNA.^{4–6}

If sufficiently sensitive, a selective testing approach in which plasma HIV RNA is measured only when screening criteria suggest increased risk of virological failure might be used to reduce expense related to HIV RNA testing. However, several screening rules proposed to

date, based on either clinical data alone or some combination of clinical data and self-reported adherence data, had low sensitivity (20–67%), particularly in validation populations.^{4–6} Pharmacy refill data were reported to classify failure with greater accuracy than self-reported adherence data,⁷ suggesting that alternate approaches to measuring adherence may improve a selective HIV RNA testing strategy.

Previous studies examining associations between adherence and virological failure have generally focused on a single summary of adherence data, such as average adherence over some interval preceding HIV RNA assessment.^{7–12} However, adherence patterns are myriad and the association of any one pattern with virological failure is unclear.⁸ Further, measurement of adherence, whether based on self-report, pill count, or electronic medication container opening, is inherently imperfect. Finally, electronic adherence monitoring methods, some of which transmit data in real time, are becoming more widely available. In summary, it is unknown how best to combine adherence and clinical data to predict virological failure.^{12,14–17} To address these challenges, we built prediction models for virological failure using Medication Event Monitoring System (MEMS) and clinical data analyzed with Super Learner, a data-adaptive algorithm based on cross-validation (i.e., multiple internal data splits).^{18,19}

We investigated the potential for pill container openings recorded using MEMS to correctly classify virological failure in a large clinically and geographically heterogeneous population of HIV patients treated with ART in the United States. We investigated 1) the extent to which a machine learning method (Super Learner) applied to MEMS adherence and clinical data improved classification of failure beyond a single time-updated MEMS adherence summary; 2) the extent to which addition of MEMS to basic clinical data improved classification of failure; and, 3) the potential for the resulting risk score to reduce frequency of HIV RNA measurements while detecting at least 95% virological failures.

Methods

Patient population and outcome

We studied HIV positive patients in the Multisite Adherence Collaboration on HIV-14 (MACH14) who underwent ART adherence monitoring with MEMS between 1997 and 2009. MEMS monitoring consists of a date and time stamp recorded electronically with each pill container opening, subsequently downloaded to a database via USB. Details on the MACH14 population have been reported elsewhere.²⁰ In brief, 16 studies based in 12 US states contributed longitudinal clinical, MEMS, and HIV RNA data to the consortium. Subjects were eligible for inclusion in our analyses if they had at least two plasma HIV RNA measurements during follow-up, with at least one of the postbaseline measurements preceded by MEMS monitoring in the previous month.

We considered two virological failure outcomes, defined as a single HIV RNA level >400 copies/ml and >1000 copies/ml, respectively, the latter to improve specificity for true failure versus transient elevations (i.e., “blips”²¹) and facilitate comparison with previous studies.^{5,7,22} Baseline (first available) HIV RNA level was used as a predictor variable in some models, as detailed below. Post-baseline HIV RNA levels were included as outcome

variables if they were preceded by MEMS monitoring in the previous month. Follow-up was censored at first detected virological failure after baseline. Subjects could contribute multiple postbaseline HIV RNA tests as outcomes.

Candidate Predictor Variables

The first prediction model (“Clinical”) was built using the following non-virological candidate variables: time since study enrollment, baseline, nadir, and most recent CD4 count, time since most recent CD4 count, ART regimen class (non-nucleoside reverse transcriptase inhibitor, boosted protease inhibitor, non-boosted protease inhibitor, and “other” regimens; Table 1), an indicator of change in regimen class in preceding two months, and time since any regimen change. Study site was excluded because, were the algorithm used in practice, site-specific fits would not generally be available. Additional potential predictor variables, including time since diagnosis and full treatment history, were missing on a large proportion of subjects (Table 1) and were not included as candidate predictors in this analysis.

The second prediction model (“Electronic Adherence/Clinical”) augmented the set of candidate predictors used in the non-virological clinical model with summaries of past MEMS data. MEMS data were used to calculate adherence summaries accounting for daily prescribed dosing over a set of intervals ranging from 1 week to 6 months preceding each outcome HIV RNA test. Dates of reported device non-use (due, for example, to hospitalization or imprisonment) were excluded from adherence calculations.

Adherence summaries included percent doses taken (# of recorded doses/# of prescribed doses), number of interruptions (consecutive days with no recorded doses) of 1, 2, 3, and 4 days duration, variance across days of percent doses taken per day, and versions of these summaries using only weekday and (for all but interruptions of >2 days) only weekend measurements. MEMS summaries for intervals over which no MEMS measurements were made were carried forward as their last observed value, and indicators of carry forward were included in the predictor set. This approach, unlike multiple imputation, reflects the data that would be available to inform a testing decision for a given patient. Number of drugs monitored, drug and drug class monitored, and number of days monitored with MEMS during each interval were also included.

Comparison of “Clinical” and “Electronic Adherence/Clinical” model performance evaluated the added value of MEMS data. We also investigated further gains from adding a single baseline HIV RNA measurement (“Baseline HIV RNA/Electronic Adherence/Clinical”) and, because extent of past virological suppression can reflect adherence behavior and impact likelihood of resistance, full HIV RNA history (“HIV RNA/Electronic Adherence/Clinical”) to the set of candidate predictor variables. The latter included most recent HIV RNA value, time since most recent HIV RNA test, number of previous tests, and an indicator of regimen switch since last test.

Construction of Prediction Models

The Super Learner algorithm was applied to each set of candidate variables in turn to build four prediction models for each failure definition (HIV RNA level > 400 copies/ml and >

1000 copies/ml). Super Learner is described in detail elsewhere.^{18,19} In brief, we specified a library of candidate prediction algorithms that included Lasso regression,²⁴ preliminary Lasso screening combined with a generalized additive model,^{25,26} a generalized boosted model,²⁷ multivariate adaptive polynomial spline regression,²⁸ and main term logistic regression. For comparison, percent of prescribed doses taken over the past 3 months was evaluated as a single time-updated predictor. To achieve a more balanced dataset and reduce computation time, all HIV RNA levels above failure threshold and an equal number of randomly sampled HIV RNA levels below threshold were used to build the models, with weights used to correct for sampling. The ability of this approach to build predictors with comparable or improved performance on the full sample is supported by theory and prior analyses,²⁹ and was verified for this dataset using selected prediction models. Ten-fold cross-validation was used to choose the convex combination of algorithms in the library that achieved the best performance on data not used in model fitting, with the negative log likelihood used as loss function (i.e. performance metric).³⁰

Evaluation of Prediction Models

The performance of the Super Learner prediction model was evaluated using a second level of cross-validation, ensuring that performance was evaluated with data not used to build the model. The data were split into ten folds (partitions) of roughly equal size, with all repeated measures from a given individual contained in a single fold. The Super Learner algorithm was run on each training set (containing nine of the ten folds) and the performance of the selected model was evaluated with the corresponding validation set, which consisted exclusively of subjects not used to build the prediction model. Cross-validated performance measures averaged performance across validation sets.

Area under the Receiver Operating Characteristic Curve (AUROC) was used to summarize classification performance across the range of possible cutoffs for classifying a given test based on predicted failure probability. Standard error estimates, accounting for repeated measures on an individual, were calculated based on the influence curve.³¹ For each of a range of cutoffs, we also calculated sensitivity, specificity, negative and positive predictive value, and the proportion of HIV RNA tests for which the predicted probability of failure was below the cutoff. We refer to the latter metric as “capacity savings” because it reflects, for a given cutoff, the proportion of HIV RNA tests that would have been avoided under a hypothetical selective testing strategy in which HIV RNA tests were ordered only when a subject’s predicted probability of failure on a given day on which an HIV RNA test was scheduled was greater than the cutoff.

We estimated capacity savings corresponding to a range of sensitivities. For example, for each validation set we selected the largest cutoff for which sensitivity was $\geq 95\%$, calculated the proportion of HIV RNA tests that had a predicted probability less than this cutoff, and averaged these proportions across validation sets. This provided an estimate of the proportion of HIV RNA tests that could have been avoided under a hypothetical selective testing rule chosen to detect at least 95% of failures without additional delay.

Costing

We conducted a basic analysis to provide a rough estimate of the cost per month at which an electronic adherence monitoring system would remain cost neutral. For each candidate prediction model, we first estimated gross cost savings that would be achieved with cutoff chosen to maintain sensitivity $\geq 95\%$ as the number of post-baseline HIV RNA tests below cutoff times the estimated combined cost of an outpatient visit and HIV RNA test. We used an estimated unit cost for an outpatient primary care visit based on the 2014 Medicare National Physician Fee Schedule (\$90–\$140 for CPT-4 code 99214) and unit cost for an HIV RNA test of \$50–\$90. We divided gross cost savings by total person time under MEMS follow-up. Person time was calculated from beginning of MEMS monitoring to the minimum of last post-baseline HIV RNA test date and end of MEMS monitoring.

Analyses were performed using MatLab and R.^{32,33}, including the ROCR, SuperLearner, and cvAUC R packages.^{34–36}

Results

Sample Characteristics

Of the 2835 patients in the MACH14 cohort, 1478 patients met inclusion criteria and contributed a total of 3096 post-baseline HIV RNA tests to prediction of failure defined using a threshold of 1000 copies/ml. Consecutive postbaseline HIV RNA tests were a median of 45 days apart (IQR:28, 99; Table 1). Baseline and final HIV RNA tests were a median of 163 days (IQR 82, 282) apart. When failure was defined as >400 copies/ml, 2751 post-baseline HIV RNA tests met inclusion criteria; baseline and final HIV RNA tests were a median of 143 days (IQR 66, 273) apart.

At baseline subjects had known of their HIV status for a median of 9.2 years (IQR 4.0, 16.1), and the majority (83%) had past exposure to ART (Table 1). A median of 78 days of MEMS data (IQR 28, 91) were available in the three months preceding HIV RNA tests conducted after the baseline assessment. Over the 3 months preceding each postbaseline HIV RNA test, median percent doses taken, as measured by MEMS, was 84% (IQR 50, 98) and subjects had experienced a median of four interruptions ≥ 24 hours in MEMS events (IQR 1,11). Using a threshold of 1000 and 400 copies/ml, respectively, 20% and 27% of post-baseline HIV RNA tests were failures and 41% and 51% of patients failed.

Classification of Failure

The Super Learner Algorithm applied to each of the three predictor sets that included MEMS data resulted in a higher cross-validated AUROC than MEMS-based measurement of three-month percent doses taken (Figure 1 and Table 2; $p < 0.001$ for each pairwise comparison). The AUROC for classification of failure by three-month percent doses taken was 0.64 (95% CI 0.61, 0.67), and 0.60 (95% CI 0.57, 0.63) for HIV RNA failure thresholds of 1000 and 400 copies/ml, respectively.

With failure defined as >1000 copies/ml, the Super Learner model based on non-virological clinical predictors achieved moderate classification performance (“Clinical” AUROC: 0.71,

95% CI: 0.69, 0.74). Addition of MEMS data significantly improved performance (“Electronic Adherence/Clinical” AUROC: 0.78, 95% CI: 0.75, 0.80, $p < 0.001$). Addition of a single baseline HIV RNA level (“Baseline HIV RNA/Electronic Adherence/Clinical” AUROC: 0.82, 95% CI: 0.79, 0.84) or full HIV RNA history (“HIV RNA/Electronic Adherence/Clinical” AUROC: 0.86, 95% CI: 0.84, 0.88) improved performance further ($p < 0.01$ for each pairwise comparison). With failure defined as >400 copies/ml, corresponding AUROCs were slightly higher (Table 2). Figure 2 shows cross-validated sensitivity, specificity, positive and negative predicted value for each of the Super Learner models across a range of cutoffs.

Selective HIV RNA Testing

Figure 3 shows cross-validated estimated capacity savings for a range of sensitivities. With failure defined as >1000 copies/ml and cutoff chosen to maintain sensitivity $\geq 95\%$, a testing rule based on non-virological clinical predictors alone achieved a capacity savings of 15%. Capacity savings increased with addition of MEMS data to 22%, with further addition of a single baseline HIV RNA level to 31%, and with addition of full HIV RNA history to 41%. Estimated cost savings per person-month MEMS follow-up increased from \$9–\$14 with non-virological clinical data alone, to \$18–\$39 dollars after incorporating MEMS and HIV RNA data (Table 2). Capacity savings at 95% sensitivity were slightly lower with a failure threshold of 400 copies/ml (Table 2).

Discussion

We examined how well electronic adherence (MEMS) and clinical data, analyzed using Super Learner, could classify virological failure (HIV RNA >400 or >1000 copies/ml) in a heterogeneous population of HIV positive patients in the United States. MEMS-based measurement of percent doses taken resulted in poor classification of failure (AUROC 0.60–0.64), as reported by others (AUROC < 0.67).⁸ Super Learner applied to MEMS and non-virological clinical data significantly improved AUROC (0.78–0.79). MEMS data also significantly improved classification of failure over non-virological clinical data alone, while addition of past HIV RNA data improved classification further.

While MEMS does not transmit data in real-time, it provides proof-of-concept that real-time electronic adherence monitoring, which is now available through other devices, could inform frequency and timing of HIV RNA testing.^{13,37,38} In order to investigate the potential for an electronic adherence-based algorithm to reserve HIV RNA testing for patients with non-negligible risk of failure, we evaluated the extent to which testing could have been reduced while detecting almost all failures without additional delay. Using a selective testing approach based on MEMS and clinical data, we estimated that 18–31% of HIV RNA tests could have been avoided (depending on failure definition and inclusion of a single baseline HIV RNA level) while missing no more than 5% of subjects failing on a given date. We were unable to evaluate the extent to which any missed failures would have been detected at a later date under a selective testing strategy because in our data failures that would have been missed were instead detected and responded to. Our analyses therefore censored data at date of first detected failure.

Addition of past HIV RNA data improved capacity savings further to 34–41%. If a selective testing strategy were applied in practice, a reduced set of HIV RNA data would be available to guide real time decision-making (some HIV RNA tests would not have been ordered). Therefore, capacity savings using an algorithm that incorporates full HIV RNA history should be seen as an upper bound.

Our results suggest an immediate gross cost savings of roughly \$16–\$29 per person-month of MEMS use, assuming access to a single baseline HIV RNA measurement. This estimate represents an upper limit for what the health care system might reasonably spend on all components of a real-time adherence monitoring system, which would need to include devices, monitoring staff, and data transmission, while remaining cost neutral. Our cost savings estimate was conservative, however, in that only HIV RNA tests with MEMS monitoring in the prior month could be deferred, and reported interruptions in monitoring were included in total MEMS time. We did not account for costs incurred by changes in second line regimen use, resistance testing, or disease progression, and we assumed that a reduction in HIV RNA testing would save both laboratory and associated clinic visit costs; excluding visit costs would reduce our estimated savings. Our estimates might also differ substantially under different HIV RNA monitoring schedules, in populations with different virological failure rates, and in settings where MEMS was used with fewer interruptions and over longer durations.

Furthermore, we conservatively assumed that real-time monitoring conferred no benefit, and evaluated a hypothetical system under which MEMS could be used to defer regularly scheduled tests, but not to trigger extra or early tests. In practice, real-time electronic adherence monitoring could be used to trigger tests between scheduled testing dates, which while increasing cost could introduce additional potentially significant benefits including opportunities to both detect failure earlier and to prevent it from occurring.³⁹ While frequency of HIV RNA monitoring in our cohort was similar to that seen in standard clinical care, our results should thus be viewed as initial proof-of-concept for future investigation including full cost-effectiveness analysis.

More efficient strategies for HIV RNA testing are especially important in resource-limited settings where standard HIV RNA testing is cost prohibitive. A number of possible strategies for selective HIV RNA testing in resource-limited settings have been proposed. WHO-recommended CD4-based criteria have been repeatedly shown to have poor sensitivity for detecting failure,^{40–43} while a number of alternative clinical and CD4-based rules proposed for selective HIV RNA testing also exhibited moderate to poor sensitivity in validation data sets.^{4–6} However, a recent clinical prediction score that incorporated self-report adherence, regular hemoglobin and CD4 monitoring, and new onset papular pruritic rash achieved more promising performance in Cambodia,^{22,44} as did pharmacy refill data, alone or in combination with CD4 counts, in South Africa.⁷ Robbins et. al. also achieved good classification of failure among US patients using adherence data abstracted from clinical notes, drug and alcohol use, and past appointment history in addition to HIV RNA, CD4 and ART data.⁴⁵ While these studies were conducted using predictor variables not available in the current study, they provide additional proof-of-concept for a strategy of selective HIV RNA testing that incorporates adherence measures. The utility and cost-

effectiveness of incorporating electronic adherence data in such a strategy remains to be determined. We found that a selective testing approach based on CD4 count and ART measures alone could achieve a 14–15% potential capacity savings while maintaining sensitivity at 95%. The performance of our classification rule might have been improved by including more extensive clinical history and self-reporting of non-adherence. Thus, while our study provides insight into classification performance achievable in settings where rich and carefully documented longitudinal clinical data are not available but electronic adherence data are, further investigation of the value added by electronic adherence data in settings with more extensive patient data remains of interest.

Our analyses focused on a machine-learning method (Super Learner) to build optimal predictors of virological failure, and evaluated the extent to which the resulting risk scores could be used to reduce HIV RNA testing while maintaining sensitivity at a fixed level if all patients with risk scores above a cutoff were tested. We focused on capacity savings rather than specificity as any initial false positives would be correctly classified with subsequent HIV RNA testing using this approach. Recently developed methods instead take the risk score and a constraint on testing as given and develop optimal tripartite rules, in which only patients with an intermediate risk score are tested.⁴⁶ Combined application of these approaches is an exciting area of future research.

The patient population in this study included patients with varied clinical histories receiving care in a range of locations throughout the United States; value added by electronic adherence data may vary in different populations. Further, subjects were followed as part of different studies, with distinct protocols for data collection and quality control. We excluded any reported periods of MEMS nonuse from our adherence calculations; however, gaps in MEMS events might still reflect device malfunction or non-use rather than missed doses. Real time reporting of electronic adherence data may improve ability to distinguish between these possibilities through real time patient queries, and thus may improve classification performance achievable with MEMS data still further.^{13,38}

In summary, our results support Super Learner as a promising approach to developing algorithms for selective HIV RNA testing based on the complex data generated by electronic adherence monitoring in combination with readily available clinical variables. A patient's risk of current virological failure, based on time-updated clinical and MEMS data, could be made available to clinicians in real time (e.g., as an automated calculation in an electronic medical record or smart phone application) to help determine whether a clinic visit and/or HIV RNA test is indicated, allowing for personalized testing and visit schedules. Our results provide initial proof-of-concept for the potential of such an approach to reduce costs while maintaining outcomes.

Acknowledgments

Source of Funding: This research was supported by the Doris Duke Charitable Foundation Grant #2011042, CFAR P30-AI50410, and the multi-site adherence collaboration in HIV (MACH14) grant R01MH078773 from the National Institute of Mental Health Office on AIDS. The original grants of individual participating studies are: R01DA11869, R01MH54907, R01NR04749, R01NR04749, R01MH068197, R01DA13826, K23MH01862, K08 MH01584, R01AI41413, R01MH61173, AI38858, AI069419, K02DA017277, R01DA15215, P01MH49548,

R01MH58986, R01MH61695, CC99-SD003, CC02-SD-003 and R01DA015679. JH was supported by K23MH087228.

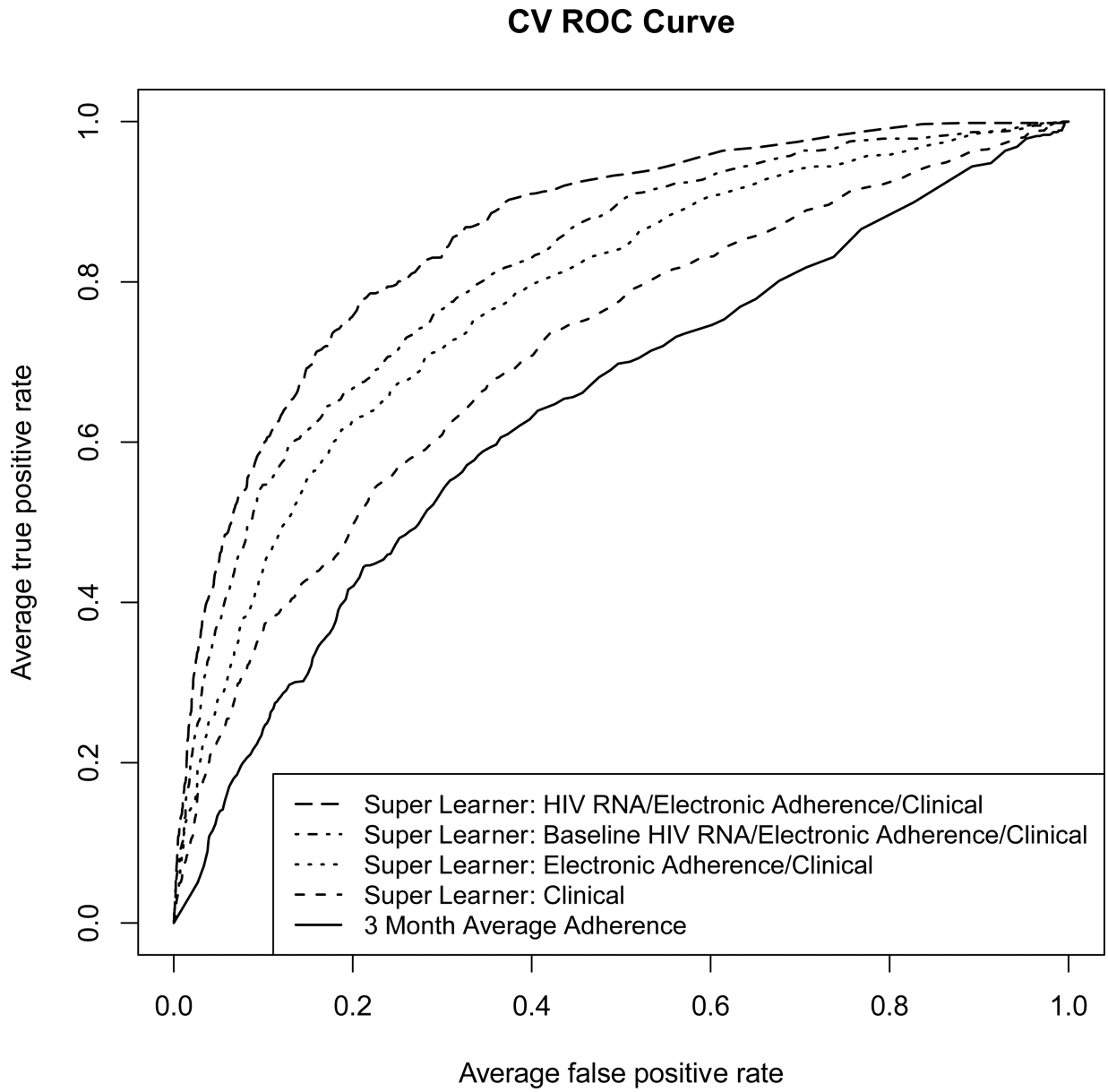
References

1. Hatano H, Hunt P, Weidler J, et al. Rate of viral evolution and risk of losing future drug options in heavily pretreated, HIV-infected patients who continue to receive a stable, partially suppressive treatment regimen. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2006 Nov 15; 43(10):1329–1336. [PubMed: 17051502]
2. Petersen ML, van der Laan MJ, Napravnik S, Eron JJ, Moore RD, Deeks SG. Long-term consequences of the delay between virologic failure of highly active antiretroviral therapy and regimen modification. *AIDS*. 2008 Oct 18; 22(16):2097–2106. [PubMed: 18832873]
3. Barth RE, Wensing AM, Tempelman HA, Moraba R, Schuurman R, Hoepelman AI. Rapid accumulation of nonnucleoside reverse transcriptase inhibitor-associated resistance: evidence of transmitted resistance in rural South Africa. *AIDS*. 2008 Oct 18; 22(16):2210–2212. [PubMed: 18832885]
4. Abouyannis M, Menten J, Kiragga A, et al. Development and validation of systems for rational use of viral load testing in adults receiving first-line ART in sub-Saharan Africa. *AIDS*. 2011 Aug 24; 25(13):1627–1635. [PubMed: 21673555]
5. Meya D, Spacek LA, Tibenderana H, et al. Development and evaluation of a clinical algorithm to monitor patients on antiretrovirals in resource-limited settings using adherence, clinical and CD4 cell count criteria. *Journal of the International AIDS Society*. 2009; 12:3. [PubMed: 19261189]
6. Colebunders R, Moses KR, Laurence J, et al. A new model to monitor the virological efficacy of antiretroviral treatment in resource-poor countries. *The Lancet infectious diseases*. 2006 Jan; 6(1): 53–59. [PubMed: 16377535]
7. Bisson GP, Gross R, Bellamy S, et al. Pharmacy refill adherence compared with CD4 count changes for monitoring HIV-infected adults on antiretroviral therapy. *PLoS medicine*. 2008 May 20.5(5):e109. [PubMed: 18494555]
8. Genberg BL, Wilson IB, Bangsberg DR, et al. Patterns of antiretroviral therapy adherence and impact on HIV RNA among patients in North America. *AIDS*. 2012 Jul 17; 26(11):1415–1423. [PubMed: 22767342]
9. de Boer IM, Prins JM, Sprangers MA, Nieuwkerk PT. Using different calculations of pharmacy refill adherence to predict virological failure among HIV-infected patients. *J Acquir Immune Defic Syndr*. 2010 Dec 15; 55(5):635–640. [PubMed: 21934556]
10. McMahon JH, Manoharan A, Wanke CA, et al. Pharmacy and self-report adherence measures to predict virological outcomes for patients on free antiretroviral therapy in Tamil Nadu, India. *AIDS Behav*. 2013 Jul; 17(6):2253–2259. [PubMed: 23435750]
11. Goldman JD, Cantrell RA, Mulenga LB, et al. Simple adherence assessments to predict virologic failure among HIV-infected adults with discordant immunologic and clinical responses to antiretroviral therapy. *AIDS Res Hum Retroviruses*. 2008 Aug; 24(8):1031–1035. [PubMed: 18724803]
12. Liu H, Golin CE, Miller LG, et al. A comparison study of multiple measures of adherence to HIV protease inhibitors. *Ann Intern Med*. 2001 May 15; 134(10):968–977. [PubMed: 11352698]
13. Haberer JE, Robbins GK, Ybarra M, et al. Real-time electronic adherence monitoring is feasible, comparable to unannounced pill counts, and acceptable. *AIDS Behav*. 2012 Feb; 16(2):375–382. [PubMed: 21448728]
14. Oyugi JH, Byakika-Tusiime J, Ragland K, et al. Treatment interruptions predict resistance in HIV-positive individuals purchasing fixed-dose combination antiretroviral therapy in Kampala, Uganda. *AIDS*. 2007 May 11; 21(8):965–971. [PubMed: 17457090]
15. Parienti JJ, Ragland K, Lucht F, et al. Average adherence to boosted protease inhibitor therapy, rather than the pattern of missed doses, as a predictor of HIV RNA replication. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2010 Apr 15; 50(8):1192–1197. [PubMed: 20210643]

16. Parienti JJ, Das-Douglas M, Massari V, et al. Not all missed doses are the same: sustained NNRTI treatment interruptions predict HIV rebound at low-to-moderate adherence levels. *PLoS one*. 2008; 3(7):e2783. [PubMed: 18665246]
17. Parienti JJ, Massari V, Descamps D, et al. Predictors of virologic failure and resistance in HIV-infected patients treated with nevirapine- or efavirenz-based antiretroviral therapy. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2004 May 1; 38(9):1311–1316. [PubMed: 15127346]
18. Polley, EC.; van der Laan, MJ. Super Learner in Prediction; U.C. Berkeley Division of Biostatistics Working Paper Series; <http://biostats.bepress.com/ucbbiostat/paper266/2010>
19. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007; 6 Article25.
20. Liu H, Wilson IB, Goggin K, et al. MACH14: a multi-site collaboration on ART adherence among 14 institutions. *AIDS Behav*. 2013 Jan; 17(1):127–141. [PubMed: 22864921]
21. Havlir DV, Koelsch KK, Strain MC, et al. Predictors of residual viremia in HIV-infected patients successfully treated with efavirenz and lamivudine plus either tenofovir or stavudine. *Journal of Infectious Diseases*. 2005; 191(7):1164–1168. [PubMed: 15747253]
22. Phan V, Thai S, Koole O, et al. Validation of a clinical prediction score to target viral load testing in adults with suspected first-line treatment failure in resource-constrained settings. *J Acquir Immune Defic Syndr*. 2013 Apr 15; 62(5):509–516. [PubMed: 23334504]
23. Rosenblum M, Deeks SG, van der Laan M, Bangsberg DR. The risk of virologic failure decreases with duration of HIV suppression, at greater than 50% adherence to antiretroviral therapy. *PLoS one*. 2009; 4(9):e7196. [PubMed: 19787058]
24. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*. 1996; 58(1):267–288.
25. Hastie T, Tibshirani R. Generalized Additive Models. *Statistical Science*. 1986; 1(3):297–318.
26. Hastie, T. Generalized additive models. In: Chambers, JM.; Hastie, TJ., editors. *Statistical Models in S*. Boca Raton, FL: Chapman & Hall/CRC Press; 1992. p. 249-304.
27. Friedman JH. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*. 2001; 29(5):1189–1232.
28. Friedman JH. Multivariate adaptive regression splines (with discussion). *The Annals of Statistics*. 1991; 19:1–141.
29. Rose, S.; Fireman, B.; van der Laan, M. Nested Case-Control Risk Score Prediction. In: van der Laan, M.; Rose, S., editors. *Targeted Learning: Causal Inference for Observational and Experimental Data*. New York, NY: Springer; 2011. p. 239-245.
30. van der Laan M, Dudoit S, Keles S. Asymptotic Optimality of Likelihood-Based Cross-Validation. *Statistical Applications in Genetics and Molecular Biology*. 2006; 3(1)
31. LeDell, E.; Petersen, ML.; van der Laan, MJ. Computationally Efficient Confidence Intervals for Cross-validated Area Under the ROC Curve Estimates; U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 312; 2012.
32. MATLAB 8.0 and Statistics Toolbox 8.1. Natick, Massachusetts, United States: The MathWorks, Inc.;
33. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
34. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: Visualizing the performance of scoring classifiers. R package version 1.0-4. 2009 <http://CRAN.R-project.org/package=ROCR>.
35. Polley E, van der Laan MJ. SuperLearner: Super Learner Prediction. R package version 2.0-9. 2012 <http://CRAN.R-project.org/package=SuperLearner>.
36. LeDell E, Petersen ML, van der Laan MJ. cvAUC: Cross-Validated Area Under the ROC Curve Confidence Intervals. R package version 1.0-0. 2013 <http://CRAN.R-project.org/package=cvAUC>.
37. Haberler JE, Kahane J, Kigozi I, et al. Real-time adherence monitoring for HIV antiretroviral therapy. *AIDS Behav*. 2010 Dec; 14(6):1340–1346. [PubMed: 20809380]

38. Haberer JE, Kiwanuka J, Nansera D, et al. Real-time adherence monitoring of antiretroviral therapy among HIV-infected adults and children in rural Uganda. *AIDS*. 2013; 27(13):2166–2168. [PubMed: 23751260]
39. Freedberg KA, Hirschhorn LR, Schackman BR, et al. Cost-effectiveness of an intervention to improve adherence to antiretroviral therapy in HIV-infected patients. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2006; 43:S113–S118.
40. Westley BP, DeLong AK, Tray CS, et al. Prediction of treatment failure using 2010 World Health Organization Guidelines is associated with high misclassification rates and drug resistance among HIV-infected Cambodian children. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012 Aug; 55(3):432–440. [PubMed: 22539664]
41. Ferreyra C, Yun O, Eisenberg N, et al. Evaluation of clinical and immunological markers for predicting virological failure in a HIV/AIDS treatment cohort in Busia, Kenya. *PLoS one*. 2012; 7(11):e49834. [PubMed: 23185450]
42. Ingole N, Mehta P, Pazare A, Paranjpe S, Sarkate P. Performance of immunological response in predicting virological failure. *AIDS Res Hum Retroviruses*. 2013 Mar; 29(3):541–546. [PubMed: 23137294]
43. Keiser O, MacPhail P, Boule A, et al. Accuracy of WHO CD4 cell count criteria for virological failure of antiretroviral therapy. *Tropical medicine & international health : TM & IH*. 2009 Oct; 14(10):1220–1225. [PubMed: 19624478]
44. Lynen L, An S, Koole O, et al. An algorithm to optimize viral load testing in HIV-positive patients with suspected first-line antiretroviral therapy failure in Cambodia. *J Acquir Immune Defic Syndr*. 2009 Sep 1; 52(1):40–48. [PubMed: 19550349]
45. Robbins GK, Johnson KL, Chang Y, et al. Predicting virologic failure in an HIV clinic. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2010 Mar 1; 50(5):779–786. [PubMed: 20121574]
46. Liu T, Hogan JW, Wang L, Zhang S, Kantor R. Optimal Allocation of Gold Standard Testing under Constrained Availability: Application to Assessment of HIV Treatment Failure. *J Am Stat Assoc*. 2013 Jan 1; 108(504):1173–1188. [PubMed: 24672142]

a.



b.

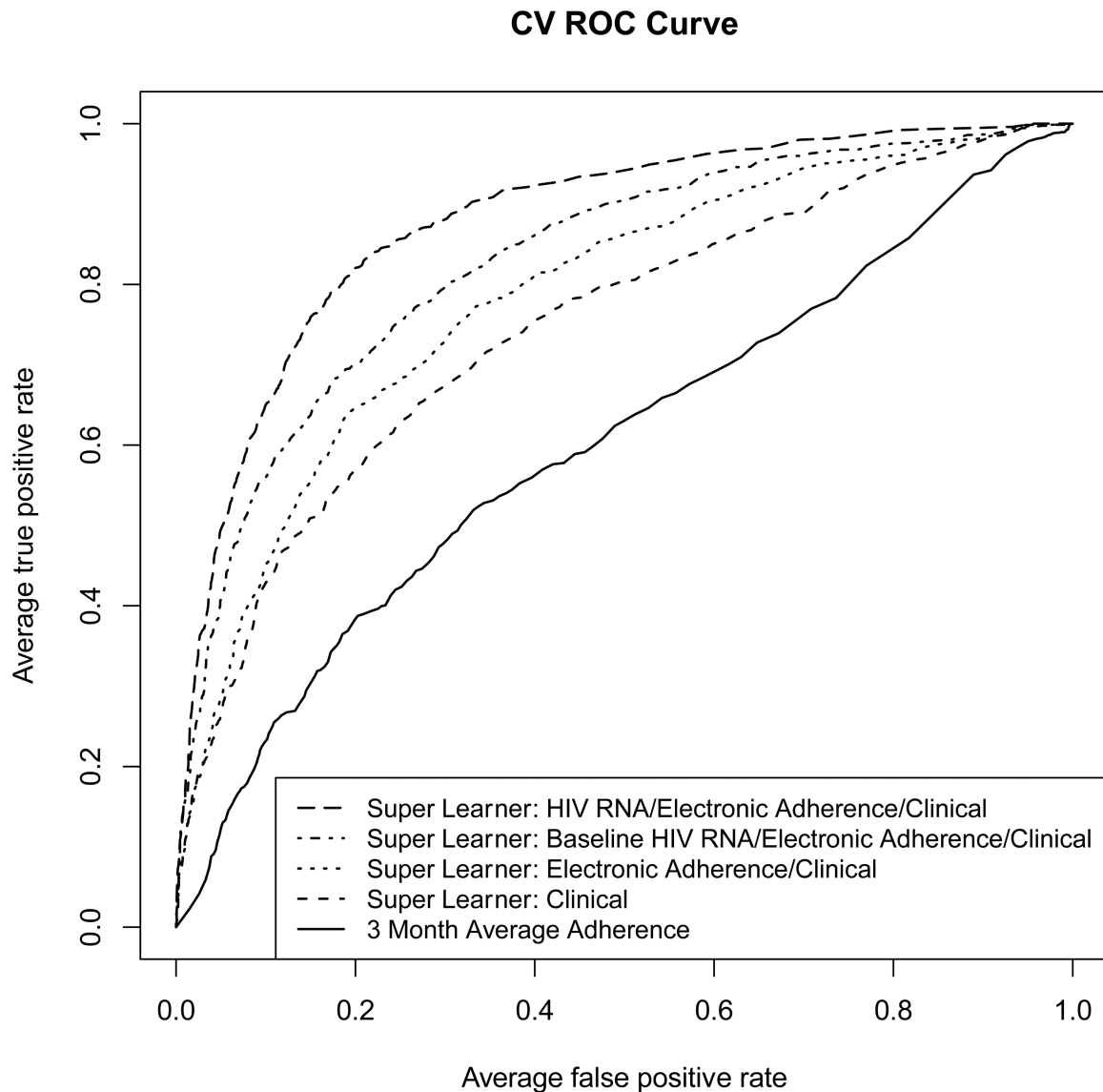
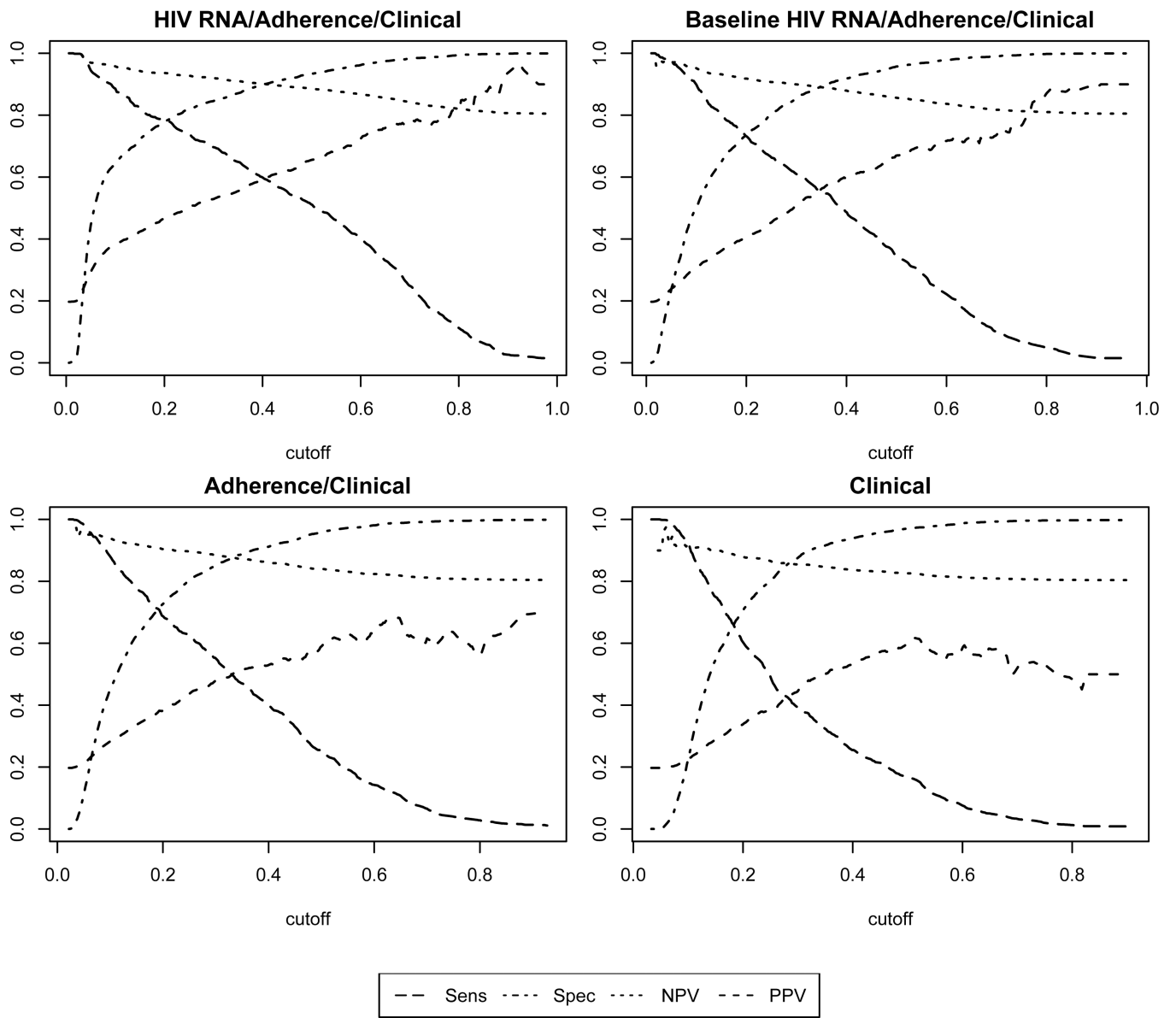


Figure 1. Cross-validated ROC Curves for classification of virological failure using four Super Learner prediction models and three-month “average adherence” (percent prescribed doses recorded by MEMS).

1a. Failure defined as HIV RNA level >1000 copies/ml

1b. Failure defined as HIV RNA level >400 copies/ml

a.



b.

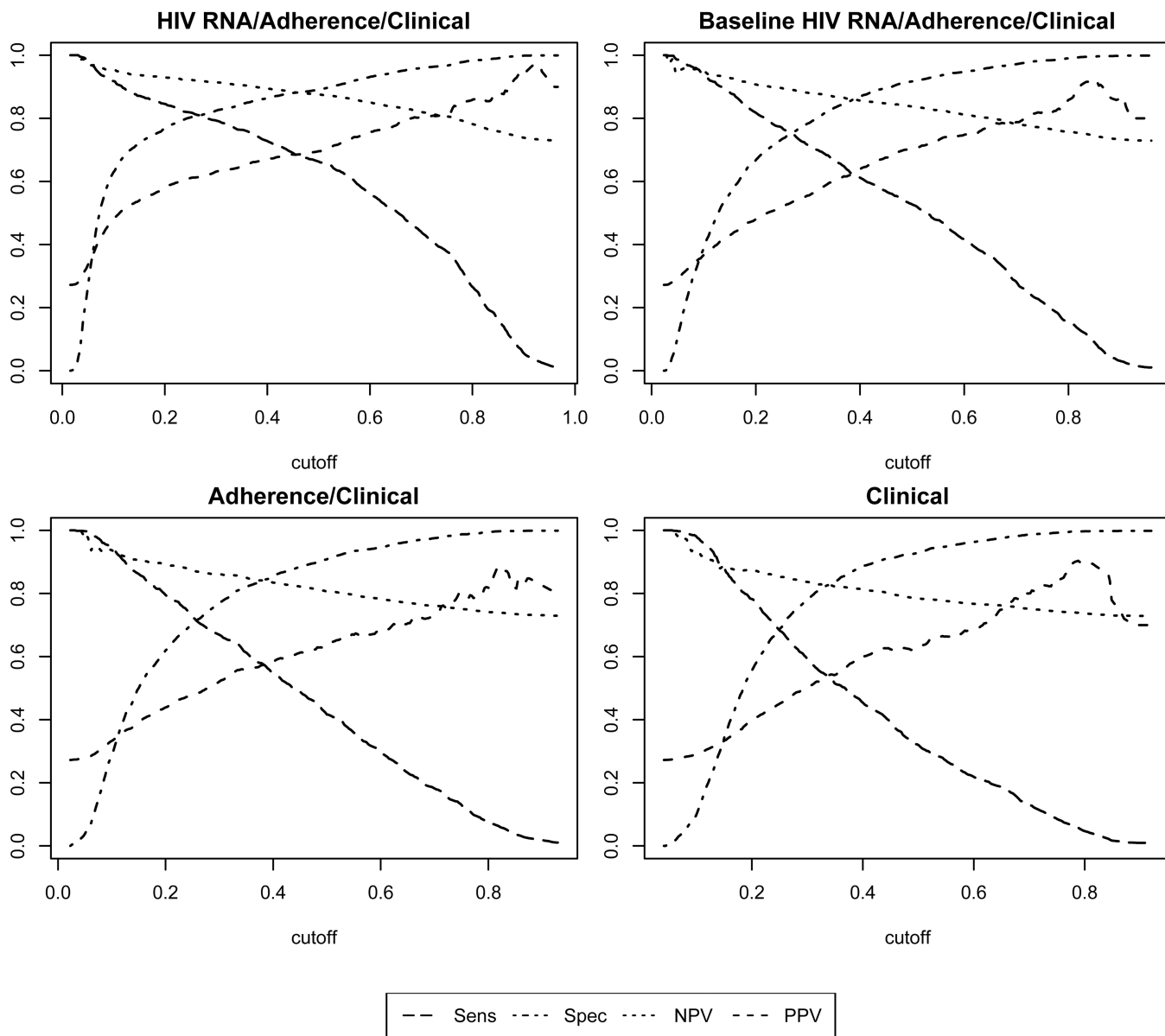


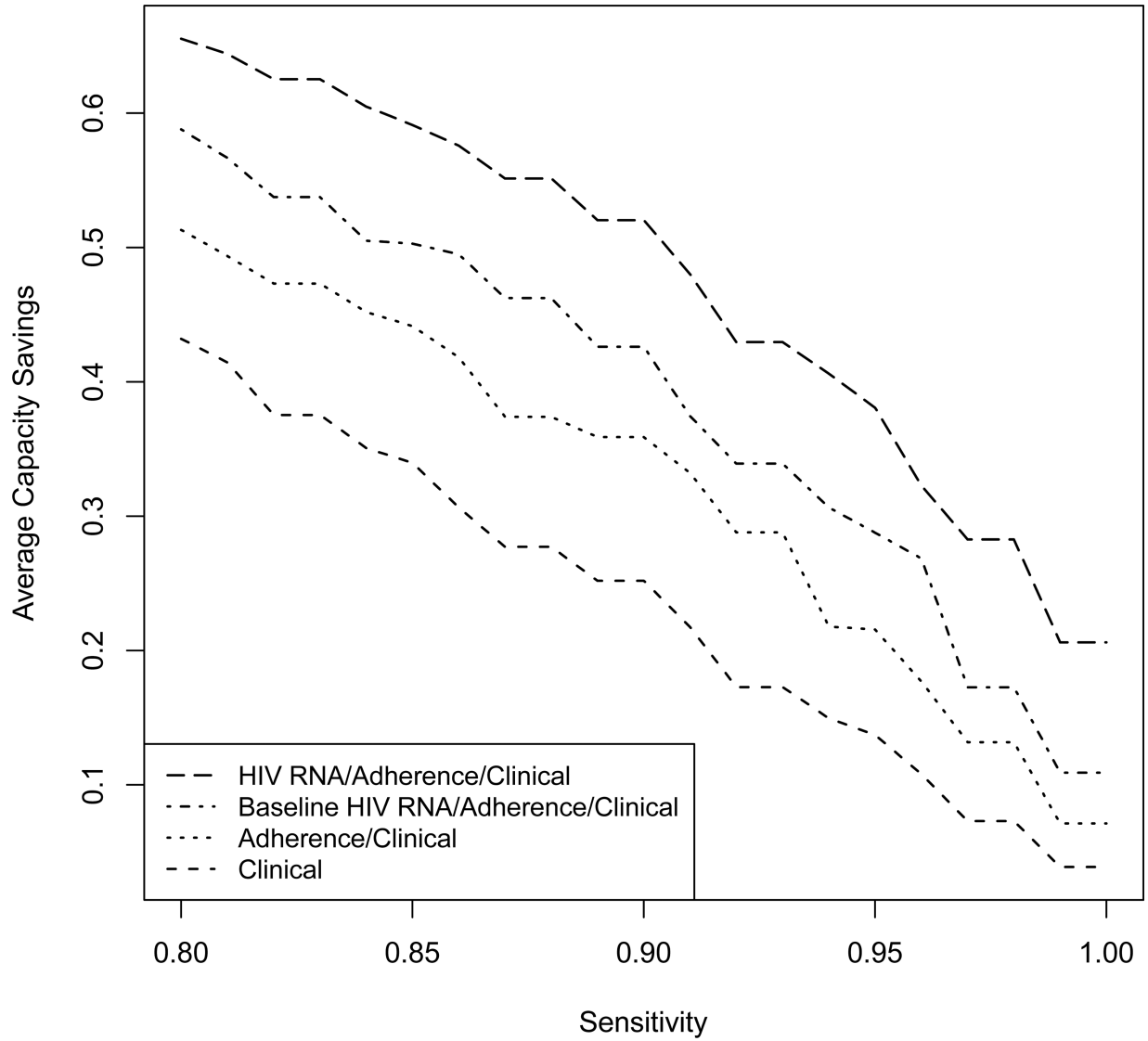
Figure 2. Cross validated performance measures for Super Learner-based classification of failure1

2a. Failure defined as HIV RNA level >1000 copies/ml

¹Sens=Sensitivity; Spec= Specificity; NPV=negative predictive value; PPV=positive predictive value

2b. Failure defined as HIV RNA level >400 copies/ml

a.



b.

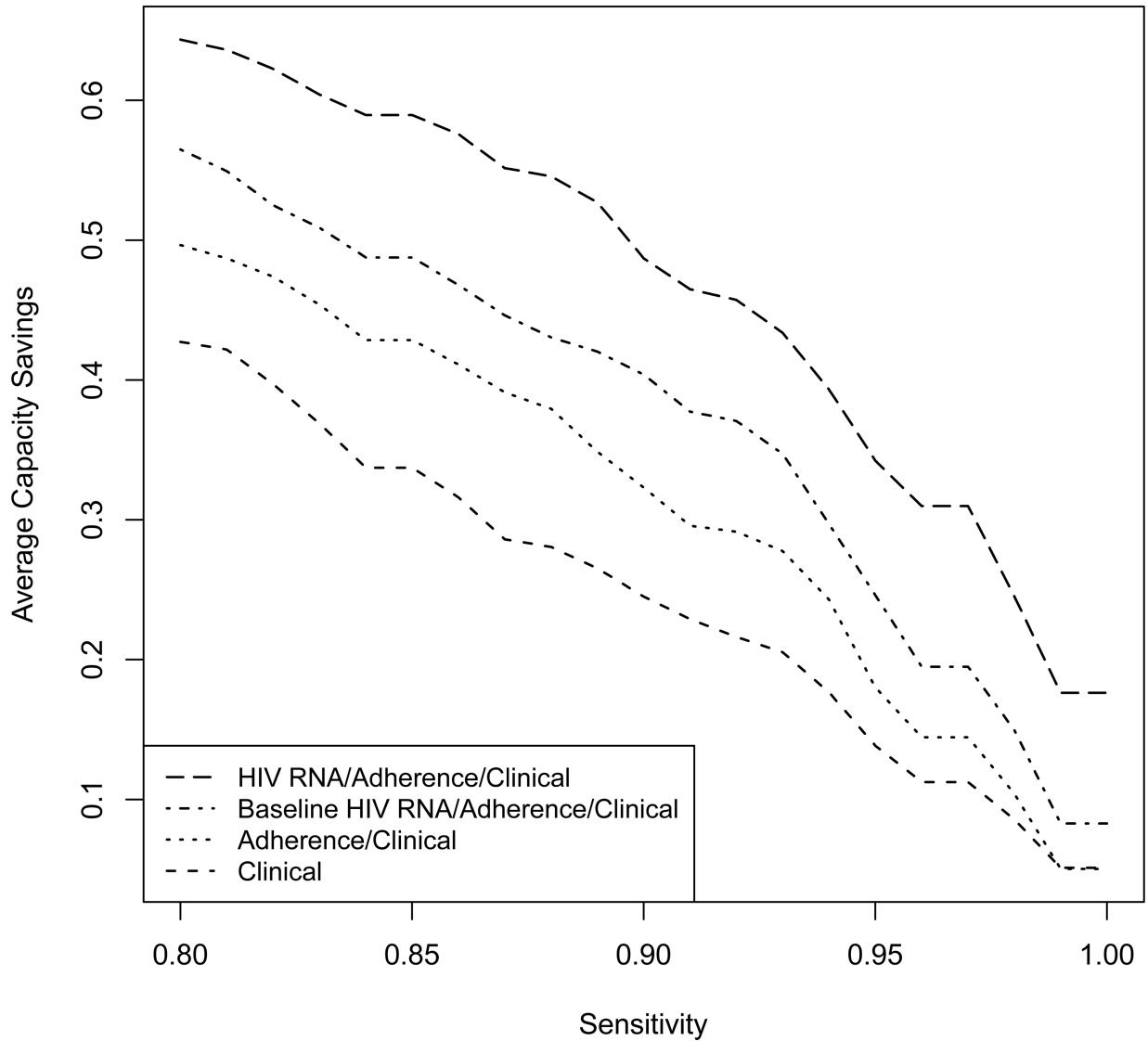


Figure 3. Estimated capacity savings (proportion of HIV RNA tests avoided) achievable by a selective testing strategy with cutoff chosen to maintain a range of sensitivities.

3a. Failure defined as HIV RNA level >1000 copies/ml

3b. Failure defined as HIV RNA level >400 copies/ml

Table 1Sample Characteristics¹

Baseline Variables²	
Age (years) ³	Median 41 (IQR: 36, 47)
Duration of known HIV status at time of enrollment (years) ⁴	Median 9.2 (IQR: 4.0, 16.1)
HIV RNA level (copies/ml)	Median 336.5 (IQR: 50, 2600)
CD4+ T cell count (cells/ μ l)	Median 295 (IQR: 143, 444)
Male ⁵	N=855 (64.7%)
Man who has sex with men ⁶	N=499 (42.7%)
Injection drug use before HIV diagnosis ⁷	N=186 (19.3%)
ART Naïve at time of enrollment ⁸	N=217 (16.9%)
Time-varying Variables⁹	
HIV RNA level>1000 copies/ml	N=607 (19.6%)
Most recent CD4+ T cell count (cells/ μ l)	Median 338 (IQR: 176, 500)
Time since most recent CD4+ T cell count (days)	Median 142 (IQR: 79, 274)
Time since most recent HIV RNA (days)	Median 45 (IQR: 28, 99)
Days monitored using MEMS in previous 3 months	Median 78 (IQR: 28, 91)
Percent doses taken (%) in previous 3 months	Median 84 (IQR: 50, 98)
Number of interruptions \geq 24 hours in previous 3 months	Median 4 (IQR: 1, 11)
Number of interruptions \geq 72 hours in previous 3 months	Median 0 (IQR: 0, 1)
NNRTI Regimen	N=806 (26.0%)
Non-boosted PI Regimen	N=553 (17.9%)
Boosted PI Regimen	N=901 (29.1%)
Other ¹⁰	N=836 (27%)

¹ For analysis of failure=HIV RNA>1000 copies/ml; results for failure=HIV RNA>400 copies/ml differ slightly due to censoring at first failure.

² Among 1478 patients meeting inclusion criteria

³ 55 (10.5%) missing age

⁴ 166 (11.2%) missing duration of known HIV status

⁵ 157 (10.6%) missing sex

⁶ 310 (21.0%) missing whether a man who has sex with men

⁷ 510 (34.6%) missing injection drug use

⁸ 192 (13.0%) missing whether ART naïve at time of enrollment

⁹ Summarized over 3096 postbaseline HIV RNA tests meeting inclusion criteria.

¹⁰ Included regimens with two nucleoside reverse transcriptase inhibitors but no other reported drug, regimens based on fusion inhibitors, and unspecified/unknown regimens.

Cross-validated area under the ROC curve (AUROC), capacity savings, and cost savings for prediction models.

Table 2

Candidate model	Failure= HIV RNA Level >1000 copies/ml			Failure=HIV RNA Level >400 copies/ml		
	AUROC (95% CI) ^{1/1}	Capacity Savings ^{1/2}	Cost Savings (US Dollars per Person-Month): Low ^{1/3} , High ^{1/4}	AUROC (95% CI) ^{1/1}	Capacity Savings ^{1/2}	Cost Savings (US Dollars per Person-Month): Low ^{1/3} , High ^{1/4}
Average Adherence: Percent doses taken over past 3 months	0.64 (0.61, 0.67)	18.9%	\$11.05, \$18.15	0.60 (0.57, 0.63)	18.8%	\$11.94, \$19.61
Super Learner: Clinical	0.71 (0.69, 0.74)	15.0%	\$8.74, \$14.36	0.75 (0.72, 0.77)	13.8%	\$8.77, \$14.42
Super Learner: Electronic Adherence/Clinical	0.78 (0.75, 0.80)	21.8%	\$12.73, \$20.91	0.79 (0.76, 0.81)	18.0%	\$11.42, \$18.76
Super Learner: Baseline HIV RNA/Electronic Adherence/Clinical	0.82 (0.79, 0.84)	30.6%	\$17.91, \$29.42	0.83 (0.81, 0.85)	24.6%	\$15.60, \$25.63
Super Learner: HIV RNA /Electronic Adherence/Clinical	0.86 (0.84, 0.88)	40.6%	\$23.73, \$38.99	0.88 (0.86, 0.89)	34.2%	\$21.70, \$35.65

^{1/1} p-values for all pairwise comparisons within a given failure definition <0.01

^{1/2} Proportion of tests below cutoff, with cutoff chosen to maintain 95% sensitivity.

^{1/3} Unit Cost \$90 per visit +\$50 per HIV RNA test

^{1/4} Unit Cost \$140 per visit +\$90 per HIV RNA test