Effectiveness of a Machine Learning Risk Score for Personalising Care and Reducing Treatment Interruptions in Nigeria

Executive Summary

Our implementation study (Dec 2023–Oct 2024) in Taraba and Kwara states in Nigeria compared 30 sites using ML-driven IIT risk scores with 19 control sites using a standard mobile app, assessing feasibility, acceptability, and effectiveness.

Case managers spent 76% more time on the list of high-risk patients on average than on the list of low-risk patients. Enrolled patients saw a 4x larger decline in IIT than patients at control facilities. ML-driven scores consistently identified patient risks. Health workers occasionally overrule AI, particularly in high-risk cases (5% overruled). Case managers found the app user-friendly and effective.

Multi-month dispensing (MMD6) reduced IIT but was less effective for high-risk patients, suggesting the need for targeted use. Urban sites had higher IIT rates than rural sites (24% vs. 18%), while private, not-for-profit facilities performed better (18%), highlighting organisational differences in outcomes.

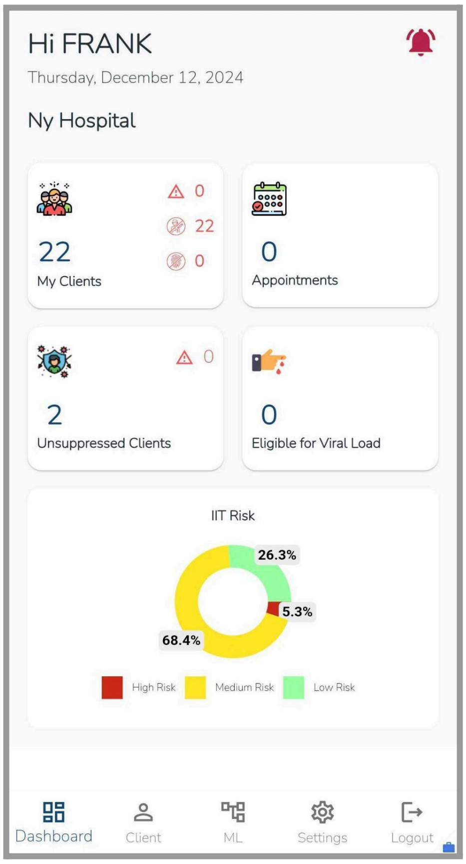
Limitations include a drop in usage post-promotion in March–April in Kwara, due to programme changes. The evaluation of effectiveness was constrained by voluntary usage of the risk score too.

Study Background

Interruption in treatment (IIT) among people living with HIV (PLHIV) in Nigeria significantly impacts viral suppression rates and health outcomes. Despite 96% of diagnosed individuals being on antiretroviral therapy (ART), only 84% achieve viral suppression, partly due to IIT, defined as being more than 28 days late for an appointment. With IIT rates as high as 49% in some regions, current care models struggle to *proactively identify and address at-risk patients*, leading to preventable health setbacks and drug resistance.

Traditional IIT management relies on reactive measures, such as tracing back-to-care strategies, implemented *after* treatment interruptions occur. These resource-intensive approaches fail to personalise care based on patients' risk levels. As a result, resources are overstretched, and high-risk patients may still require more support. A proactive, data-driven approach is needed to identify at-risk patients early and provide tailored interventions before they have an IIT.

USAID supported Jhpiego through the RISE and ACE4 projects; Jhpiego built the Client Management (CM) application (figure right) to provide case managers with easier access to client files. Palindrome Data and Jhpiego



Nigeria developed a machine learning (ML) IIT risk-scoring algorithm using existing electronic medical record (EMR) data to predict IIT per patient. The algorithm was integrated into the CM application to enhance proactive care and improve patient outcomes.

This study evaluated how leveraging EMR data, ML-driven risk scores, risk-based DSD interventions, and digital tools empowered healthcare workers to shift IIT management from reactive to proactive. The integrated approach lays a foundation for efficient, patient-centred HIV care in Nigeria, with outcomes influenced by the interdependence of these components.

Method

This implementation study, conducted from December 2023 to October 2024, assessed the effectiveness of using an ML-based IIT risk scoring model to inform proactive, personalised care for reducing IIT rates and improving ART patient outcomes in Nigeria. The study used a quasi-experimental design, with health facilities randomised by site. All ART patients at the participating facilities were included, with 30 intervention sites utilising a modified CM application featuring the IIT risk score. In comparison, 19 control sites used the standard version of the application, which didn't have any risk scoring.

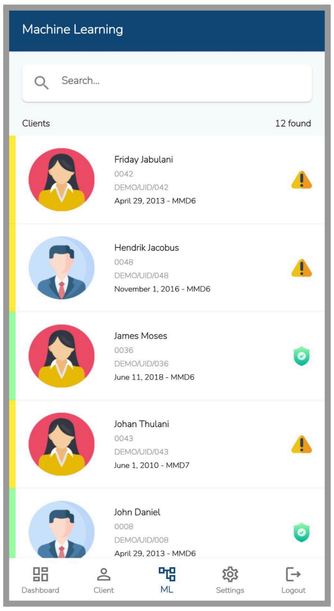
Patrick J McGovern FOUNDATION
jhpiego Johns Hopkins University Affiliate
PALINDROME DATA

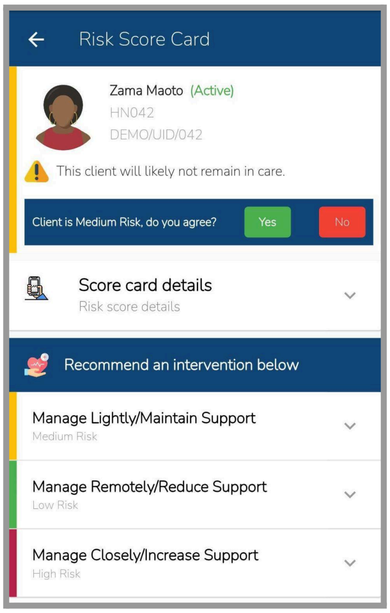
Evaluation Framework

1	**Uptake & Feasibility** of Risk Prediction in EMR-based application
2	**Acceptability** of Informing Case Managers with Risk Prediction
3	**Accuracy** of Risk Prediction
4	**Effectiveness** of Case Management Informed by Risk Prediction on IIT
5	**Differentiation** of Care by Risk

The IIT risk score at intervention sites enabled case managers to deliver tailored care based on individual patient risk levels, aiming to prevent treatment interruptions before they occurred. Control sites maintained routine care practices using the unmodified application, providing a comparative baseline. Both qualitative and quantitative data were collected to evaluate the tool's feasibility, acceptability, and effectiveness, focusing on adoption rates, user experiences, and client health outcomes.

Patients were enrolled in the study when a case manager viewed, responded to, or offered a service via a patient's risk score page on the CM application.



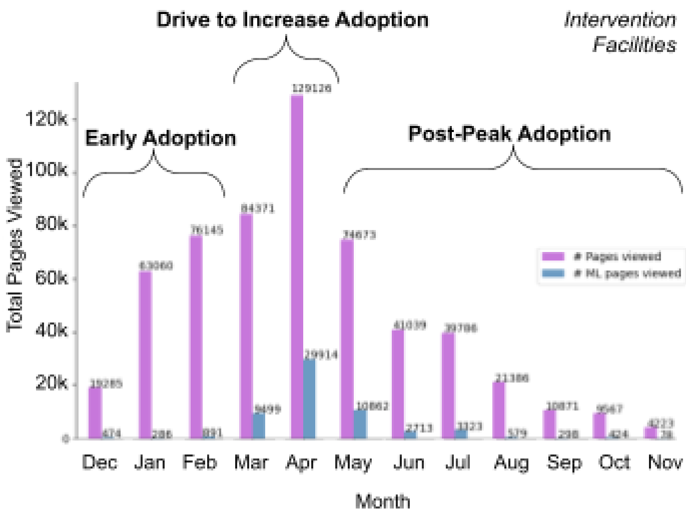


Results

ENROLLED COHORT: The visit data received from the EMR system provided the most recent visit data for all patients; thus, we included visits from 10 December 2023 to mid-November 2024 (N=167,193 visits). Before the study, 50,888 patients visited facilities in either state, compared to 66,308 patients during the study, with Taraba having a three-fold higher caseload than Kwara (pre-study: 50,918 versus 13,305; during the study: 46,737 versus 13,520). **17,999 patients were enrolled across 30 sites.** Most patients were between 30 and 50 years old (58%, N=97,593; an average of 38 years). Most had been on treatment for more than 12 months by their last visit (58%, N=96,230) with at least two drug refills (87%, N=146,138).

In Taraba, IIT rates in the intervention arm were consistently lower than those in the control arm across study periods (pre-study: 21% versus 24%; during the study: 18% versus 26%), while in Kwara, the IIT rates pre-study were higher than is reasonable to report pre-study, likely due to a low number of observations (August to 9 December 2023), and 16% in both arms during the study. Urban facilities had consistently higher IIT rates than rural facilities (24% versus 18%), and private, not-for-profit facilities had lower IIT rates (18%), further indicating that organisational differences may be a key driver of treatment success. Patients 20-30 years old or late for their prior visit were at a higher risk of an IIT (26% and 37%, respectively).

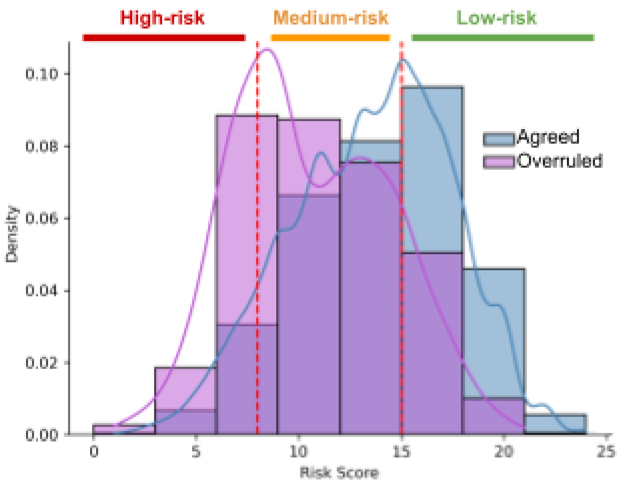
1. UPTAKE & FEASIBILITY



Usage of all risk scoring-related (ML) pages was low during the early adoption of the CM application (December 2023 to February 2024). Retraining and sharing infographics on WhatsApp increased the number of active users and usage during the adoption drive (March to April 2024), but usage slowly declined after these efforts. However, active users declined slower in this period (88 total users and 47 ML users in June versus 50 total and 10 ML in November). Post-peak drops in usage are partly due to changes in Kwara, where changeovers of implementing partners discontinued usage after 5 July 2024.

From monthly surveys, case managers found the application easy to use and effective in saving time and simplifying case management. Users' positive feedback highlighted features like *high-risk client categorisation* and classifying risk using ML. Users identified occasional syncing issues and suggested improvement, including adding biometric security, enhancing data accuracy, and refining client categorisation. While well-received overall, targeted updates could enhance the functionality.

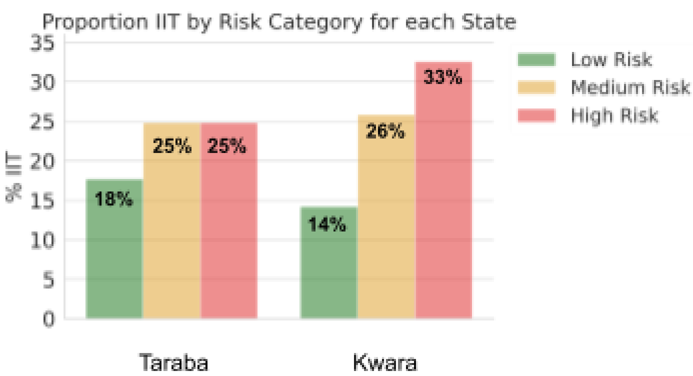
2. **ACCEPTABILITY & IMPLEMENTATION of AI**



Case managers overruled risk scoring for 5% of visits (N=893) and 16% of high-risk visits (N=214). Overall, overruled decisions were less accurate than agreeing with the scores, with an average accuracy of 43% (N=697) compared to 72% (N=7,884) for accepted scores. However, when high-risk scores were overruled, accuracy improved significantly to 77% (N=198). At General Hospital Omu-Aran, the rates of overruling were notably higher, with 30% of cases (N=264) being overruled. High-risk scores were frequently challenged at this facility, with 53% of high-risk assessments (N=48) being overridden. Interestingly, while the facility demonstrated a high accuracy rate of 96% for overruling high-risk scores, it struggled to accept them, achieving only 7% accuracy in these cases.

In summary, the risk score was most accurate when case managers agreed. When they disagreed and overruled a high-risk assessment, the updated risk category was typically more accurate than the original high-risk score.

3. **ACCURACY of AI**

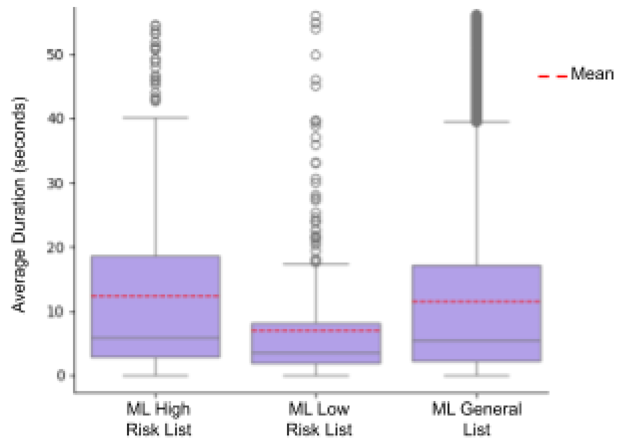


The rate of IITs was 7% higher among visits predicted to be followed by an IIT (27% versus 20%). The low-risk category successfully identified patients at a lower risk of an IIT for both states (Taraba: 18% across scored visits; Kwara: 14% across scored visits). The risk of the medium and high-risk categories was unexpectedly similar in Taraba. Still, both were 1.4x the risk of the low-risk visits, while in Kwara, the medium-risk category was 1.8x, and the high-risk category was 2.3x the risk of the low-risk visits.

4. **DIFFERENTIATION OF CARE**

One key aspect of systematically scoring patients is the ability to measure and visualise whether interventions are going to the right patients at the right time. When developing various care packages with program management staff and intervention designers, we identified interventions most suitable for high-, medium-, and low-risk patients.

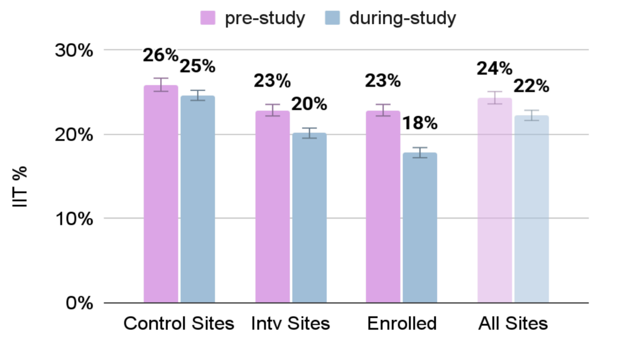
Of the 27 interventions offered via a client's risk scoring page, 10 were slightly differentiated by IIT risk. Multi-month dispensing, fast track, and identification of treatment support partners/groups were preferentially offered to low-risk patients. At the same time, disclosure counselling, contact detail validation, and referrals to third-party services were more likely to be given to medium- and high-risk patients.



Case managers spent 76% more time on the list of high-risk patients on average than on the list of low-risk patients (an average of 12.4 seconds on the high-risk list and 7.0 seconds on the low-risk list).

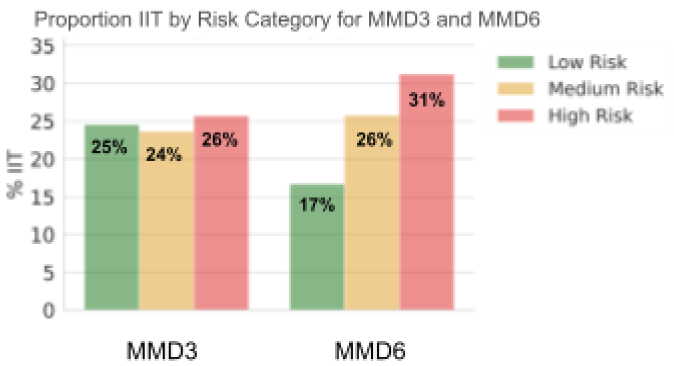
5. EFFECTIVENESS

During the study, 86,635 outcomes were observed across control facilities and enrolled patients. Of the predicted IIT, 22% resulted in an IIT, compared to 24% pre-study (N=75,184). **Risk scoring reduced IIT by 3.8 ± 0.1%** (enrolled patients compared to control facilities; 4.7 ± 0.4% in Taraba and 0.8 ± 0.7% in Kwara). This means **enrolled patients saw a 4x larger decline in IIT than patients at control facilities.**



Note: Enrolled patient data represents (a) pre-study: all visits at intervention sites and (b) during study: only visits at intervention sites where the risk score was engaged with.

6. MULTI-MONTH DISPENSING & RISK



Most patients were prescribed MMD6 (76%, N=184,199), aligning with policy changes from 2023, which provide longer scripts to “lower-risk” patients. Most other patients received MMD3 (23%, N=54,719). Although MMD6 was associated with lower IIT risk than MMD3 (22% for scored patients), it was ineffective for high- and medium-risk patients (31% and 26%, respectively). Contrastingly, MMD3 performed similarly across the risk categories.

7. QUALITATIVE FEEDBACK

Before the app's introduction, case managers often struggled to determine which patients needed urgent support. They relied on lengthy line lists and unprioritised follow-ups, making identifying high-risk individuals difficult. Many admitted that with caseloads of up to 500 patients, those who needed help the most were often overlooked.

The risk scoring addressed this by highlighting patients requiring greater attention. Through this, case managers uncovered critical cases they might have missed:

- One patient flagged as high-risk was found to have stopped treatment after becoming immobile due to an accident. The case manager could step in, provide tailored support, and prevent further deterioration.
- Another patient, also identified as high-risk, revealed struggles with marital issues and depression during follow-ups prompted by the app. Continuous support helped the patient regain stability, and eventually, they began contacting the case manager proactively.

The app has not only made it easier to prioritise care but has also strengthened relationships between case managers and their patients, ensuring that those most in need are not left behind.

Policy Relevance

This study shows evidence for four major takeaways on the use of advanced data operationalisation in large-scale ART programs:

1. Operationalising Advanced Analytics in HIV Care
Successfully implementing machine learning models in Nigeria demonstrates that advanced data-driven decision-making tools, previously limited to the private sector, can be effectively adapted for low-resource healthcare environments. These tools empowered case managers to make timely, data-informed interventions, impacting the health outcomes of over 17,000 patients.

2. New Progress Toward 95-95-95 Goals
By reducing treatment interruptions by 3.8%, the risk-scoring intervention demonstrates new progress on global targets to retain patients in care and achieve viral suppression. Differentiating care based on patient risk profiles enhances resource allocation, ensuring high-risk patients receive intensive support, which is vital for sustaining ART adherence and viral suppression.

3. Improving Healthcare Worker Effectiveness
Before using the application, case managers often struggled to prioritise patients effectively, leading to missed opportunities to support those most in need. The application's ability to flag high-risk patients transformed their workflows, enabling them to focus on those requiring urgent attention while managing lower-risk patients. This reduced the burden of managing large patient cohorts.

4. Health Worker Behavior Change & Personalizing Patient Care
The application shifted how case managers approached patient care, enabling a more personalised differentiated service delivery model. Low-risk patients were decanted to less intensive interventions, such as multi-month dispensing. In contrast, high-risk patients received more time-intensive support, such as enhanced counselling and follow-up referrals. These practices demonstrate how ML can personalise care, optimise resources, and improve patient outcomes.

Scope & Limitations

Some known challenges and limitations were identified:

Not all users participated for the length of the study due to natural workforce turnover and staffing cycles.

Since another implementing partner supported Kwara, the study team faced organisational barriers limiting their influence and access to data. Furthermore, a change in implementing partners in June disrupted Kwara's usage, as new agreements were required.

Since feasibility and acceptability are components of the study's aims, the usage of the CM application and the ML components was optional, meaning that the effectiveness of the scoring cannot be evaluated for every patient at intervention facilities.

Funding Acknowledgement and Ethics

The Patrick J. McGovern Foundation funded the study, and ethics approval was granted by the NHREC of Nigeria (NHREC/01/01/2007) and the Johns Hopkins Bloomberg School of Public Health IRB (26761).