Machine-Learning-Guided Pre-Visit Outreach to Patients at High Risk to Miss HIV Visits

CROI 2026 February 22–25 • Denver, Colorado

1076

Mosong B[2], Mwangi A[1,2,8], DeLong AK[1], Bacher I[3], Achilla E[2], Cafu C[2], Kimaina A[2], Nyabuto G[2], Diero L[2,4], Gaitho D[2], Wachira J[2,4], Ashiono E[2,4], Kimaiyo S[2,4], Kantor R[7], Fraser H[1], **Hogan JW[1,2]**

[1]Department of Biostatistics, Brown University, Providence, USA, [2]Academic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya, [3]Department of Global Health, University of Washington, Seattle, USA, [4] Department of Medicine, School of Medicine, College of Health Sciences, Moi University, Eldoret, Kenya, [5] Department of Emergency Medicine, Harvard Medical School, Boston, USA, [6]Department of Medicine, Indiana University School of Medicine, Indianapolis, USA, [7]Department of Medicine, Brown University, Providence, USA, [8]Department of Mathematics, Physics and Computing, School of Science and Aerospace Studies, Moi University, Eldoret, Kenya. **NIH Funding:** R01 AI 167694 and P30 AI 42852

BACKGROUND AND RATIONALE

Lifelong HIV care requires consistent clinic engagement. Identifying patients at higher risk of missing appointments, and providing timely outreach before scheduled visits, may improve attendance and lead to better clinical outcomes. **We analyzed data from the pilot phase of a stepped-wedge cluster randomized trial to evaluate association between risk-prioritized pre-visit outreach and on-time return to clinic (RTC).**

RESEARCH QUESTIONS

- Among those flagged as high risk for missed visit**, is attempted contact by an outreach worker associated with on-time RTC?**
- Among those flagged as high risk and called by outreach worker, **is being successfully reached by outreach worker associated with on-time RTC?**

METHODS

Study Setting (Figure 1)

- **Academic Model Providing Access to Healthcare (AMPATH), Eldoret Kenya**
- **Pilot Data:** 46,409 patient encounters, 8 clinical sites, 1 Oct 2023 to 31 Dec 2024

Risk Prediction and Patient Outreach

- **Developed** and cross-validated a machine-learning-based prediction model on patient encounter data to quantify probability of missing next scheduled clinical visit
- **Installed** model-based EHR-based clinical decision support system at AMPATH
- **Implemented** weekly risk-prioritized pre-visit outreach using AMPATH patient reminder protocol for those at highest risk for a missed clinic visit

Prediction Model Details

- **27-variable ML model** used to predict risk of missed clinic visit using SuperLearner
- Patient is **flagged as 'high risk'** if risk score is top 20%-ile at population or clinic level
- **Features with highest variable importance:** Number of missed visits in last 3 scheduled encounters; days since last viral load; number of visits since enrollment.

Outcome and process indicators (Figure 2)

- **Pilot study outcome**: return to care (RTC) on or before scheduled visit date
- **Process indicators**: flagged as high risk, called / reached by outreach worker

Data Analysis: Multilevel Logistic Regression

- **Outcome variable**: on-time return to care (yes/no)
- **Process variables**: attempted contact **(Model 1),** successful contact **(Model 2)**
- **Adjustment covariates**: ML-based risk score, gender, age, care program, calendar month, days since last appointment, random effects for patient and facility

RESULTS

- Attempted outreach and successful outreach associated with RTC **(Table 1)**
- Pilot study indicates reasonable predictive accuracy for ML model **(Table 2)**
- ML-predicted risk scores can be calculated and displayed at point of care **(Figure 3)**
- Risk scores, outreach capacity vary considerably across clinics **(Table 3, Figure 4)**

Pre-visit outreach to clients at high risk for missed visit may improve likelihood of on-time return to care.

An ongoing stepped-wedge cluster-randomized trial will rigorously evaluate its impact.

Table 1: Multilevel logistic regression: association between outreach process indicators and on-time return to care

Predictor	Model 1 OR (95% CI)	Model 2 OR (95% CI)
Attempted contact	1.22 (1.10 - 1.35)	---
Successful contact	---	1.77 (1.55 - 2.03)
Male gender	0.99 (0.90 - 1.09)	0.95 (0.84 - 1.08)
Age/10	1.00 (0.95 - 1.05)	0.99 (0.93 - 1.06)

Table 2: Predictive accuracy of an ML model of return to care

ML Predictive Accuracy Metrics	Training Phase	Pilot Phase
% visits missed	36%	35%
AUROC	0.72	0.73
PPV	63%	48%

Figure 1: AMPATH Catchment Area

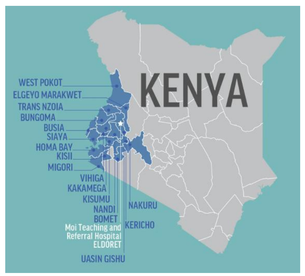
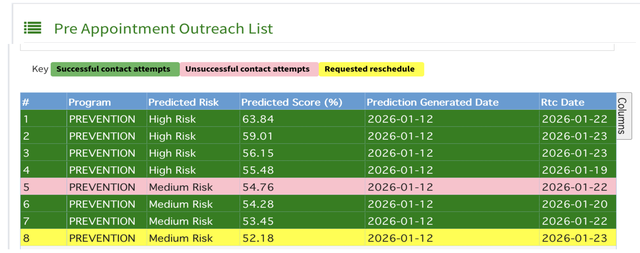


Figure 2: Process and patient flow related to high-risk identification, outreach, and return to care for 46,409 patient visits over 15 months of observation at 8 pilot study clinics.

Figure 3: Weekly list of risk-prioritized patients in AMPATH EHR. Successful follow-ups appeared in green, yellow if they requested a reschedule and pink if outreach failed.



Pre Appointment Outreach List

Key: Successful contact attempts | Unsuccessful contact attempts | Requested reschedule

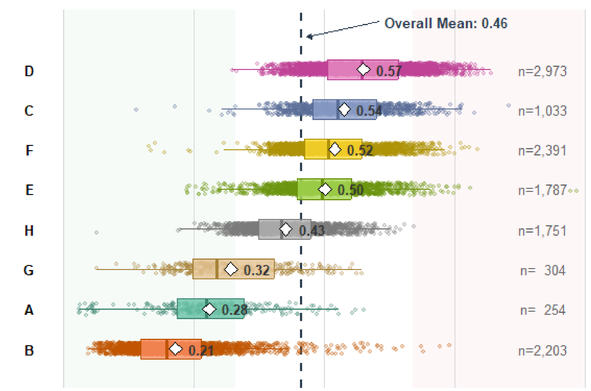
#	Program	Predicted Risk	Predicted Score (%)	Prediction Generated Date	Rtc Date
1	PREVENTION	High Risk	63.84	2026-01-12	2026-01-22
2	PREVENTION	High Risk	59.01	2026-01-12	2026-01-23
3	PREVENTION	High Risk	56.15	2026-01-12	2026-01-23
4	PREVENTION	High Risk	55.48	2026-01-12	2026-01-19
5	PREVENTION	Medium Risk	54.76	2026-01-12	2026-01-22
6	PREVENTION	Medium Risk	54.28	2026-01-12	2026-01-20
7	PREVENTION	Medium Risk	53.45	2026-01-12	2026-01-22
8	PREVENTION	Medium Risk	52.18	2026-01-12	2026-01-23

Columns

Table 3: Summary of Outreach Attempts by Clinic

Clinic	Scheduled Visits	Flagged as high-risk	Called, not reached	Called, reached
A	1,254	254 (20%)	53 (21%)	118 (46%)
B	11,422	2,203 (19%)	400 (18%)	779 (35%)
C	3,092	1,033 (33%)	60 (5.8%)	127 (12%)
D	4,701	2,973 (63%)	654 (22%)	606 (20%)
E	7,259	1,787 (25%)	394 (22%)	349 (20%)
F	9,525	2,391 (25%)	576 (24%)	1,230 (51%)
G	1,349	304 (23%)	8 (2.6%)	166 (55%)
H	7,807	1,751 (22%)	305 (17%)	1,008 (58%)

Figure 4: Risk score distribution by clinic



CHALLENGES & LIMITATIONS

- Maintaining airtime for outreach worker calls
- Managing added work load
- These are pilot data – potential for confounding