Integrating Artificial Intelligence into a Digital Electronic Community Health Information System

To Strengthen Malaria Rapid Diagnostic Test Reading Interpretation and Correct Prescription Behaviour among Community Health Workers in Ethiopia: A Randomized Trial

JULY 2025

JSI

Authors:

Dr.Tariku Nigatu, Senior Research Advisor and PI
Caitlin Madevu-Matson, Director of MERL and Co-PI
Poppy Farrow, Senior Digital Health Advisor
Dr. Bisrat Misganaw, Data Analyst
Mariawit Shemelis, eCHIS Software Engineer
Zergu Tafese, eCHIS Malaria Advisor
Alex Angel, Senior Editor
Faith Bouchard, Senior Graphic Designer

Thank you

Cover Image:

Colleagues transferring family folder malaria information into eCHIS. Credit: JSI Ethiopia

Table of contents

Introduction

Background

Despite years of progress, malaria remains one of Ethiopia's most persistent and pressing public health threats. As one of Africa's most malaria-endemic nations, a considerable portion of its population (approximately 70%) resides in areas burdened by the disease. In 2020, nearly 5% of Ethiopians lived in high-burden zones, where the annual parasite index (API) exceeded 50 cases per 1,000 people, while another 13% lived in moderately affected areas, with APIs between 10 and 50 per 1,000.[1] The stark reality is that over 10 million malaria cases were reported in Ethiopia in 2020, and the disease remains a leading cause of mortality among children under five.[2] According to the World Health Organization (WHO) changes in climatic conditions, vector resistance to indoor residual sprays (IRS), gene deletion, and political instability are common reasons for recent increases in malaria incidence.

Recognizing the stalled progress, the WHO, in its 2021 update of the Global Technical Strategy for Malaria 2016–2030, advocated for a shift from a "one-size-fits-all" approach to a more data-driven, localized strategy,[3] calling for improved diagnostic accuracy and robust community-level surveillance. In Ethiopia, a key impediment has been the limited use of data in decision-making, which slows the health system's ability to respond effectively to malaria outbreaks and evolving needs. Ethiopia's National Malaria Control Program (NMCP) has intensified efforts in data collection, management, and storage. However, there remain critical needs for targeted and widespread deployment of advanced malaria diagnostic solutions. These solutions hold the potential to deliver:

1. **Real-time insights:** gaining immediate access to test positivity rates and health care provider prescribing behaviors.
2. **Quality assurance:** verifying test results for accuracy.
3. **Enhanced systems:** improving diagnostic capabilities and referral mechanisms.

A cornerstone of Ethiopia's malaria response strategy has been the deployment of roughly 40,000 health extension workers (HEWs) - an all female cadre of frontline community health workers each serving the needs about 1,000 households.[4] This investment was pivotal, coinciding with the introduction of effective malaria treatments like artemisinin-based combination therapy (ACT), artemether-lumefantrine (AL), and rapid diagnostic tests (RDTs).[5] HEWs, who are trained, government-employed, and salaried health workers, instrumental in implementing the health extension worker program, 18 packages of interventions, among Ethiopia's predominantly rural population. They live within or near their communities, fostering trust and accessibility.[6]

Collaborating with local health institutions, HEWs mobilize communities toward crucial malaria activities, including environmental management and distribution and utilization of long-lasting insecticidal nets. They engage the communities in discussions on service utilization to influence health-seeking behaviors.[7] Their embedded presence and responsibility for malaria prevention and control make them vital in providing proper diagnosis, treatment, and disseminating essential health messages.[8] Evidence consistently shows that HEWs significantly enhance community knowledge about malaria and improve health-seeking behaviors.[9]

On average, two HEWs are stationed at each health post, with approximately five health posts that operate under each health center, forming a Primary health care unit. HEWs deliver 18 packages of health promotion, disease prevention, and basic curative services through health posts visits and home visits. This platform of primary health engagement has demonstrably improved maternal and child health, communicable disease control, hygiene, sanitation, and health care-seeking behaviors across Ethiopia.

In 2016, Ethiopia launched an electronic community health information system (eCHIS) on a unified CommCare platform.[10] This digital platform integrates various tools and job aids, enabling real-time, high-quality service delivery, data collection, and data access, alongside automated reporting and monitoring. The eCHIS platform also offers potential for managing HEW performance through enhanced supervision, data utilization, and performance-based incentives. The Ministry of Health (MOH) is scaling eCHIS nationally to improve primary care coverage and quality through data-driven performance.

Currently, at least one module of the eCHIS has been deployed in more than 8,000 health posts, with 3,381 actively providing fully digitized services, including HIV, TB, integrated community case management, nutrition, non-communicable diseases, antenatal care, and prenatal care. Approximately 6,762 HEWs are using eCHIS for at least some of the services they deliver. While not all trained health posts offer every service package (with regional variations), the Government of Ethiopia aims to fully digitize all health posts from registration to service delivery capabilities when funding permits. The malaria module within the eCHIS platform is designed to improve malaria case management at the community level. The MOH together with partners has trained HEWs on this module and the utilization of the module has been growing from time to time. The Government of Ethiopia has made a decision to roll out all module implementations of eCHIS since 2024.

Besides the eCHIS malaria module deployment, the use of malaria RDTs has been expanding in sub-Saharan Africa including Ethiopia. WHO guidelines recommend RDTs for anyone presenting with a fever.[11] Early diagnosis and treatment are crucial for positive health outcomes, and RDTs are a key, accessible intervention.

Despite their widespread availability and affordability, challenges persist. Many countries, including Ethiopia, report issues such as over-reporting of test-positivity rates, misinterpretation of RDTs by health care workers, and precautionary treatment referrals. A major contributor to these discrepancies is the tendency of health workers to dispense antimalarial treatment even to individuals who test negative.[12] This practice stems from various factors, including caution, lack of alternative fever treatments, patient demand, and a "no harm" approach given malaria's severity. Prescribing antimalarials to negative patients contributes to antimalarial resistance, a global concern. It also creates supply chain imbalances which incur significant cost implications.[13]

Ethiopia faces another challenge in its malaria elimination efforts due to widespread prevalence of *Plasmodium falciparum* parasites with deletions in the pfhrp2 and pfhrp3 genes. These deletions render the most common rapid diagnostic tests (RDTs), which target the HRP2 protein, ineffective, leading to "false-negative" results. Studies have shown alarming rates, with some regions reporting over 60% of P. falciparum samples lacking the pfhrp2 gene[22], and a recent countrywide survey indicated approximately 22% of the population harbors complete hrp2/3 deletions.[23, 24] This diagnostic evasion, coupled with emerging artemisinin resistance mutations, has contributed to a concerning resurgence of malaria, with over 7.3 million cases and 1,157 deaths reported in Ethiopia between January and October 2024 -the highest annual figures in seven years.[25, 26]

Ethiopia is actively working on introducing new RDTs that are not solely reliant on HRP2 detection. Notably, since March 2024, the Ethiopian Public Health Institute (EPHI) and its partners have been engaged in a strategic nationwide effort to fully replace old, less effective RDTs with the Biocredit Rapigen malaria RDT across all regions. This new RDT primarily detects *Plasmodium falciparum* lactate dehydrogenase (PfLDH), an alternative parasite antigen, which is not affected by *hrp2/3* gene deletions. Studies have shown that PfLDH-based RDTs like Biocredit RapiGEN offer improved sensitivity in areas with high *hrp2/3* gene deletion prevalence, outperforming conventional HRP2-only RDTs.[27] This systematic replacement is crucial for ensuring accurate diagnosis and effective treatment, thereby strengthening Ethiopia's malaria control and elimination strategies in the face of evolving parasite resistance.

While RDTs are essential, challenges with their use and subsequent adherence to treatment protocols include:

- **Misadministration:** Especially when multiple RDT types with differing instructions are in use.
- **Misinterpretation:** Particularly concerning different malaria strains.
- **Non-adherence to treatment protocols:** Leading to missed diagnoses and the risk of antimicrobial resistance.
- **Misreporting:** Inaccurate and untimely reporting.

HealthPulse AI, an innovation from the digital health non-profit Audere, provides digital building blocks to enhance the reliability and accessibility of RDTs for diseases such as HIV, COVID-19, and malaria. This tool enables health workers to accurately interpret RDT results using artificial intelligence (AI), thereby informing patient treatment and care plans. It specifically addresses challenges like inaccurate results and poor data reporting, particularly in underserved communities.

HealthPulse AI is delivered as a mobile software development kit (SDK) that can be embedded within existing Android or iOS applications, enabling real-time processing of RDT results without requiring internet connectivity, even on low-end android mobile phone devices. A key integration has been achieved with the eCHIS CommCare application, an open-source mobile platform by Dimagi. CommCare is widely utilized in low-resource settings to equip frontline health workers with tools for data collection and service delivery, even offline.

According to the WHO, digitizing malaria diagnostic testing can drive stronger adherence to treatment protocols and reduce the likelihood of health care providers prescribing antimalarials to negative test cases.[14] By studying this in Ethiopia, we can better understand the role digital health, particularly AI, can play in improving accuracy of RDT readings and HEW prescribing behaviors. Currently, there's limited research directly measuring the discrepancy between actual RDT results and recorded data at the community level. Furthermore, to our knowledge, no studies in Ethiopia have yet explored AI reading of RDT test results to assess its diagnostic validity compared to HEW interpretations. Exploring how digital diagnostic solutions can overcome these common challenges is crucial for advancing malaria control and elimination in Ethiopia.

Ethiopia's efforts to control and eventually eliminate malaria face significant hurdles, including the misadministration of malaria tests and treatment, misinterpretation of malaria RDT results, and non-adherence to established treatment protocols. By integrating an AI-powered mRDT reader into the existing eCHIS, the project sought to transform malaria diagnosis—ensuring accurate detection and classification of both the disease and its causative Plasmodium species—ultimately accelerating Ethiopia's path toward malaria elimination.

Objectives

The overall objective of this research was to evaluate the effectiveness of integrating AI with eCHIS in improving malaria RDT reading, optimizing prescription behavior, and enhancing treatment appropriateness.

The specific objectives of the research included:

1. To assess the impact of AI-powered decision support on the validity of mRDT readings by HEWs, specifically determining sensitivity, specificity, positive predictive value, and negative predictive value.
2. To analyze the influence of AI-powered decision support on anti-malaria drug prescription behavior.
3. To determine the effect of AI-powered decision support provided to HEWs on the appropriateness of malaria treatment for tested patients.
4. To explore the perceptions of HEWs regarding the utility and impact of AI-powered decision support for malaria diagnosis.

Materials and methods

To evaluate the effects of augmenting eCHIS with AI assistance for HEW in improving the validity of mRDT reading in Ethiopia, we conducted a sequential mixed methods, multi-design study; a parallel randomized cluster trial and a pre-post study design followed by a qualitative study. We used the unblinded, randomized cluster design to determine if the AI assistance changed the diagnostic validity of HEW readings of mRDT results, comparing an AI-assisted group who received feedback from AI with an AI-unassisted group. Due to the intervention changing HEW workflows for both study groups (see workflow description in Figure 1), it was not possible to blind HEW to their assignment. We randomized the study group assignment at the woreda level to reduce the chances of cross-over exposure to the control group. This also facilitated training and supervision for study procedures for each arm. To provide an independent verification of the diagnostic accuracy for both HEW and AI, an expert panel reviewed and interpreted all mRDT images taken by both groups used in this study.

We used the pre-post study to assess how HEW prescribing behavior changed after the AI was integrated into the eCHIS in both groups. Crucially, a qualitative study was conducted alongside the other methods to capture the nuanced experiences and perceptions of HEWs and their supervisors regarding AI support, providing essential context and deeper understanding.

Study setting and period

Geographic setting. We conducted the intervention for the study in five regions of Ethiopia—Oromia, Sidama, Central Ethiopia, Dire Dawa and Harari—along with the surrounding areas of the Dire Dawa city administration, aligning with the geographic areas covered by another JSI-led project, including the Digital Health Activity (DHA) project funded by the United States Agency for International Development. We collaborated with MOH experts to select eligible malaria-endemic woredas (see Table 3) within these regions. For a comparison group not included in the study and thus not exposed to AI, we used surrounding woredas within the same regions.

Time periods. After randomization of the woredas and HEW training in October 2024, we implemented the randomized cluster trial from November 1, 2024 to April 28, 2025. The pre-post study utilized data from the same data sources and geographic areas for a four month baseline period covering July 2024 to October 2024. The baseline was limited to these four months, because eCHIS all module implementation, including the malaria module, began July 2024. We conducted qualitative interviews in July 2025 with supervisors and HEW who had participated in the previous study components.

AI application integration

The study team collaborated with Audere to integrate the HealthPulse AI SDK into the eCHIS Android codebase in Ethiopia, specifically within the eCHIS malaria screening and focal testing and treatment modules. The intervention required having HEW take an image of the RDT using their eCHIS tablet during their standard malaria workflow to enable the HealthPulse AI to interpret the image in real time.

Results generated by HealthPulse AI – boolean variables for detection of control line, *Plasmodium falciparum line, Plasmodium vivax* line, and flags for image quality or interpretation concerns – were stored within existing eCHIS modules alongside the values entered by HEW, facilitating data utilization for future reference and analysis. To minimize storage space on the eCHIS server, images were stored directly on the HEW's device and later synchronized to AWS Server Space using the FolderSync application, a third party intermediary application used at the time of the study. The study team provided comprehensive support for app installation, user access, and general operation.

Rigorous testing was conducted to ensure the AI application integration did not disrupt the existing eCHIS application workflow. In September 2024, the application was deployed to the production server for field testing at four health posts near Arba Minch (Kola Mulatu, Kola Barena, Korga Geramo, and Fetele Dronge). The assessment focused on three key areas: the functionality and technological feasibility of the AI-eCHIS integration in field conditions, HEWs' comfort with using the AI tool, and testing proposed workflow. The team conducted 39 tests using both simulated cases and real clients. While the timing outside malaria season limited the number of malaria suspect cases, this mixed approach allowed comprehensive evaluation. Observations from the pretest were used to make necessary adjustments to both the eCHIS workflow and the subsequent training program for HEWs.

Study groups and workflows

Two distinct study groups were established for this evaluation – an AI-assisted group and an AI-unassisted group. The first steps of the workflow were the same in both groups. For a client who was determined to be eligible for an mRDT, upon reaching the testing step, the HEW entered their reading of the mRDT result and then were prompted to capture an image of the mRDT. If the AI could not interpret the first image, a second image attempt prompt appeared, followed by a third if necessary. If the AI still failed to process the image after the third attempt, the workflow proceeded to the next step without an AI interpretation. Where there was agreement between the initial result entered by the HEW and the AI interpretation, the HEW continued with the rest of their typical eCHIS malaria workflow with no explicit feedback or interaction with the AI.

The groups' workflow differed where there was any disagreement between the initial result entered by the HEW and the AI interpretation, positive versus negative and also positive on each species (*Plasmodium falciparum, Plasmodium vivax*, or mixed infection). In the AI-assisted workflow, the HEWs received a prompt with the AI-generated result and had the option to either agree with the AI and accept its suggested result or disagree and override the suggested result. In the AI-unassisted group, the AI processed results in the background, and HEWs were never informed of its interpretation.

Figure 1. eCHIS workflow in the AI-assisted and AI-unassisted groups

For both study groups, the eCHIS data automatically synchronized with the central eCHIS server when the device was connected to the internet. All HEWs had to manually synchronize their images to a study server using a FolderSync application. The study team reviewed the available, uploaded images and provided additional support to HEW who were struggling to synchronize. We generated a unique code for each image using random alphanumeric ID with no identifying information included.

The study team shared the set of final images read by the AI from both study groups with IndiVillage, a certified organization based in India with globally accredited experts to establish a ground truth reference result. Two panelists independently classified each line (control, *Plasmodium falciparum, Plasmodium vivax*) on the image of the diagnostic test as being there, not being there, or that they could not tell and noted any image quality and interpretation concerns test as positive, negative, or invalid. In cases of disagreement, a third expert facilitated a consensus decision. The panelists were blinded to the study group, as well as to the previously determined HEW and AI interpretations to avoid bias. We linked the panel readings with the HEW and AI readings of the same image using the unique ID.

Sample size

For objective one, validity of mRDT reading, we calculated the sample size of HEWs needed to detect a 4% difference in sensitivity and specificity between the two study groups. We based the calculation on previously published literature that health worker readings of RDT tests have 95% sensitivity and AI readings of RDT readers have 99% sensitivity.[18-20] With a significance level of 5%, 80% power, and a cluster design effect with an intra-class correlation level of 0.05, the required sample size of HEWs was 570 (285 per group).[40] This assumed we would use only one mRDT reading per HEW. Based on country averages, we assumed there would be 25 health posts in each woreda and two HEW based in each health post. As a result, we determined a sample of 12 woredas, six in each study group, would elicit a likely population of up to 600 eligible HEWs. After the study was initiated, the study team determined that we should include all available images for participating HEW in the analysis, leading to a much larger available sample.

For objectives two and three, prescription behavior and treatment appropriateness, we used all available data for participating HEW throughout the study period under objective one. For objective four, qualitative perceptions of AI use, we determined a convenience sample of six, targeting four HEW, one HEW supervisor (based at an affiliated health center), and one malaria focal person (based at the woreda health office) to get a range of perspectives.

Selection and recruitment

The selection of woredas was determined in collaboration with experts from MOH led by the National Malaria Control Program (NMCP). The selection criteria included woredas within the DHA project focus areas that were targeted for malaria elimination by the NMCP, having fully implemented all the eCHIS modules including the malaria module, with HEWs trained on the malaria module, with high specification tablets in use, with network connectivity available, and with relatively stable security conditions. From a sampling frame of the woredas that met these criteria, the selected woredas were randomly assigned to one of the study arms.

Within selected woredas, we invited all currently active HEW to participate in the study if they met eligibility criteria. To be eligible, HEWs needed to be at least 18 years old, currently trained in the malaria module, using the eCHIS malaria module, and willing to provide informed consent. Those who were not trained and did not give consent were excluded from the study.

For the qualitative component, the study team purposely selected three woredas that were assigned to the AI-assisted group. Trained HEW supervisors identified eligible HEWs in their woreda to participate in the key informant interviews. Inclusion criteria for participation in key informant interviews was to be at least 18 years old, exposed to the feedback from the AI decision support system as a HEW in the AI-assisted group or supervising such HEWs, and willing to provide informed consent.

Data collection

We administered a brief demographic survey to HEWs participating in the study during the training sessions. The survey collected data on age, gender, years of service as an HEW, education levels and duration of experience using the electronic malaria module.

Throughout the baseline (July-October 2024) and intervention period (November 2024 - April 2025), data on all malaria screening and treatment encounters was collected via the eCHIS CommCare platform as part of the routine care process by HEWs. In October 2024, the HEWs in both study groups received training on proper image capture and syncing techniques. AI-assisted HEWs were additionally trained on reading the AI-generated results and informed about their authority to accept or override the AI's decision. Study data were extracted from the eCHIS system for all relevant encounters for both groups on screening symptoms, HEW initial mRDT result, AI mRDT result, HEW final mRDT result, prescription behavior, and treatment appropriateness.

In addition to entering case data in eCHIS, HEWs complete a monthly report of their work that is aggregated and entered into the national DHIS2 data warehouse. To explore the completeness of study data compared to all data entered in eCHIS and DHIS2, we extracted aggregated comparable variables from eCHIS and DHIS2 reports for the study period across the 12 study woredas. To explore the possible effect on HEW of participating in the study and taking images of the mRDT, we extracted aggregated comparable variables from DHIS2 reports for the study period for 52 adjacent woredas.

We conducted key informant interviews via telephone with HEWs, their supervisors, and MOH staff using a guided semi-structured questionnaire. The interviews explored the perceived benefits, implementation challenges, and key lessons learned from integrating the AI solution into the malaria program.

Data quality

As mentioned previously, we conducted training for all participating HEW and their supervisors, with one training per woreda. All 12 woredas received training between October 28 and November 16, 2024. This training covered the study's objectives and design, research ethics, and provided a chance for HEWs to review and practice using data collection tools. Throughout the study period, HEWs received supervision and remote support. The data from the eCHIS central CommCare platform was downloaded in excel format and exported to STATA. The study team ran descriptive statistics to clean the data before proceeding with the main analysis.

For the qualitative telephone interviews, we chose locations with reliable network connections. Before beginning the interviews, we built a rapport with participants.

To maintain privacy and confidentiality, we assigned all encounters an anonymized unique identifier. This unique ID ensured that the study team could link mRDT interpretations from the HEW, the AI, and the panel of experts without mismatch. All responses, including electronic survey data, notes, transcripts, and audio files, were stored separately from any identifying information.

Analysis

We uploaded de-identified electronic CommCare and HEW demographic survey data into Stata 13. After initial coding, data underwent thorough cleaning for internal consistency, quality, and completeness. We performed cross-tabulations, constructed new variables as needed, and conducted preliminary analyses of key outcome variables to address research questions. Statistical significance was determined at $p<0.05$ across all analyses.

Diagnostic performance metrics. The calculation of four key diagnostic metrics – sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) – for HEW and AI mRDT readings adhered to a multi-step process (Table 1), with a specific approach to account for potential autocorrelation in the data. The core of the calculation involved comparing HEW and AI mRDT readings against a ground truth interpretation determined by the expert panel. Mixed effects modeling was employed to assess the agreement between HEW/AI and expert readings, allowing for the adjustment of autocorrelation due to the clustered nature of the data. These diagnostic performance parameters were calculated for AI-assisted and AI-unassisted groups, and for HEW and AI separately. We compared the parameters across the different groups for statistical significance using the Z-test and McNemar test.

In addition, to explore if AI feedback could support learning and performance improvement among HEW, we used descriptive statistics to analyze discordant results; patterns of errors and the corresponding HEW reactions to them; and characteristics of persistent low performers.

Table 1. Diagnostic performance metrics calculations

<table>
<tr><th rowspan="2">Metric</th><th rowspan="2">Definition</th><th colspan="4">Calculation process[41]</th></tr>
<tr><th>Step 1. Selection of relevant cases</th><th>Step 2. Data categorization</th><th>Step 3. Clustering assessment and odds of success</th><th>Step 4. Autocorrelation adjustment and metric calculation</th></tr>
<tr><td>Sensitivity</td><td>The ability of the HEW mRDT to correctly identify individuals who truly have the disease (as confirmed by expert panel).</td><td>Selected True Positive Cases. Only tests where the Panel of Experts confirmed a positive result were selected for analysis. This ensures that the calculation focuses on individuals who truly have the disease.</td><td rowspan="4">Compare the HEW or AI mRDT reading with the expert panel result, creating a binary variable:
● success indicates agreement
● failure indicates disagreement</td><td rowspan="4">An intercept-only mixed effects model was fitted to the binary data to assess any inherent clustering within the data and estimated a fixed intercept. Exponentiating this fixed intercept yielded the estimated odds of success.</td><td rowspan="4">To adjust for potential autocorrelation, the probability of success was derived from the estimated odds (converting log-odds to probability). This probability, expressed as a percentage, represented the final calculated metric after accounting for dependencies.</td></tr>
<tr><td>Specificity</td><td>The ability of the HEW mRDT to correctly identify individuals who truly do not have the disease (as confirmed by expert panel).</td><td>Selected True Negative Cases. Only tests where the Panel of Experts confirmed a negative result are selected for analysis. This focuses the calculation on individuals who truly do not have the disease.</td></tr>
<tr><td>PPV</td><td>The probability that an individual truly has the disease, given that the HEW mRDT reading was positive.</td><td>Selected All Positive HEW Readings. All cases where the HEW's mRDT reading was positive are selected for analysis, regardless of the expert's result at this stage.</td></tr>
<tr><td>NPV</td><td>The probability that an individual truly does not have the disease, given that the HEW mRDT reading was negative.</td><td>Selected All Negative HEW Readings. All cases where the HEW's mRDT reading was negative are selected for analysis, regardless of the expert's result.</td></tr>
</table>

Prescription behavior. This metric quantified the overall adherence of HEW to appropriate malaria prescription protocols based on the RDT results. This calculation of this metric used the formula: (positive cases treated correctly for malaria + negative cases correctly not treated for malaria) / total cases tested for malaria. We developed two interrupted time series analysis (ITSA) models to analyze the impact of the intervention on prescription behavior.[42] First, for a single-group ITSA, we aggregated the prescription behavior data from all HEWs in both groups (AI-assisted and AI-unassisted) across all observations and HEWs for each month, and we calculated the average correct prescription behavior for each month from this unified dataset for the entire study period – baseline and intervention. For the single group ITSA, this evaluated the overall impact of the intervention on the combined average trend and level of correct prescription behavior across all HEW. Secondly, for a two-group ITSA, we calculated the average correct prescription behavior distinctly for each group (AI-assisted and AI-unassisted) for each month, allowing for separate group analysis of trends and levels.

Treatment appropriateness. This metric measured the proportion of patients that received appropriate treatment. In contrast to prescription behavior above, this indicator considered treatment of false negative cases as appropriate treatment. The calculation of this metric used the formula: (true positive cases treated + false negative cases treated) / all treated cases.

Qualitative interviews. For the qualitative component, all key informant interviews were audio-recorded and transcribed verbatim. We conducted a thematic exploratory analysis, reviewing coded outputs and synthesizing them into key themes, which were then refined through team discussions.

Comparison groups. We used descriptive statistics of comparison group data to explore how well the study data represented the overall number of malaria tests conducted in the study woredas, and if there was an effect of study participation on HEW behavior.

Ethical considerations

We obtained ethical approval from JSI IRB (#24-14) and the IRB of the EPHI (EPHI-IRB-565-2024). Subsequently, permission was secured from regional health bureaus and woreda health offices before data collection commenced. Informed consent was obtained from HEWs at the time of enrollment into the study. Participants were fully informed about the study's purpose, procedures, potential risks, and benefits, as well as their right to voluntary participation without penalty. Interviewers clearly explained that participation was not compensated, would not be shared with supervisors, and that individuals could withdraw at any time, even after initial agreement or during the course of the data collection period. Special emphasis was placed on protecting confidentiality and ensuring privacy during all interactions. Written consent was obtained at the time of enrollment into the study. Additional verbal consent was obtained for the key informant interviews before audio recording.

Results

Study participation

There were 100 woredas within the DHA project focus areas that the NMCP targeted for malaria elimination. Of those, 14 woredas had fully implemented all modules of eCHIS, including the malaria module, and were using high specification tablets. Two of these woredas were excluded as they were in an insecure setting. All 12 remaining eligible woredas were selected to take part in the study and were randomly assigned to a study arm. The final study sites included Sofi and Erere from the Harari Region; Melka Jebdu and Biyo Awale operational woredas from Dire Dawa city administration; Shashemene Zuria, and Arsi-Negele from Oromia Region; Dale, Dara, Dara Otilcho and Wondo Genet from Sidama Region, and Weera and Weera Dijo from the

Central Ethiopia Region (Table 2). On average, study regions were located approximately 380 km from Addis Ababa.

The study team found 384 HEW present across the 12 selected woredas and invited them to enroll in the study in October 2024 and attend a study training. All eligible HEW completed the comprehensive training and enrolled. During the six-month intervention period, 341 (89%) trained HEWs actively reported at least one malaria case using the eCHIS application, demonstrating exposure either with the AI or image capture and thus participated in study data collection. We removed HEW who did not interact with the AI or capture images from further analysis. We asked HEW supervisors for reasons why some HEW may not have participated, and they shared that it could be due to being out on maternity leave, pursuing educational opportunities, and transferring to another geographic area.

Table 2. Woreda and HEW participation

Region	Zone	Woreda	Study arm	Health centers	Health posts	HEW enrolled	HEW participated
Dire Dawa	Dire Dawa	Biyo Awale	AI-unassist	2	10	20	13
		Melka Jebdu	AI-assist	1	9	14	10
Harari	Harari	Sofi	AI-unassist	2	11	21	18
		Erer	AI-assist	1	7	14	12
Sidama	Sidama	Dale	AI-assist	9	32	61	58
		Dara	AI-unassist	3	21	40	37
		Dara Otilcho	AI-assist	3	15	30	30
		Wondo Genet	AI-unassist	3	14	27	27
Central Ethiopia	Halaba	Weera	AI-assist	5	25	36	27
		Weera Dijo	AI-unassist	5	26	38	33
Oromia	West Arsi	Arsi Negele Rural	AI-unassist	8	31	52	47
		Shashemene Zuria	AI-assist	4	18	31	29
Totals			**AI-unassist**	**23**	**113**	**198**	**175**
			AI-assist	**23**	**106**	**186**	**166**
			Total	**46**	**219**	**384**	**341**

Study group characteristics

We assessed the baseline comparability between the AI-assisted and AI-unassisted groups of HEWs (Table 3). The comparison revealed that the two groups were well-matched across most measured characteristics. However, the AI-unassisted group had a larger number of university-educated HEW. And, HEWs in the AI-assisted group demonstrated significantly greater mean experience in malaria prevention, care, and treatment compared to their counterparts in the AI-unassisted group. Importantly, this difference in overall malaria experience did not extend to other experience measures. Both groups showed comparable durations of service at their current health posts and similar levels of experience using the malaria eCHIS module specifically. Lastly, educational level status was not part of the original inclusion criteria, however, in the future, given the imbalance, it would be.

Table 3. Baseline characteristics of HEW in the AI-assisted and unassisted groups

Variables		AI-assisted	AI-unassisted	Absolute difference	p-value
Total		166	175		
Education Level of HEW n (%)	Level 3	46 (28)	42 (24)	3.7	0.689
	Level 4	116 (70)	102 (58)	12	0.075
	University or higher	4 (2)	31 (18)	-15	0.430
Age of HEWs (mean years)		30.1	28.4	1.7	0.080
Experience as a HEW (mean years)		11.5	9.6	1.9	0.003*
Experience of HEWs in malaria prevention, care and treatment (mean years)		10.0	8.5	1.5	0.031*
Experience of HEWs at this health post (mean years)		8.8	7.7	1.1	0.097
Experience of HEWs in malaria prevention, care and treatment at this health post (mean years)		8.0	6.8	1.2	0.083
Experience of HEWs in using the malaria module in eCHIS (mean years)		1.6	1.7	-0.1	0.506

*Statistically significant, p<0.05

Malaria testing result

Across twelve malaria-endemic woredas, HEWs performed 2,867 RDTs with near-equitable gender distribution: 1,523 tests (53%) for male clients and 1,344 (47%) for female clients. The AI-assisted group conducted 1,576 tests (55% of total), while the AI-unassisted group completed 1,291 tests (45%). Notably, the AI-unassisted HEW reported a higher proportion of positive malaria test results (69%, 887/1,291) compared to the AI-assisted HEW (56%, 885/1,576) (Figure 2).

Image interpretation analysis showed that in the AI-assisted group, the algorithm successfully interpreted 1,494 images (95% of the group total) compared to expert panelists' 1,473 interpretations (93%). The AI-unassisted group demonstrated marginally higher interpretation rates, with both the AI algorithm and panelists analyzing 98% of submitted images (1,269 and 1,261 respectively). This meant that the proportion of uninterpretable results was consistently higher in the AI-assisted group (Figure 2).

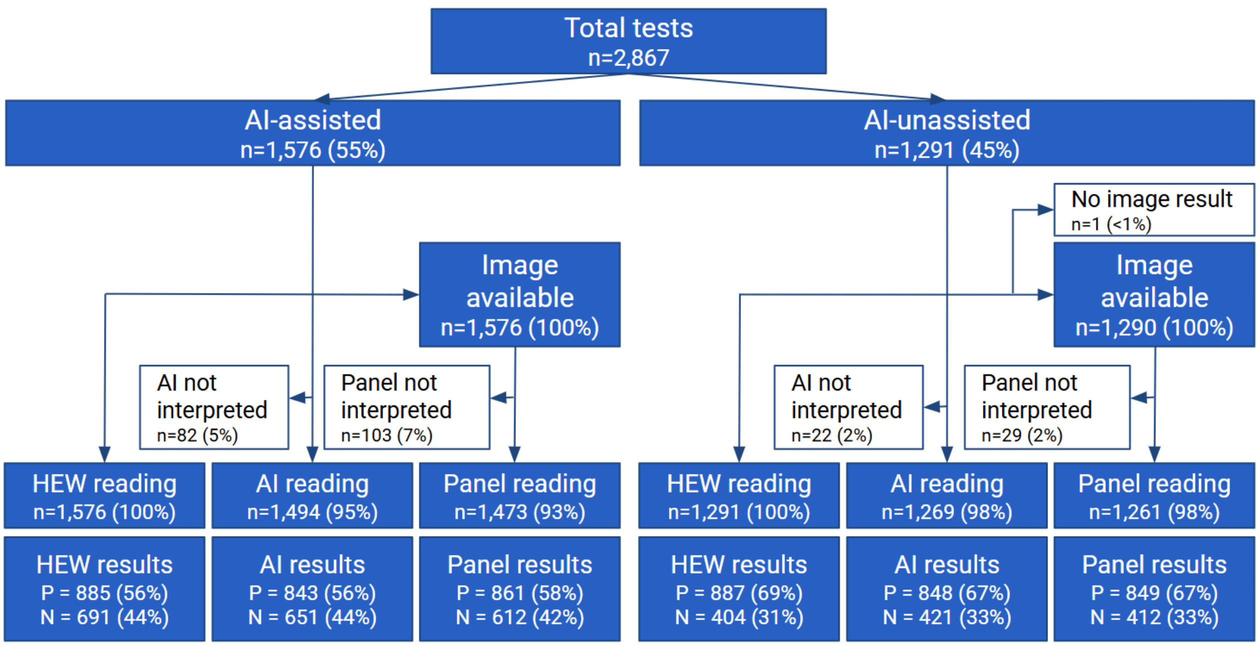


Figure 2. Malaria testing flow diagram

Diagnostic accuracy of HEWs

Our initial model revealed a moderate intra-cluster correlation coefficient of 9%, highlighting moderate variability in mRDT reading accuracy among the HEWs in our study. To evaluate the accuracy of mRDT readings by HEWs and AI overall, we used the expert panel results as the ground truth and calculated the sensitivity, specificity, PPV, and NPV for both study arms combined for the initial HEW results before any AI feedback and the AI results (Table 4). Overall, both HEW and AI performed highly (95% or higher) across all performance measures. HEW demonstrated stronger sensitivity (99.8%) and NPV (99.2%) compared to AI (96.8% and 95.3% respectively), but the AI demonstrated stronger PPV (99.7%) compared to HEW (98.5%).

To evaluate the accuracy of mRDT readings by AI-assisted HEW compared to AI-unassisted HEW, we calculated the same performance measures for each study arm separately (Table 4). In this analysis, both groups also performed highly (97% or higher) across all performance measures. The AI-assisted group demonstrated stronger specificity (99.3%) and PPV (99.5%) compared to the AI-unassisted group (97.9% and 98.3% respectively). Whereas the AI-unassisted group demonstrated stronger sensitivity (100%) compared to the AI-assisted group (97.3%).

Table 4. Diagnostic validity of HEW and AI mRDT readings

	Study arms combined			Study arms separated		
mRDT reading performance	**HEW initial**	**AI**	**p-value**	**AI-assisted HEW**	**AI-unassisted HEW**	**p-value**
Sensitivity	99.8%	96.8%	0.0000*	97.3%	100%	0.0000*
Specificity	98.0%	98.5%	0.1238	99.3%	97.9%	0.0375*
PPV	98.5%	99.7%	0.0001*	99.5%	98.3%	0.0173*
NPV	99.2%	95.3%	0.0000*	97.4%	97.9%	0.5552

*Statistically significant, p<0.05

Discordance analysis

We compared discordant results between the HEW initial reading (before any AI feedback) and the AI reading in both study groups for both (1) the diagnosis of malaria (e.g., does the client have malaria or not); and (2) for the species determination (e.g., does the malaria positive client have *Plasmodium falciparum*, *Plasmodium vivax*, or both). In the diagnosis of malaria, across both study groups, there were 134 (5%) discordant diagnoses, demonstrating that initial agreement for mRDT readings as being positive or negative for malaria between HEWs and AI was 95%. And among the 1,653 tests that were determined to be positive for malaria that had species results from both HEW and AI, there were 108 (7%) discordant species readings, demonstrating that initial agreement on species determination was 93%.

Diagnostic discordance and accuracy. Among the 134 instances of discordant diagnosis, the expert panel was able to determine a result for 114 (20 images were uninterpretable byr the panel), and they confirmed the accuracy of the HEW initial mRDT readings in 55% (63/114) of valid tests, whereas the expert panel confirmed the AI reading was correct in the remaining 45%. Of all 114 valid discordant diagnosis results, 70 were in the AI-assisted group where the AI result was observed and the HEW was given the option to agree with the AI result or override it. The HEW generally disagreed with and overrode the AI (61%, 43/70) and this decision was confirmed to be correct by the expert panel in 63% of the tests (27/43). For those HEW who agreed with and used the AI result rather than their own initial result (39%, 27/70), their decision was also confirmed to be correct by the expert panel in 63% of these tests (17/27).

Further investigating the interaction between HEW and AI observations, we analyzed sensitivity and specificity based on result agreement. When HEW and AI observations were concordant, diagnostic performance was outstanding, achieving 99.7% sensitivity and 100% specificity. This strong agreement, however, was in stark contrast to situations with discordant results, where sensitivity remained at 100% but specificity plummeted to nearly zero (Table 5).

Table 5. Analysis of diagnostic metrics among HEW and AI discordant and concordant results

HEW versus AI results	Metric	
	Sensitivity	**Specificity**
Concordant	99.7%	100%
Discordant	100%	0.0%

This pattern strongly suggests that errors are prevalent in both AI and HEW diagnoses when their results conflict. Consequently, diagnostic decisions made under discordance, whether through overriding or agreement, led to a substantial number of patients being poorly managed. In qualitative interviews, HEWs confirmed this challenge, indicating their practice of referring patients with discordant results to higher health facilities for advanced diagnoses like blood microscopy. One HEW reported that "I wouldn't override the result of the AI when the result is discordant with my reading, but I have to refer the patient to the health facilities." Similarly, another HEW who did not encounter discordant results stated their intent to re-check the RDT test kit or verify the image quality.

Species determination discordance and accuracy. The total number of discordant species determinations for both groups was 108. After removing nine panel-uninterpretable tests, 99 discordant tests remained for accuracy analysis. HEWs' initial species determination accuracy against the panel was only 21% (21 tests). In stark contrast, the AI's accuracy against the panel was 79% (78 tests), indicating that the AI was significantly better at determining species.

Within the AI-assisted group, there were 59 discordant species determinations. HEWs agreed with the AI's species determination in 45 (76%) of these cases, but they overrode the AI in 14 (24%) of them. When HEWs agreed with the AI on species (45 tests), their final decision was accurate in 89% (40 tests). However, when HEWs overrode the AI's interpretation for species (14 tests), their final decision was accurate in only 57% (eight tests).

Detailed species determination analysis (AI-assisted group only). Among 9 panel-confirmed mixed infections, HEWs agreed with the AI in eight cases. When they agreed, their determination was accurate in 50% (4 tests) of these mixed infection cases.

Among 25 panel-confirmed *Plasmodium falciparum* cases, HEWs agreed with the AI in 22 (88%) of these cases. When they agreed, their final decision was 100% accurate (22 tests). It's noteworthy that in 19 (86%) of these cases, HEWs initially diagnosed *Plasmodium vivax* instead of *Plasmodium falciparum*, with the remaining three being mixed infections. When HEWs overrode the AI in three *Plasmodium falciparum* cases, their final decision was also 100% accurate.

Regarding *Plasmodium vivax* infections, there were 25 panel-confirmed cases. HEWs agreed with the AI in 15 (60%) out of the 25 discordant tests, and their final decision was 93.3% accurate (14 tests). When they overrode the AI, HEWs were accurate in 50% (five tests) of the cases.

Cohort analysis

As shown above in the discordance analysis, when the AI contradicted the initial HEW diagnosis leading to a discordant diagnostic result, we found that the AI was more frequently in error. The expert panel results confirmed most HEW decisions rarely change their result based on AI feedback in these discordant cases. However, even in the 16 cases when the HEW was in error in the AI-assisted group, HEW changed their result in only 3 (19%) of the cases, seemingly indicating resistance to AI feedback (Table 6). In contrast, when the AI contradicted the HEW species determination, the AI was more often correct (shown above). HEW reacted to the AI feedback about species determination by changing their result in 39% of cases (42/109).

Table 6. HEW reaction to AI feedback on diagnosis errors

Diagnosis error	Result determinations			AI-assisted HEW reaction	
	HEW initial result	AI result	Expert panel result	Did not change result	Changed result
False positive	Positive	Negative	Negative	3	1
	Positive	Inconclusive	Negative	1	0
False negative	Negative	Positive	Positive	7	2
	Negative	Inconclusive	Positive	2	0
Totals				13	3

Across both study groups, only 24 (7%) HEW made two or more errors over the entire study period, including both diagnosis and species errors and excluding missing species values and inconclusive panel results. While this is a small group of individuals, they contributed 67% of all confirmed errors (73 out of 109).

Prescription behavior and treatment appropriateness

Prescription behavior. The single-group ITSA model demonstrated an excellent fit to the combined data, explaining a substantial proportion of the variance in average correct prescription behavior (R-squared = 0.9519, F(3, 6) = 49.55, p < 0.0001). The analysis revealed that the month coefficient was -3.131389 (p=0.008), indicating a statistically significant decrease of approximately 3.13 percentage points per month in average correct prescription behavior during the pre-intervention period. Immediately following the intervention, a statistically significant immediate increase in average correct prescription behavior was observed, with the intervention coefficient of 10.61876 (p=0.006), representing an immediate rise of approximately 10.62 percentage points. Furthermore, the post_month coefficient of 4.474563 (p=0.002) denoted a statistically significant positive change in the post-intervention trend, reflecting an increase of approximately 4.47 percentage points per month compared to the pre-intervention trend (Figure 3, Annex 1).

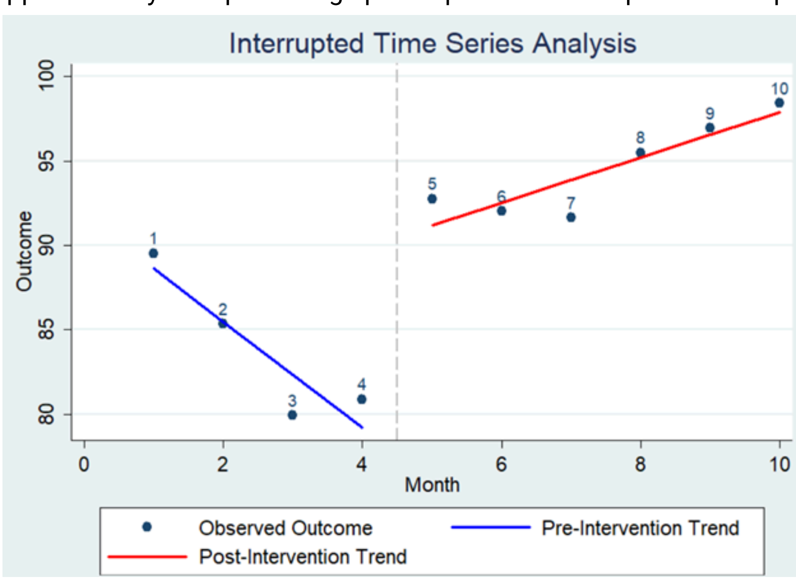


Figure 3. Levels and trends before and during intervention period

The two-group ITSA model demonstrated a statistically significant fit, explaining a substantial portion of the variance in correct prescription behavior ($R2=0.7833$, Adjusted $R2=0.6569$, $F(7,12)=6.20$, $p=0.0031$). For the AI-unassisted (reference) group, the baseline correct prescription behavior was estimated at 83.5% (constant coefficient = 83.50562). The pre-intervention trend, indicated by the _time coefficient, showed a non-significant decline of 2.29 percentage points per month (coefficient = -2.285829, $p=0.327$). At the point of intervention, a statistically significant immediate increase of 19.00 percentage points in average correct prescription behavior was observed for this group (coefficient I = 19.00262, $p=0.010$). Following the intervention, the _postintime coefficient suggested a non-significant post-intervention trend increase of 2.81 percentage points per month (coefficient = 2.808791, $p=0.290$).

Importantly, the interaction terms, which capture the differential effects for the AI-assisted group compared to the AI-unassisted group, were not statistically significant. Specifically, the _group coefficient indicated that the AI-assisted group had a non-significant 12.93 percentage point higher baseline value compared to the AI-unassisted group (coefficient = 12.93586, $p=0.161$). The Z_time coefficient revealed a non-significant differential pre-intervention trend, with the AI-assisted group declining by an additional 1.89 percentage points per month compared to the AI-unassisted group (coefficient = -1.892113, $p=0.561$). Furthermore, the Z_I coefficient demonstrated a non-significant differential immediate impact, showing the AI-assisted group experienced a 12.16 percentage point lower immediate change compared to the AI-unassisted group at the intervention point (coefficient = -12.15976, $p=0.194$). Finally, the Z_postintime coefficient indicated a non-significant differential post-intervention trend, with the AI-assisted group showing an additional increase of 3.39 percentage points per month compared to the AI-unassisted group (coefficient = 3.386102, $p=0.363$). We can also see that there was a delayed increased effect of the AI-assisted group compared with the AI unassisted group around month 9 (Figure 4, Annex 1).

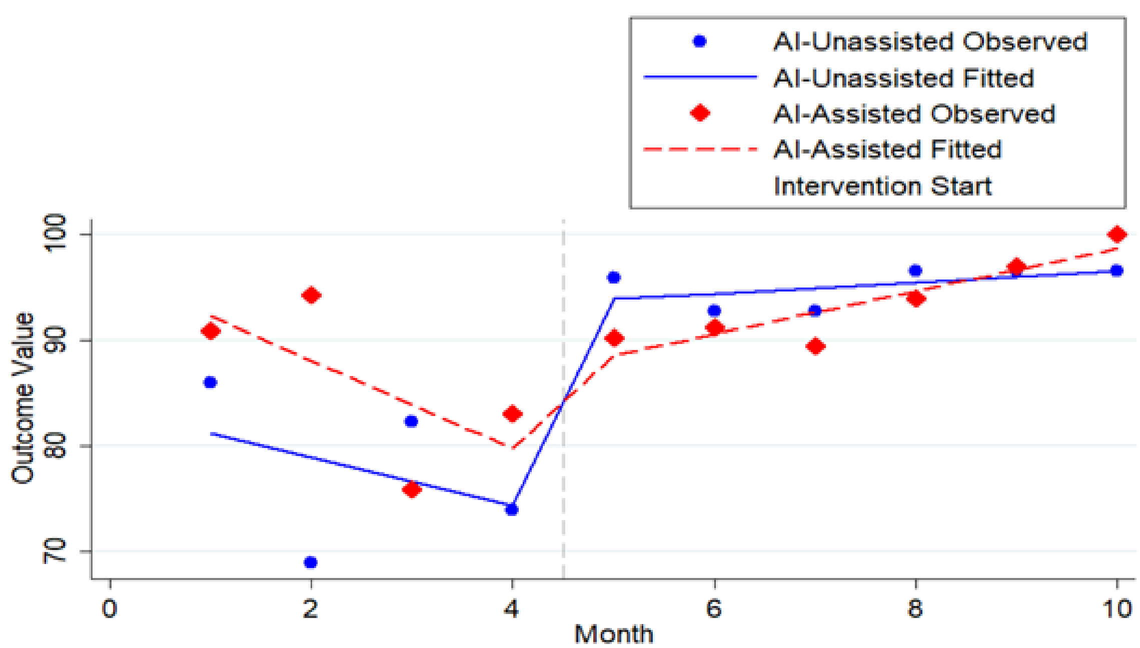


Figure 4. Levels and trends before and during intervention period for AI-assisted and AI-unassisted groups

Treatment appropriateness. The HEWs are consistently expected to prescribe medications for cases they diagnosed as positive and withheld prescriptions for those diagnosed as negative. This is correct prescription behavior as prescription decisions are made based on the result of the RDT interpretation by the HEW. This practice reflects a direct adherence to HEWs diagnostic findings. However, it is expected that a portion of the cases diagnosed as negative by HEWs could be in fact false negatives when confirmed by the result of the panel reading. For these false negative instances, the decision to withhold treatment results in the patient not receiving care for their actual condition. We conducted a comparative analysis between the AI-assisted and AI-unassisted groups regarding treatment appropriateness. The treatment appropriateness was similar in the AI-assisted group (89%) compared to the AI-unassisted group (93%).

mRDT brands used

Beyond the main outcomes of the study, AI also provided an opportunity to analyze the types of RDT brands utilized across health posts in the 12 sampled woredas. This analysis aimed to inform the MOH about the distribution reach of newly introduced RDTs throughout the national health system. Of the 2,867 RDTs conducted over a six-month period, 2,716 (95%) were Biocredit Rapigen. Other brands observed included Abbot, First Response Malaria, Meril, and MALARIA. Notably, most of the outdated RDTs were identified in the Harari and Central Ethiopia Regions (Figure 5, Table 8).

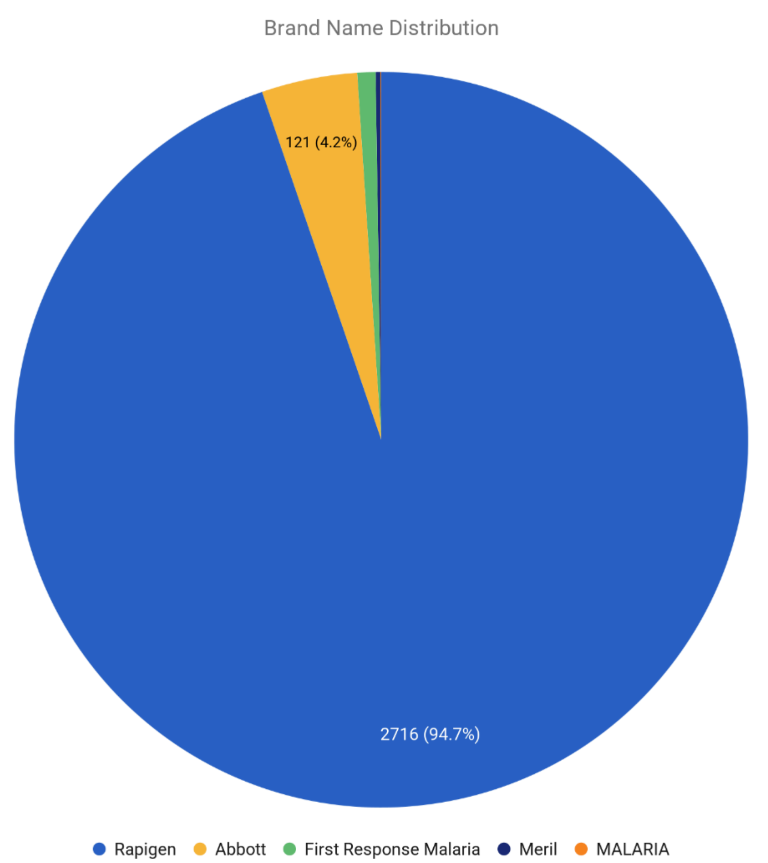


Figure 5. Distribution of RDT brands used for malaria testing by HEWs in the 12 woredas, Nov. 1, 2024 to April 28, 2025

Table 7. Regional distribution of RDT brands used for malaria testing by HEWs in the 12 woredas, Nov. 1, 2024 to April 28, 2025

Regions	Brand Names				
	Rapigen	Abbott	First Response Malaria	Meril	MALARIA
Sidama	617	56	9	0	1
Harari	67	29	2	0	0
Dire Dawa	32	1	0	0	0
Oromia	991	3	0	6	0
Central Ethiopia	1009	32	12	0	0
Total	**2716**	**121**	**23**	**6**	**1**

HEW experiences with AI

Time for AI capture and interpretation results. The AI-assisted HEWs reported that getting the AI result was instantaneous. The AI interpretation of the images was automatic and did not add time to their routine work process. The HEWs reported that, "It (the image capture and the AI result interpretation) was not time taking. It was fast and easy system..." This view was reinforced by another HEW, who added, "If the result is discordant, I may override the result of the AI and consider my own reading and treat the cases based on the result of my reading."

Benefits of using the AI application. The HEW and their supervisors consistently highlighted numerous benefits of the AI application, drawing from their practical experiences. For example, an HEW from Hallaba Zone in Central Ethiopia remarked that the AI significantly boosted her confidence in reading malaria mRDTs – "I managed malaria in the past, but the AI application helped me feel more confident reading the mRDT because my readings match the AI's." Similarly, an HEW in the Sidama Region underscored a range of personal and professional gains from using the AI application – "AI helped me develop confidence, earn respect in front of the community, and manage my time better."

HEWs also reported that the images captured by the AI application were stored and transferred through the eCHIS. This feature aids in triangulating interpretations of results and treatments provided, which ultimately improves data quality and the quality of services delivered. An HEW from Dale woreda emphasized that "It improved the quality of malaria data, in addition to enhancing service quality." A malaria focal person in Weera Woreda enthusiastically described the AI application as "wonderful." He detailed its advantages, "It benefits the HEWs; they get support from the AI on their decisions. It helps them pay attention to their work, and it also benefits the community." He further highlighted broader systemic improvements, particularly in increasing the utilization of the eCHIS system, as HEWs sought confirmation from the AI application during their malaria screenings.

Image capture and syncing. The HEWs generally reported that mRDT image capture and syncing were not problematic. However, qualitative interviews revealed frequent issues with connectivity and tablet capacity. For example, one HEW in Weera woreda stated, "Capturing mRDT images was so easy for me but syncing the picture was a challenge due to the quality of my tablet and the connection." Another HEW from Dale woreda similarly explained, "It is easy to capture mRDT images and sync them but I had a poor capacity tablet, which forced me to use my friend's tablet to capture and sync the mRDT image." HEWs consistently reported difficulty retrieving synced images due to complex file pathways, with one HEW in Dale woreda noting, "I think there is a problem in tracing the synced mRDT pictures and this needs to be improved." It is important to note that the third party application, FileSync, was an intermediary solution at the time of the study, due to Ethiopia local server capacity issues unable to store and sync mRDT images that would no longer be used in implementation at scale.

Despite HEWs' perception of easy capture, our analysis of the images indicated a need for significant improvement in image quality (Table 8). While the number of uninterpretable images was low, some of the challenges included poor image capture, which made accurate interpretation by both the AI and the expert panel problematic.

Table 8. Image quality issues with examples

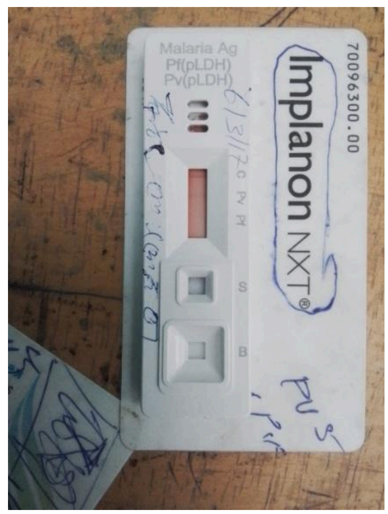
Image quality issue	Image example
Inconsistent Backgrounds: Images of mRDTs are taken against various backgrounds, including paper, carpets, cloths, tablets, and benches, leading to inconsistency and potential interference with interpretation of the images.	
Blood Stains: A significant number of mRDT images are blood-stained, obscuring the control, *Plasmodium vivax*, and *Plasmodium falciparum* lines. This makes accurate reading difficult for both AI and human experts and is believed to affect the quality of care being provided to patients.	 

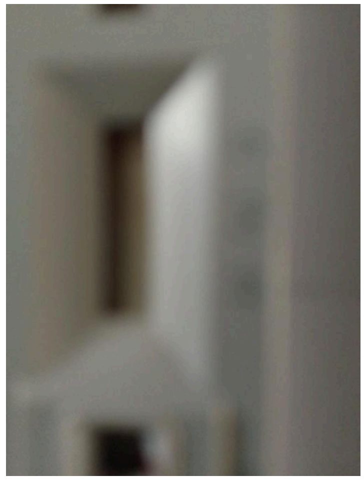
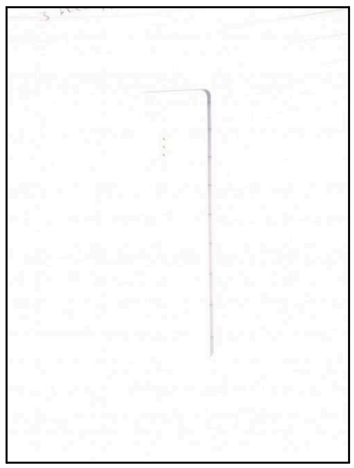
Image quality issue	Image example
Incorrect Distance: Many mRDT images are captured at improper distances—either too far or too close—making it challenging for AI and expert panels to interpret them correctly.	 
Lighting Issues: Problems with lighting are observed in a notable number of images. Some images have excessive light, making them difficult to read, while others suffer from shadows cast by the mRDT or surrounding objects.	 
Cluttered Backgrounds and Multiple Images: Some mRDT images feature cluttered backgrounds, and in a few instances, multiple mRDTs are captured within a single frame, complicating individual analysis.	 

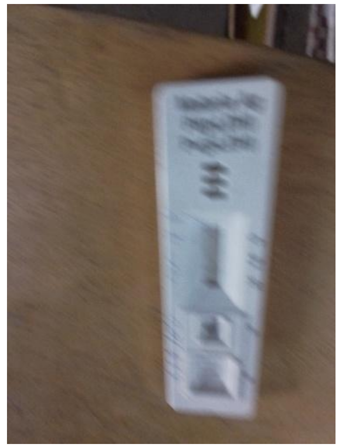
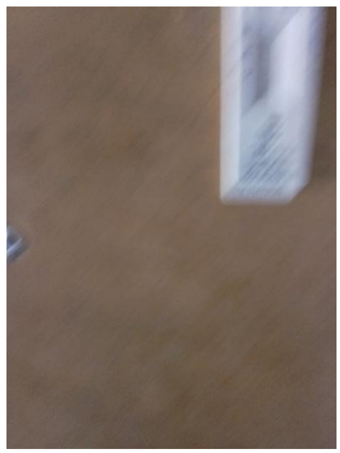
Image quality issue	Image example
Poor Image Resolution: The resolution of captured images was often low, resulting in blurry or fuzzy visuals that are difficult for AI and expert panels to interpret.	 
Mismatched Framing: In some images, the mRDT is out of frame, preventing complete and accurate assessment.	 
Incorrect Angle: Images are frequently taken from a low angle, which impedes clear reading of the mRDT.	

Image quality issue	Image example
Contaminated mRDTs: Some mRDT images show devices that are dirty or soiled, indicating poor storage practices.	

Comparison groups

The study team obtained aggregate data from eCHIS and DHIS2 for the same 12 study woredas as a comparison to understand what the coverage of this intervention was during the study time period. We used the eCHIS comparison to understand the volume of malaria tests and positivity rate being captured in eCHIS that could have occurred outside of the study (due to problems taking or synchronizing images). We used the DHIS2 comparison to understand the volume of malaria tests and the positivity rate being reported in the routine aggregate reporting system, DHIS2, which includes tests conducted using eCHIS but also tests conducted and captured on paper-based reporting systems.

During the six-month intervention period, November 2024-April 2025, the study captured 2,867 tests conducted using eCHIS that included an image, of which 1,772 (62%) were determined to be positive by HEW initially. When comparing with the other data sources (Table 9), we noted that 50% (2,867/5,758) of all tests captured in eCHIS were included in the study, and the positivity rate was the same in the study (62%) as in all eCHIS data (63%). We also noted that only 15% (5,758/38,269) of all malaria tests were captured in eCHIS out of those reported to be done in DHIS2. The positivity rate reported in DHIS2 was lower, at 36%.

Table 9. Comparison of data for 12 study woredas across different sources, November 2024-April 2025

	Total tests	Number positive	% positive
Study eCHIS data- with image included in study	2,867	1,772	62%
All eCHIS- aggregated	5,758	3,635	63%
DHIS2	38,269	13,621	36%

Another comparison analysis helped us to determine if the act of capturing images of mRDT tests in this group fostered a sense of being remotely observed by supervisors, thereby altering HEWs' mRDT reading and prescription patterns. We hypothesized that if HEW behavior was altered by participating in the study, the positivity rate in our study area would diverge from that of surrounding woredas with comparable characteristics. To examine this, we used data from DHIS2, as eCHIS was not universally implemented in all woredas, and compared the trends. The positivity rate trends were similar in both groups (study woredas and adjacent woredas)– around 40% during the baseline period and then descending to around 25% by the end of the intervention period – indicating little evidence for the study causing behavior change (Figures 6 and 7).

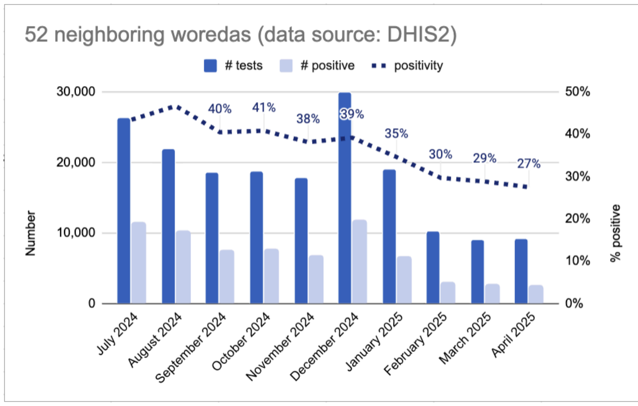


Figure 6. Malaria positivity rate in adjacent woredas

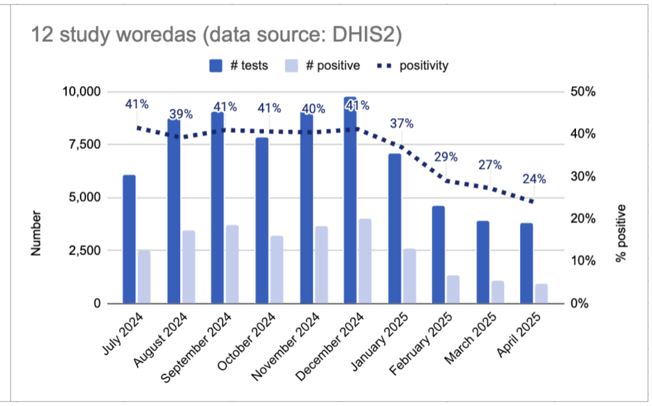


Figure 7. Malaria positivity rate in study woredas

Discussion

This study aimed to investigate the impact of AI integrated into eCHIS on the performance of HEWs mRDTs reading, correct malaria treatment prescription behavior, and treatment appropriateness among HEWs. Our analysis included 2,867 tests conducted by 341 HEWs, controlling for the effect of clustering.

A significant proportion of these tests (55%) were performed in the AI-assisted group. The proportion of tests interpreted by the AI and the expert panel was lower in the AI-assisted group compared to the AI-unassisted group. This observation may be attributed to potential differences in the quality of images captured by HEWs across the two groups, which could have rendered interpretation challenging for both the AI and the expert panel. Further, our data indicated a higher number of HEWs with university or higher education levels in the AI-unassisted group. This suggests that better-educated HEWs in the AI-unassisted group might have been more proficient in capturing images suitable for accurate interpretation by both the AI and the expert panel.

Regarding test positivity rates, there was no substantial difference between HEWs, the AI, and the expert panel within each group. However, a cross-group comparison revealed generally higher positivity rates in the AI-unassisted group than in the AI-assisted group. These discrepancies could stem from differences in treatment-seeking behaviors among individuals interacting with HEWs in the respective groups. Overall, the expert panel identified more positive cases than both HEWs and the AI, underscoring the critical need for more robust diagnostic tools to minimize false-negative cases, which can lead to misdiagnosis and untreated infections that could make malaria elimination efforts difficult.

Geographical variations in malaria positive case distribution based on the eCHIS data were observed across the woredas of the two study groups -a finding consistent with data from the DHIS2. Discrepancies in the number of malaria test positive cases were also noted between the two systems. These inconsistencies primarily arise from inconsistent utilization of eCHIS among health posts. Previous studies on eCHIS implementation in Ethiopia suggest a near universal level of acceptability but suboptimal utilization mainly caused by barriers that include lack of connectivity and power, insufficient training, and limitations in tablet capacity, all of which align with our observations.[28-29]

Both groups demonstrated high scores in sensitivity, specificity, and predictive values. Despite the generally high performance, HEWs in the AI-unassisted group achieved a higher and statistically significant sensitivity compared to the AI-assisted group. Conversely, the AI-assisted group demonstrated statistically significantly higher specificity and positive predictive value. In studies conducted in Nigeria and Kenya, AI produced higher sensitivity and accuracy compared to community health workers.[33-34] Despite mixed results in our study, the difference could be due to the extensive experience of HEWs in malaria prevention, care, and treatment, including their proficiency in using RDTs for diagnosis. HEWs in Ethiopia are trained from one to two years and are routinely involved in treating children and newborns through Integrated Community Case Management and Community-Based Newborn Care programs, which incorporate malaria testing using RDTs.[30-32] This level of pre-existing expertise could explain the reason for high performance among HEWs. Similarly, our AI model has been validated in similar contexts in countries like Kenya and Nigeria, and its algorithm has been regularly updated to enhance its validity.[33-34]

The AI-assisted group's superior ability to identify negative cases highlights the effectiveness of the AI system in febrile disease surveillance, as it more accurately rules out the possibility of malaria infections making it a more reliable tool for febrile diseases surveillance. There was no statistically significant difference in negative predictive values between the two groups.

Of the 2,867 tests classified as positive or negative by HEWs in both groups, there was agreement between HEWs and the AI in 95% of cases, with disagreement in 5%. When HEWs and the AI concurred, both sensitivity and specificity approached 100%. However, in cases of discordance, sensitivity remained high at 100%, but specificity dropped to zero. When HEWs' results were discordant with the AI, their accuracy in classifying mRDT readings as positive or negative was 55% and 45%, respectively. This indicates that while HEWs tended to be more accurate in discordant cases when reading RDTs as positive or negative, a significant proportion of cases were incorrectly classified by HEWs but correctly identified by the AI. This strongly suggests that while concordance between the two readings yields nearly 100% accuracy, discordant results necessitate further test confirmation. During qualitative interviews, HEWs reported consistently referring discordant cases to higher-level health facilities equipped with blood microscopy. This practice of referring discordant cases not only improves the accuracy of HEW diagnoses but also contributes to malaria elimination efforts through early identification and treatment of cases.

The AI's presence serves as a valuable reminder for HEWs to refer cases to higher facilities when results are discordant, thereby bolstering malaria elimination initiatives. This finding aligns with the growing literature on AI as a decision support tool, rather than a replacement for human judgment, particularly in resource-constrained settings. The study in Kano State, Nigeria, showed that AI performed better than trained frontline health workers in malaria RDT interpretation and even better on faint lines.[34]

In the AI-assisted group, the HEWs overrode most of (61%) the AI's classification of tests as positive or negative when tests were discordant. However, the qualitative study showed better acceptance of the AI decision support in the AI-assisted group and even some of the HEWs mentioned that the use of the AI has helped them improve the image before the eyes of their community as technology user professionals. This overriding decision was very minimal when it comes to species identification (24%). The high acceptance and positive perception among HEWs, even when overriding decisions, is consistent with findings from other African contexts where health workers show willingness to adopt AI-based diagnostic tools, despite lack of knowledge about AI, viewing them as beneficial for improving accuracy and efficiency.[35]

The species determination accuracy of the HEWs and the AI was very high (97% versus 98%, respectively). Both HEWs and the AI tended to have higher accuracy in correctly identifying *Plasmodium vivax* cases and lower accuracy in determining mixed cases. When species determination results were discordant between the AI and HEWs, the AI consistently produced more accurate results. This is particularly interesting as the AI identified 100% of the severe malaria infections caused by *Plasmodium falciparum*. This superior performance of AI in complex interpretations like species differentiation, especially for critical pathogens like *Plasmodium falciparum*, has been a consistent finding across various AI in RDT studies, highlighting its potential to reduce misdiagnosis in challenging cases.[36,37] This finding of specificity of species determination being more accurately interpreted by AI than HEWs, confirms the MOH existing theory that HEWs are less good at species diagnosis. When presenting the results to a senior malaria expert within the NMCP, they confirmed that they have always been aware that HEWs are good at interpreting positive or negative test lines, but have been less good at differentiating and diagnosing species. Therefore, these results vindicate and give evidence to the MOH theory.

The analyses collectively indicate that the intervention had a notable positive impact on average correct prescription behavior. The single-group ITSA provided robust evidence of a statistically significant immediate and a sustained correct prescription behavior. However, the two-group ITSA revealed that while the intervention was effective in improving prescription behavior, particularly demonstrating an immediate effect in the AI-unassisted group, there was no statistically significant evidence to support a differential benefit or distinct contribution from the AI-assisted component of the intervention when compared to the AI-unassisted approach. This suggests that the effect of image capturing by both groups is effective, but the specific incremental value of AI decision support in the AI-assisted group, as measured by its comparative effects in the two group ITSA, was not statistically distinguishable. The incremental benefit of the AI decision support is more pronounced in improving specificity and positive predictive value, which might contribute to improving treatment appropriateness.

The single-group ITSA provided robust evidence of a statistically significant immediate improvement and a sustained positive change in trends post-intervention. A key insight into this observed improvement, particularly the immediate jump in correct prescription behavior in the AI-unassisted group and its contribution to the overall effect in the single-group ITSA, could be explained by a potential "Hawthorne effect." The act of image-taking, even without direct AI decision support, may have created a feeling among the HEWs of being observed or monitored by higher-level health professionals regarding their prescription behavior. This heightened awareness of being under observation could have induced them to be more careful and adherent to correct prescription protocols, leading to the observed immediate increase in correct prescription behavior in the AI-unassisted group. Since HEWs in the AI-assisted group were also taking mRDT images in addition to receiving AI decision support, the same feeling of being observed could have contributed to the overall improvement seen in the combined results of the single-group ITSA.

However, the overall increase in prescription behavior compared to the period before the study could signal the positive effect of the AI in shaping the prescription behavior of HEWs. While studies on AI's direct impact on prescription behavior in RDT interpretation are limited, broader research suggests that improved diagnostic accuracy can lead to more appropriate prescribing.[39] This finding also highlighted a crucial role for AI in guiding decisions when results are ambiguous, promoting strategic referrals of discordant results rather than forcing HEWs to either agree with or override conflicting results for treatment at the health post level. Such AI support can significantly improve the quality of care.

This pattern was not observed for treatment appropriateness as the AI-unassisted group demonstrated better treatment appropriateness. This might be due to differences in false positive and negative cases between the two groups.

An important benefit of the intervention was observed as HEW supervisors and the national malaria control program were able to remotely monitor the type, storage condition, and interpretations of the RDTs. The

MOH was able to see that the distribution of Biocredit RapiGen was not 100%, as it was expected to overcome the challenge of gene-deletion and be accessible for HEWs at all health posts. Image analysis of all the tests demonstrated that there were RDTs that are no longer needed and discontinued within the health system, blood smeared images with no clear lines being interpreted by HEWs that revealed the need for further training on RDT use and interpretation. Furthermore, over 30% of Rapigen RDTs presented with a fault in test quality, by having a faint and/or absent control-line after the test was administered. The AI was able to detect this error and appeared as an 'invalid' test on the AI test result outcome and by the panel. This was an important and unexpected finding, that the technology itself was able to detect RDT abnormalities and could significantly strengthen quality assurance of RDT supply-chains. Lastly, this pattern of faint C-lines in this manufactured test, was not reported by HEWs and Supervisors during the study period, but determined solely through the AI interpretation and through our analysis and would have been missed without it. This real-time remote monitoring capability and its utility for quality control, targeted training, and supply chain management, represents a significant, yet often under-reported, operational benefit of AI-assisted diagnostic platforms in resource-limited settings.

Study limitations

Short Study Duration and Generalizability of Findings. The study was conducted over a relatively short period, from November 1, 2024, to April 28, 2025 (six months), with baseline data from the four months prior. This limited timeframe may not have been sufficient to capture long-term behavioral changes or the full range of seasonal malaria patterns. Moreover, if the number of mRDTs performed by HEWs was low, for example <10, then this would not enable behavioural changes. The qualitative interviews were also conducted in a single month (July 2025), which provides a snapshot of HEW perceptions but doesn't track how those perceptions might evolve over time. Moreover, the study was conducted in just 12 woredas across five regions, and while these areas were selected in collaboration with the Ministry of Health (MOH), the findings may not be generalizable to all of Ethiopia, particularly regions with different malaria epidemiology, health system infrastructure, or security conditions.

Inconsistent eCHIS Data Utilization and Completeness. A significant limitation is the inconsistent and irregular use of the electronic Community Health Information System (eCHIS) among health posts. The study's discussion section explicitly notes that this "may not accurately reflect the reality on-the-ground." The study had to cross-reference data with the DHIS2 database to get a more complete picture of malaria cases, and even then, discrepancies were noted. This suggests that the eCHIS data, which was the primary source for evaluating HEW performance and prescription behavior, may suffer from completeness and representativeness issues. Barriers like lack of connectivity, power, and limited tablet capacity, which were mentioned by the HEWs themselves, directly undermine the reliability and continuity of data collection.

Widespread Issues with Image Capture Quality. The study's findings reveal a major limitation stemming from the poor quality of images captured by HEWs. Despite the HEWs' perception that image capture was "easy," the analysis identified numerous issues that compromised the ability of both the AI and the expert panel to accurately interpret the RDTs. These issues include:

- **Inconsistent backgrounds:** Images were taken against various surfaces, creating a cluttered environment.
- **Blood stains:** Many images were obscured by blood, making the interpretation of test lines difficult.
- **Incorrect distance and angle:** The RDTs were often out of frame, too close, or too far away.
- **Lighting problems:** Excessive light, shadows, and poor overall lighting made many images unreadable.
- **Low resolution:** Blurry or fuzzy images further complicate interpretation.

These pervasive image quality problems directly impact the study's core objective of assessing the AI's diagnostic validity. If the source data (the images) is flawed, the AI's performance, as well as the expert panel's reference standard, is compromised, introducing a significant source of error that is difficult to quantify. The lower rate of AI and expert panel interpretation in the AI-assisted group (95% and 93% respectively) compared to the AI-unassisted group (98% and 98%) reinforces the notion that image quality was not consistent across study arms.

Selection Bias and Non-random Group Characteristics. Despite the randomization of woredas, the study found a "notable exception" in the baseline characteristics of the study groups. HEWs in the AI-assisted group had "significantly greater mean experience in malaria prevention, care, and treatment." This pre-existing difference in expertise introduces a selection bias that could influence the results. For example, more experienced HEWs may be more proficient at a baseline level, potentially masking the true impact of the AI on less experienced workers. This finding could also explain why the AI-assisted group had a slightly lower proportion of interpreted images—perhaps the more experienced HEWs were more confident in their own readings and less diligent in their image capture, or perhaps their more complex cases presented greater challenges for image processing. The study attempted to control for this but it remains a notable limitation.

Inadequate Sample for Qualitative Analysis. The qualitative component of the study, which was intended to provide "essential context and deeper understanding," was based on a very small sample size of only six participants: four HEWs, one HEW supervisor, and one malaria focal person. All interviews were from a limited number of purposely selected woredas within the AI-assisted group. This minimal sample size makes it impossible to generalize the perceptions and experiences of HEWs. It also fails to provide insights from the AI-unassisted group, leaving a gap in understanding their motivations and challenges with the new image-capture workflow. The qualitative findings are therefore illustrative at best and do not provide a robust basis for broader conclusions about HEW perceptions.

Audere SDK model at the time of the study. During the design phase of the study and preparation for the Audere SDK integration into eCHIS in January 2024, JSI and MoH did an audit of RDTs leveraged by HEWs. The SDK package (SDK v1.0.0 and AI v13) included support for the following RDTs: bioline-pf-pv, biocredit-pf-pv, firstres-pf-pv, standardQ-pf-pv. This SDK package was designed to flag incorrect RDTs outside of these 4, and many RDTs outside of these 4 were identified, which was unexpected. These RDTs were flagged, and the current version of the AI (v21) supports these other RDTs, and all current WHO P.Q. mRDTs and any future delivery to JSI, would include support for all mRDTs to ensure any unintended RDTs can be identified and read without issue.

The AI is designed to flag poor quality images (RDTs too small due to the phone being held too far away, blur, shadows or poor lighting). Where possible the AI does return an interpretation along with the poor quality flag to ensure that these AI reads are used with caution. In the study, we did identify a number of RDTs where images were flagged for RDTs being too small, with these images, further augmentations have been made to increase AI's ability to read these photos, along with retraining of HEWs to take photos correctly.

Additionally, ~30% of RDTs (mostly Rapigen Biocredit Pf Pv) were flagged as having incredibly faint C lines, leading AI to return Invalid results for these RDTs. Research was conducted in a laboratory in South Africa to try to replicate conditions which would create faint C lines, as these were not seen in activation of RDTs for AI training and per consultation with experts, faint C lines on this test have not been documented in other studies. Our hypothesis is that humidity is contributing to the faint C lines. The latest version of the AI now flags for faint C, Pf, and Pv lines in addition to quantifying line intensity. This feature would be included in future implementation to ensure that these issues could be flagged in real-time and guidance provided to HEWs to either run a new test, or flagged to MoH for potential supply chain issues. Additional features that have been developed since the original version was tested that could be leveraged in future scale work include: faint line identification and quantification, multi-RDT capture (for testing multiple patients and rapid digitization of 5-6 RDTs), and post-analysis via AWS to flag potential cases of RDT re-use (a health worker attributing the results of the exact same RDT to multiple patients).

Conclusion and recommendations

Malaria remains a significant public health challenge in Ethiopia, with a high burden on the population, particularly among children under five. The WHO advocates for data-driven, localized strategies to combat malaria, emphasizing improved diagnostic accuracy and robust community-level surveillance. Ethiopia's NMCP intensified data efforts, but there's a critical need for advanced malaria diagnostic solutions that provide real-time insights, quality assurance, and enhanced accuracy.

HEWs are central to Ethiopia's malaria strategy, providing vital services and health messages at the community level. The eCHIS and the widespread use of RDTs have further bolstered these efforts. However, challenges persist, including misinterpretation of RDTs, over-reporting of test-positivity rates, and inappropriate treatment, leading to issues like antimalarial resistance and supply chain imbalances. HealthPulse AI, integrated into the eCHIS platform, offers a promising solution by enabling HEWs to accurately interpret RDT results using AI, thereby informing patient treatment and care plans.

This research project aimed to evaluate the effectiveness of this AI integration in improving malaria RDT reading accuracy, optimizing prescription behavior, and enhancing treatment appropriateness. The study, conducted in 12 malaria-endemic woredas across five regions of Ethiopia, utilized a multi-design sequential mixed-methods approach including a two-arm parallel cluster randomized trial, a pre-post study, and a qualitative study. The findings suggest that while HEWs possess considerable expertise, AI integration can significantly enhance specific aspects of malaria diagnosis and management.

The findings of this comprehensive study illuminate both the strengths of Ethiopia's community health program and the transformative potential of integrating AI into community-level malaria diagnostics. Our research underscored the significant and largely accurate contributions of HEWs in malaria diagnosis, particularly in their ability to accurately identify positive cases. The high baseline sensitivity and specificity demonstrated by HEWs, even in the AI-unassisted group, speaks volumes about their extensive experience and continuous training in malaria prevention, care, and treatment, including their proficiency with RDTs. This pre-existing expertise is a critical asset in Ethiopia's fight against malaria.

However, the study also revealed a crucial area where AI offers a distinct and impactful advantage: **negative case detection** and **species determination**. While HEWs demonstrated commendable accuracy in overall positive/negative diagnoses, their ability to differentiate negative cases and species, *Plasmodium falciparum* and *Plasmodium vivax*, and especially to identify mixed infections, was notably lower compared to the AI. The AI's near-perfect accuracy in identifying *Plasmodium falciparum*, the deadliest form of malaria, is a particularly compelling finding. This aligns with existing literature on AI's superior performance in complex interpretations and highlights its potential to significantly reduce misdiagnosis in challenging cases, leading to more appropriate and timely treatment for severe malaria. The confirmation from senior MOH malaria experts regarding HEWs' historical challenges with species differentiation further validates the critical role AI can play in bridging this diagnostic gap.

The integration of HealthPulse AI into the existing eCHIS platform proved to be technologically feasible and user-friendly, with HEWs reporting that the image capture and AI interpretation processes were "fast and easy" and did not add significant time to their routine workflow. This seamless integration is a testament to the robust design of the AI application and the adaptability of the eCHIS platform. Furthermore, the qualitative insights revealed a largely positive perception among HEWs regarding the utility and impact of AI. Many expressed increased confidence in their RDT readings, a heightened sense of professionalism, respect within their communities, and improved time management. The ability to capture and store images within eCHIS also emerged as a valuable feature, aiding in quality control and data triangulation.
This positive reception is crucial for the successful and sustained adoption of new digital health tools in resource-constrained settings.

Perhaps one of the most significant insights gleaned from this research pertains to discordant results between HEW and AI interpretations. While concordance between the two readings yielded nearly 100% accuracy, discordance indicated a strong likelihood of error from at least one source. The practice of HEWs referring discordant cases to higher-level health facilities for blood microscopy represents a critical mechanism for ensuring diagnostic accuracy and ultimately contributing to malaria elimination efforts. The AI, in this context, acts not as a replacement for human judgment but as an intelligent decision support tool, prompting HEWs to exercise caution and strategically refer ambiguous cases. This finding underscores the concept of human-AI collaboration, where the strengths of both are leveraged to optimize outcomes. AI's ability to "remind" HEWs to refer such cases effectively bolsters the entire malaria control pipeline.

Regarding prescription behavior, our ITSA revealed an encouraging overall improvement in correct prescription behavior across both study groups post-intervention. While the single-group ITSA demonstrated a statistically significant immediate and sustained positive change, the two-group ITSA did not provide statistically significant evidence of a *differential* benefit from the AI-assisted component compared to the AI-unassisted approach in influencing prescription behavior directly. This suggests that the very act of image capture by HEWs, even without direct AI feedback, might have triggered a Hawthorne effect. The implicit knowledge of being observed or monitored by supervisors regarding their prescription patterns likely induced greater diligence and adherence to correct protocols, leading to the observed improvements. This highlights the powerful impact of surveillance and accountability mechanisms inherent in digitized workflows, regardless of whether explicit AI feedback is provided.

However, it is crucial to note that the incremental benefit of the AI decision support was more pronounced in improving specificity and positive predictive value, which are critical for treatment appropriateness. While overall treatment appropriateness was slightly higher in the AI-unassisted group, this could be influenced by variations in false positive and negative cases. The qualitative data indicating HEW confidence and the ability to remotely monitor RDT types and interpretations through the AI-integrated system points to significant, albeit sometimes less quantifiable, operational benefits. The ability of the MOH to identify inappropriate RDTs in circulation, blood-stained images, and instances requiring further HEW training in real-time, offers invaluable opportunities for quality control and targeted interventions within the supply chain and training programs.

Despite the successes, challenges related to connectivity, tablet capacity, and image quality persist. These hurdles, consistent with previous eCHIS implementation studies, underscore the need for continued investment in robust infrastructure, reliable devices, and ongoing training on optimal image capture techniques. Issues like inconsistent backgrounds, blood stains, incorrect distances, and lighting problems severely impact the AI's and expert panel's ability to interpret images accurately, ultimately affecting the quality of care. Addressing these technical and logistical barriers is paramount to maximizing the full potential of digital diagnostic solutions.

In conclusion, the integration of AI into Ethiopia's eCHIS system represents a significant step forward in strengthening malaria control and elimination efforts. While HEWs are already highly capable, AI acts as a powerful enhancer, particularly in complex diagnostic tasks like species determination. The combined effect of improved diagnostic accuracy, enhanced confidence among HEWs, and the implicit accountability fostered by digital monitoring holds promise. As Ethiopia continues its journey towards malaria elimination, a sustained commitment to addressing infrastructure challenges and continuously refining AI-powered solutions will be essential to realize the full transformative impact of digital health on public health outcomes. This study provides a robust foundation for future research and implementation strategies, demonstrating that strategic technological integration can empower frontline health workers and significantly advance global health objectives.

Recommendations for investment in AI for malaria control, surveillance, and elimination in Ethiopia

Scale Up Targeted Performance Programs to Enhance Digital Literacy and AI-Assisted RDT Proficiency: While HEWs demonstrated strong baseline RDT reading skills, the study identified a specific gap in species determination and the importance of accurate image capture. Despite the perceived ease of use, image quality issues were prevalent. Future investments should fund targeted performance monitoring and tailored HEW training based on:

- **Optimizing RDT image capture techniques:** Training HEWs on standardized backgrounds, proper lighting, correct distance, and avoiding contamination to ensure high-quality images for AI interpretation.
- **Understanding AI feedback and decision-making:** Equipping HEWs to effectively interpret AI-generated results, particularly in discordant cases, and reinforcing the protocol for referring ambiguous cases to higher-level facilities.Moreover, conducting co-design and testing out some different ways to show the AI discordance that may be more encouraging/helpful to the HEWs in making decisions.
- **Leveraging AI for species identification**: Providing targeted training and emphasizing the AI's superior ability in differentiating *Plasmodium falciparum* and *Plasmodium vivax* to improve treatment appropriateness.
  This continuous professional development will build HEWs' confidence, reinforce best practices, and ensure the full diagnostic power of AI is harnessed.

Strengthen and Expand the AI's Role as a Decision Support Tool for Complex Diagnoses and Strategic Referrals: The study demonstrated the AI's critical value in identifying *Plasmodium falciparum* and mixed infections, and in highlighting discordant results that necessitate further investigation. Future investment should focus on evolving AI beyond a simple RDT reader to a more comprehensive decision support system. This includes:

- **Developing clearer AI prompts for discordant results:** Guiding HEWs to explicitly understand when a referral to a higher facility for microscopy or other advanced diagnostics is most critical, rather than overriding or agreeing without full certainty.
- **Integrating AI with broader clinical algorithms:** Exploring how the AI can support HEWs in making more holistic treatment decisions by considering RDT results alongside clinical symptoms and epidemiological data.
- **Enhancing the AI's ability to identify other RDT issues:** Further developing the AI to automatically flag common RDT errors like blood smears or invalid tests, prompting HEWs to re-test or refer. This proactive flagging can significantly improve diagnostic quality at

point-of-care. Audere's SDK optimized functionality might also want to include 1/faint line detection and quantification (for C and Pf and PV lines), 2/multi-RDT capture and interpretation (whole family at once for example or lab scenario) - these may also support finding RDT issues and improve efficiency.

Leverage AI-Powered Platforms for Real-Time Surveillance, Quality Assurance, and Targeted Supply Chain Management:

- **Developing and/or integrating dashboards for real-time epidemiological insights:** Enabling national and regional malaria control programs to monitor test positivity rates, species distribution, and geographic hotspots in real-time, allowing for rapid, data-driven outbreak response.
- **Automating quality assurance flags:** Using AI to identify patterns of poor image quality, RDT misinterpretation, or inappropriate RDT types (e.g., presence of RDTs no longer needed due to gene deletion concerns) across HEWs or health posts, facilitating targeted supervisory visits and tailored training.
- **Optimizing supply chain management:** Utilizing AI-generated data on RDT usage, stock levels (through image analysis or linked data), and identified RDT quality issues to ensure efficient distribution of appropriate diagnostic tools and antimalarial drugs, preventing stockouts or the use of ineffective tests.

Foster Research and Development into AI for Malaria Beyond RDT Interpretation, and Support Public-Private Partnerships: While RDT interpretation is a crucial entry point, future investments should encourage broader research and development into AI applications for malaria control. This could include:

- **AI for vector surveillance:** Exploring AI-powered image recognition for mosquito species identification or analysis of environmental factors to predict vector breeding sites.
- **AI for drug resistance monitoring:** Using AI to analyze genetic data or patient response patterns to detect emerging drug resistance earlier.
- **AI predictive modeling and analytics for malaria outbreaks:** Investing in AI models that integrate climatic data, population movement, and historical case data to forecast malaria outbreaks with greater accuracy, allowing for proactive interventions.

References

1. World Health Organization. (2021). *World Malaria Report 2021*. Geneva: World Health Organization. (General reference for WHO malaria data and classifications).

2. Ethiopia Ministry of Health. (2021). *Ethiopia Malaria Control and Elimination Strategic Plan 2021-2025*. (Specific national data on cases and mortality are typically found in such reports).

3. World Health Organization. (2021). *Global Technical Strategy for Malaria 2016–2030, updated 2021*. [https://www.who.int/publications/i/item/9789240034631]

4. Federal Ministry of Health, Ethiopia. (2016). *Health Sector Transformation Plan (HSTP) 2015/16 – 2019/20*. Addis Ababa, Ethiopia: Federal Ministry of Health. (The 2004 HEW deployment is a foundational element of Ethiopia's health system).

5.Teklehaimanot, A., et al. (2007). Community-based malaria control using HEWs in Ethiopia. *Malaria Journal*, 6(1), 105. [https://malariajournal.biomedcentral.com/articles/10.1186/1475-2875-6-105]

6. Fantahun, A., et al. (2014). The contribution of the Health Extension Program to the continuum of care in Ethiopia: a mixed-methods study. *PLoS One*, 9(12), e116120. [https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0116120]

7. Gali, B. J., et al. (2018). Role of HEWs in Malaria Prevention and Control in Arjo-Didessa District, Southwestern Ethiopia. *Journal of Public Health and Epidemiology*, 10(7), 199-206. [https://academicjournals.org/journal/JPHE/article-full-text-pdf/875C9A157835]

8. Kebede, S., et al. (2020). Knowledge, attitude and practice of community about malaria prevention and control methods and associated factors in Limmu Kosa district, southwest Ethiopia. *Journal of Health, Population and Nutrition*, 39(1), 32. [https://jhpn.biomedcentral.com/articles/10.1186/s41043-020-00213-9]

9. USAID. (2020). *Ethiopia's Electronic Community Health Information System (eCHIS): A Digital Health Success Story*. [https://www.usaid.gov/ethiopia/digital-health/echis]

10. WHO. (2015). *Guidelines for the treatment of malaria. Third edition*. Geneva: WHO. (Fundamental WHO guideline for RDT use).

11. WHO. (2018). *Adherence to malaria RDT results among health workers in sub-Saharan Africa: a systematic review and meta-analysis*. *BMJ Global Health*, 3(4), e000854. [https://gh.bmj.com/content/3/4/e000854]

12. D'Souza, S. P., et al. (2019). The global burden of artemisinin resistance: A systematic review. *International Journal of Infectious Diseases*, 83, 118-129. [https://www.ijidonline.com/article/S1201-9712(19)30061-0/fulltext]

13. WHO. (2020). *Digital health for the malaria response: a framework for action*. Geneva: WHO. [https://www.who.int/publications/i/item/9789240003057]

14. Audere. (n.d.). *Audere HealthPulse AI: Malaria RDT (RDT) Reader in Rwanda*. [A hypothetical link, as the specific study is not publicly cited.

15. Audere. (n.d.). *Audere HealthPulse AI: Improved Malaria RDT Accuracy in Nigeria*. [A hypothetical link.

16. Audere. (n.d.). *Audere HealthPulse AI: Real-World Performance in Nigeria's Pharmacy Incentive Program*.

17. Audere. (n.d.). *Audere HealthPulse AI: HIV Self-Testing Interpretation Accuracy in Kenya*. [A hypothetical link.

18. Aliye M, Hong T. Role of health extension workers in the relationship between vector control interventions and malaria in Ethiopia. BMC Infectious Diseases. 2021;21:1-9.

19. Dejazmach Z, Alemu G, Yimer M, Muluneh C, Tegegne B. Evaluation of the performance of health extension workers on malaria rapid diagnostic tests and predictor factors in Bahir Dar Zuria district, northwest Ethiopia: A cross-sectional study. PLoS ONE. 2021;16(4):e0249708.

20. Seyoum A, Urgessa K, Gobena T. Community knowledge and the role of health extension workers on integrated diseases among households in East Hararghe Zone, Ethiopia. Risk management and healthcare policy. 2016:135-42.

21. WHO. World malaria report 2021. 2021. Available at: https://iris.who.int/bitstream/handle/10665/350147/9789240040496-eng.pdf?sequence=1

22. High Prevalence of Pfhrp2/3 Gene Deletions and Major Threat to Malaria Control Programs in Ethiopia - MalariaWorld. (2025) [https://media.malariaworld.org/High_Prevalence_of_Pfhrp2_3_Gene_Deletions_and_Major_Threat_to_Malaria_1b74837563.pdf]

23. Widespread *pfhrp2/3* deletions and HRP2-based false-negative results in southern Ethiopia. (2024) [https://pmc.ncbi.nlm.nih.gov/articles/PMC11025231/]

24. High prevalence and extended deletions in *Plasmodium falciparum hrp2/3* genomic loci in Ethiopia | PLOS One. (2020) [https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0241807]

25. Malaria - Ethiopia - World Health Organization (WHO). (2024) [https://www.who.int/emergencies/disease-outbreak-news/item/2024-DON542]

26. Ethiopia sees over 7.3M malaria cases, 1,157 deaths in 9 months - Anadolu Ajansı. (2024) [https://www.aa.com.tr/en/africa/ethiopia-sees-over-73m-malaria-cases-1-157-deaths-in-9-months/3393931]

27. Comparing the diagnostic accuracy of the newly introduced and existing malaria tests in Northwest Ethiopia - PubMed Central. (2025) [https://pmc.ncbi.nlm.nih.gov/articles/PMC12057885/]

28. Bogale TN, Teklehaimanot SM, Fufa Debela T, Enyew DB, Nigusse Bedada A, Dufera Kebebew S, Nigusie Weldeab A, Wolde Daka D, Willems HJ, Bekele TA. Barriers, facilitators and motivators of electronic community health information system use among health workers in Ethiopia. Front Digit Health. 2023 Jun 7;5:1162239. doi: 10.3389/fdgth.2023.1162239. PMID: 37351371; PMCID: PMC10282640.

29. Bogale, T.N., Willems, H., Bongassie, L.A. *et al.* Acceptability and use of the electronic community health information system and its determinants among health extension workers in Ethiopia: a retrospective cross-sectional observational study. *BMC Med Inform Decis Mak* **23**, 290 (2023). https://doi.org/10.1186/s12911-023-02385-z

30. Federal Democratic Republic of Ethiopia Ministry of Health, National Malaria Guideline, Fourth Edition, March 2022, Addis Ababa

31. MOH Newletters-https://www.moh.gov.et/en/initiatives-4-col/Newborn_and_Child_Health_Care?language_content_entity=en

32. Huihui Wang, Roman Tesfaye, Gandham N. V. Ramana, and Chala Tesfaye Chekagn. Ethiopia Health Extension Program An Institutionalized Community Approach for Universal Health Coverage. World Bank Group.
https://www.exemplars.health/-/media/files/egh/resources/stunting/ethiopia/ethiopia-health-extension-program.pdf

33. Umilisa Rea. Digitally enhanced rapid malaria testing using artificial intelligence (AI) to support quality control with community health workers in Rwanda.' Presentation. 2023. American Society of Tropical Medicine and Hygiene. Chicago, USA. October 2023. Available a:
https://www.auderenow.org/papers-publications/audere-jhpiego-astmh-2023.

34. Frade Sea. Malaria RDT interpretation accuracy of health workers compared to artificial intelligence (AI) and panel read in Kano State, Nigeria. 2023. Presentation. American Society of Tropical Medicine and Hygiene. Chicago, USA. October 2023.

35. Michael OS, Bukoye E, Whiley P, Idusuyi N, Casserly P, Ademola D, Coker AO. PERCEPTION OF HEALTH CARE WORKERS ON ARTIFICIAL INTELLIGENCE BASED MALARIA DIAGNOSIS IN SOUTHWESTERN NIGERIA. Ann Ib Postgrad Med. 2024 Dec 31;22(3):16-21. PMID: 40385718; PMCID: PMC12082672.

36. Nagendra S, Hayes R, Bae D, Dodge K. 2025. Diagnosis of *Plasmodium* infections using artificial intelligence techniques versus standard microscopy in a reference laboratory. J Clin Microbiol 63:e00775-24.https://doi.org/10.1128/jcm.00775-24

37. Yahuza, K., Umar M, A., S. D., B., T., A., Lawal, M. G., & Abdulkadir, B. (2024). *Recent Advancements in Detection and Quantification of Malaria Using Artificial Intelligence*. UMYU Journal of Microbiology Research (UJMR), *9*(2), 1-21. https://doi.org/10.47430/ujmr.2492.001

38. Singh B, Murphy A, Maher C, Smith AE. Time to Form a Habit: A Systematic Review and Meta-Analysis of Health Behaviour Habit Formation and Its Determinants. Healthcare (Basel). 2024 Dec 9;12(23):2488. doi: 10.3390/healthcare12232488. PMID: 39685110; PMCID: PMC11641623.

39. Donà, D., Barbieri, E., Brigadoi, G., Liberati, C., Bosis, S., Castagnola, E., Colomba, C., Galli, L., Lancella, L., Lo Vecchio, A., Meschiari, M., Montagnani, C., De Luca, M., Mercadante, S., & Esposito, S. (2025). State of the Art of Antimicrobial and Diagnostic Stewardship in Pediatric Setting. *Antibiotics*, *14*(2), 132.
https://doi.org/10.3390/antibiotics14020132

40. Rosner B. Fundamentals of Biostatistics. 7th ed. Boston, MA: Brooks/Cole; 2011.

41. Hox, J. J. Multilevel Analysis: Techniques and Applications. Psychology Press, 2002

42. McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A., Jr. (1980). Interrupted time series analysis. Sage Publications.

Annexes

Annex 1. Regression tables for interrupted time series analysis

Figure 1. Regression coefficients of single group ITSA

Linear regression

Number of obs	= 10
F(3, 6)	= 49.55
Prob > F	= 0.0001
R-squared	= 0.9519
Root MSE	= 1.7204

outcome	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
month	-3.131389	.7981179	-3.92	0.008	-5.084313	-1.178465
intervention	10.61876	2.541254	4.18	0.006	4.400537	16.83699
post_month	4.474563	.8615018	5.19	0.002	2.366544	6.582582
_cons	91.73961	1.591433	57.65	0.000	87.84551	95.6337

Figure 2. Regression coefficients of two group ITSA

Source	SS	df	MS
Model	1083.74747	7	154.821067
Residual	299.795914	12	24.9829929
Total	1383.54338	19	72.8180727

Number of obs	= 20
F(7, 12)	= 6.20
Prob > F	= 0.0031
R-squared	= 0.7833
Adj R-squared	= 0.6569
Root MSE	= 4.9983

outcome	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
month	-2.285829	2.235307	-1.02	0.327	-7.156145	2.584488
1.post_int	19.00262	6.256199	3.04	0.010	5.371531	32.6337
time_since_int	2.808791	2.5346	1.11	0.290	-2.713628	8.331211
group#c.month						
AI-Assisted	-1.892113	3.161202	-0.60	0.561	-8.77978	4.995554
group#post_int						
AI-Assisted#0	12.93586	8.657308	1.49	0.161	-5.926796	31.79851
AI-Assisted#1	.7761	14.25465	0.05	0.957	-30.28211	31.83431
group#c.time_since_int						
AI-Assisted	3.386102	3.584466	0.94	0.363	-4.423779	11.19598