Case Report

Asynchronous Mobile Money Integration for Bypassing the Cash Desk for Patient Payments and Evaluating Payment Cycle Time Over 8 Months at a Rwandan Tertiary Hospital: Retrospective Observational Study

Djibril Mbarushimana, MMed; Jean Damascene Gatera, BSc; Yaqub Muganga, BSc; Oswald Benimana, BSc; Jean Nepomuscene Ntawurushimana, MSc; Felix Niyomugabo, BSc; Clarisse Kayitasire, BSc; Emmanuel Nsengiyumva, BSc; Adolphe Hitimana, BSc; Prosper Kabanguka, BSc; Liliane Ingabire, BSc; Issa Bagirinshuti, BSc; Felicite Mukamana, MMed; Theogene Twagirumugabe, MMed; Jules Ndoli Minega, MMed; Vincent Nsanganiye, BSc; Christian Ngarambe, MMed

University Teaching Hospital of Butare, Huye, Southern Province, Rwanda

Corresponding Author:
Djibril Mbarushimana, MMed
University Teaching Hospital of Butare
Mamba
Huye, Southern Province
Rwanda
Phone: 250 788 575 656
Email: djidji01@gmail.com

Abstract

Background: Patient payment queuing at hospital cashier desks is a persistent operational challenge in resource-limited health care settings. In October 2025, the University Teaching Hospital of Butare (CHUB), a public tertiary referral center in Rwanda, integrated an asynchronous Mobile Money payment module with its OpenClinic GA hospital management information system (HMIS) through the Irembo payment gateway, as an unfunded quality improvement initiative.

Objective: This study aimed to describe the integration architecture and quantify its impact on patient payment cycle time, channel adoption, and transaction reliability over 8 months of routine operation.

Methods: Performance was evaluated by retrospective analysis of all closed invoices for prepaid services (laboratory, imaging, and pharmacy) from October 1, 2025, to May 31, 2026 (n=146,203 episodes). Payment cycle time was defined as the interval from bill creation to payment confirmation. The primary channel comparison used a 1440-minute (24-hour) analytic cutoff; calendar-month–adjusted quantile regression and a sensitivity analysis on the full, untruncated duration distribution (raw medians; Kaplan-Meier estimation with log-rank test) tested whether this cutoff affected the channel comparison. Bootstrap 95% CIs were computed for median cycle times.

Results: The integration implements a fully asynchronous, nonblocking model with dual trigger pathways (patient-initiated SMS and staff-initiated HMIS form) and a 4-step Irembo API workflow. Across 146,203 invoiced episodes restricted to the prepaid (laboratory, imaging, and pharmacy) analytic cohort, Mobile Money completion grew from 3.8% (n=724) to 20.9% (n=3228) of monthly episodes. Among 131,037 within-24-hour episodes, median payment cycle time was 5 (95% CI 5-5) minutes for Mobile Money vs 72 (95% CI 71-72) minutes for cash, a difference of 67 (95% CI 66-67) minutes. This difference was materially unchanged after adjustment for calendar month (quantile regression coefficient –67.0, 95% CI –67.7 to –66.3 minutes; $P<.001$) and in a sensitivity analysis using the full, untruncated duration distribution (median 5, IQR 2-18 minutes vs median 88, IQR 34-263 minutes for cash, a 94.3% lower median; log-rank $P<.001$). Separately, the hospital-wide Mobile Money transaction log (all billing categories) comprised 44,216 unique transaction attempts with a 97.9% (43,278/44,216) completion rate and a 3.6% (3167/87,269 gateway attempts) per-attempt failure rate; 329 (0.7%) transactions failed permanently, with a generic gateway server-error response implicated in 302 of these (most plausibly reflecting insufficient patient balance). Within the prepaid cohort, 18,935 (12.9%) of 146,203 episodes were paid via Mobile Money. Insurer paper-documentation requirements were the principal identified constraint on adoption.

XSL•FO
RenderX

Conclusions: Asynchronous Mobile Money integration was associated with a substantially shorter patient payment cycle time and provides a reproducible reference architecture for sub-Saharan African hospitals. Because the payment channel was self-selected by patients and no preimplementation baseline was available, these findings are associational rather than a causal demonstration of queue elimination.

(JMIR Form Res 2026;10:e107597) doi: 10.2196/107597

KEYWORDS

asynchronous payment; digital health; hospital management information system; IremboPay; mobile money; OpenClinic GA; payment cycle time; Rwanda; sub-Saharan Africa

Introduction

Patient Payment Queuing as a Barrier to Health Care Access

Long payment queues at hospital cashier desks constitute a recognized, yet underreported, operational bottleneck in health care delivery across sub-Saharan Africa [1,2]. Patients and caregivers may spend considerable time awaiting payment clearance before receiving medications, diagnostic results, or clinical interventions, creating friction that delays care delivery and undermines patient experience [3]. In resource-limited settings, where staffing constraints and infrastructure limitations amplify service demand peaks, the consequences extend beyond individual inconvenience to systemic barriers to timely care [4,5]. A recurring, and largely untested, question in this literature is whether the residual barrier to digital payment adoption in such settings is primarily technical (the reliability of the payment rail itself) or institutional (workflow and documentation requirements that apply regardless of payment channel); this study is positioned to speak to that question directly, because it separately measures transaction-system reliability and the structural determinants of the adoption ceiling.

Mobile money platforms, operating through telecommunications networks, have achieved near-universal penetration in Rwanda and across the region [6,7]. In Rwanda and in comparable East African settings, mobile money is an increasingly common mechanism for settling health-related payments [8-10]. Integration of mobile money with hospital information systems offers the prospect of reducing reliance on physical queues for payment processing, allowing patients to settle invoices remotely via their mobile handsets [11]. A critical design distinction governs the clinical value of such integrations. Synchronous implementations block the clinical or administrative workflow pending payment confirmation, preserving the queue at a different point; asynchronous implementations, by contrast, dispatch the payment request to the patient's handset in the background, allow the workflow to continue without waiting, and confirm payment independently once the patient completes the mobile transaction [12]. The latter pattern is the subject of this report. Despite the recognized burden of hospital payment queuing and the growing availability of mobile money infrastructure, rigorous quantitative evaluations of asynchronous mobile money payment integration within routine hospital workflows, and of comparable mHealth architectures more broadly in sub-Saharan African hospitals, remain scarce in the peer-reviewed literature [13,14].

Institutional Context: CHUB

The University Teaching Hospital of Butare (CHUB; Centre Hospitalier Universitaire de Butare) is a 450-bed public tertiary referral center located in Huye, Southern Province, Rwanda [15]. The hospital operates OpenClinic GA, an open-source hospital management information system (HMIS), as its primary HMIS for patient registration, clinical documentation, laboratory information management, pharmacy dispensing, and financial billing [16,17]. The OpenClinic GA billing workflow generates patient invoices upon order fulfillment with server-recorded time stamps, providing the temporal precision necessary for cycle time analysis.

Prior to October 2025, all patient payments were processed exclusively through physical cashier desks, with no mechanism for remote or mobile payment. The queuing burden this generated was a documented operational concern, motivating the decision to implement Mobile Money integration as a continuous improvement initiative. The integration was developed in technical partnership with Irembo Ltd as payment gateway operator, without external funding.

Study Context and Objectives

This paper provides an integrated account of the intervention by describing the technical architecture of the Irembo-OpenClinic GA integration, including the payment trigger pathways, the transaction workflow, the background processing scheduler, and the system data model, and reporting the quantitative performance evaluation of the deployed system over 8 months of routine operation.

The objectives of this work were to (1) describe the integration architecture in sufficient detail to be reproducible within any OpenClinic GA deployment; (2) quantify patient payment cycle time by channel (Mobile Money vs cash) across the full study period; (3) characterize the growth of Mobile Money adoption over the 8-month postimplementation period; (4) assess the technical reliability of the Mobile Money transaction system, including failure rate, root-cause distribution, and evolution over time; (5) delineate the billing segments where server-generated time stamps permit rigorous cycle time analysis; and (6) identify the structural constraints on the adoption ceiling, testing specifically whether these constraints are technical or institutional, and formulate recommendations to further reduce cashier-desk dependence.

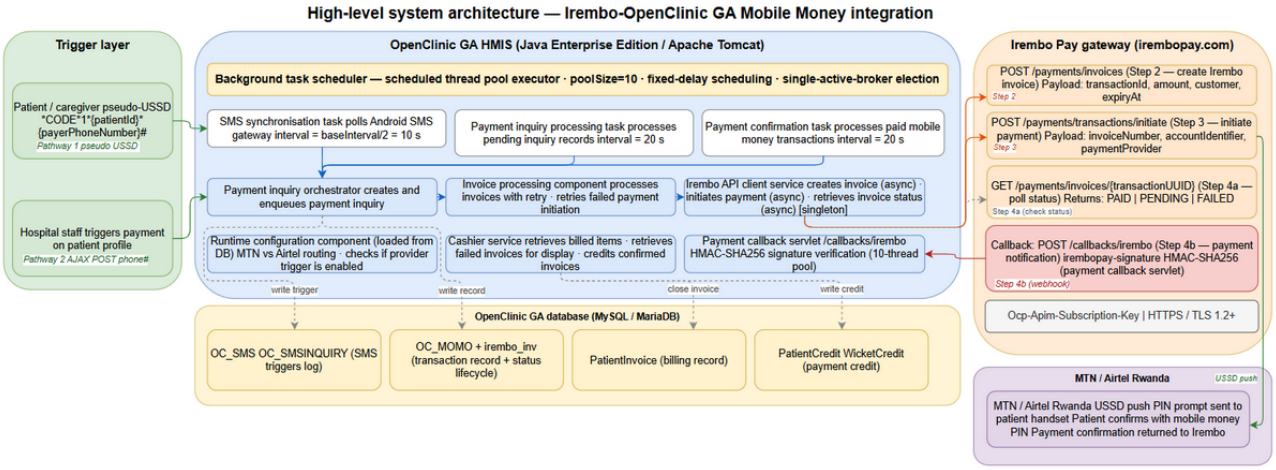
System Architecture Overview

The integration comprises 5 principal components interacting through well-defined interfaces: (1) the OpenClinic GA HMIS

XSL•FO
RenderX

application server, (2) the multithreaded background task scheduler, (3) the Irembo Pay Representational State Transfer (REST) API gateway, (4) the patient mobile handset via the MTN or Airtel mobile money network, and (5) the OpenClinic GA relational database. Figure 1 illustrates the high-level architecture. This component-based, asynchronous-broker design is consistent with the modular, adaptable-architecture principles described for other mHealth systems built for resource-constrained settings, such as the psychoeducational mHealth platform reported by Balderas-Díaz et al [18], in which a decoupled communication layer similarly separates patient-facing triggers from backend clinical logic.

Figure 1. High-level system architecture of the Irembo-OpenClinic GA Mobile Money payment integration at University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare). AJAX: Asynchronous JavaScript and XML; API: application programming interface; async: asynchronous; DB: database; GET and POST: HTTP request methods; HMAC-SHA256: hash-based message authentication code using Secure Hash Algorithm 256; HMIS: hospital management information system; HTTP(S): Hypertext Transfer Protocol (Secure); MTN: Mobile Telephone Networks; OC_MOMO, OC_SMS, OC_SMSINQUIRY: OpenClinic database tables for Mobile Money transactions, SMS triggers and SMS payment inquiries; Ocp-Apim-Subscription-Key: Azure API Management subscription key header; PIN: personal identification number; SMS: short message service; TLS: Transport Layer Security; USSD: Unstructured Supplementary Service Data; UUID: universally unique identifier.



The architecture implements a nonblocking asynchronous payment model [12,19]. Once a payment trigger is received (through either pathway described under Payment Trigger Mechanisms), the HMIS records the payment request in the database and returns control to the triggering workflow immediately. The background scheduler then processes the request, creates the Irembo invoice, and dispatches the payment initiation request to the patient's handset, all without blocking any downstream clinical or administrative step.

Full implementation detail supporting this overview, including the technology stack and configuration parameters, the 2 payment-trigger pathways, the background task scheduler, the 4-step Irembo API transaction workflow, retry and error-handling logic, and the extended OpenClinic GA data model, is provided in Multimedia Appendix 1.

Methods

Study Design and Setting

This was a retrospective, observational evaluation conducted at CHUB, covering the postimplementation period from October 1, 2025, to May 31, 2026. All analyses were conducted on routinely collected HMIS data with no patient contact or clinical data collection. This report follows the SQUIRE (Standards for Quality Improvement Reporting Excellence) 2.0 guidelines [20]; the completed checklist is provided in Multimedia Appendix 2.

Ethical Considerations

The project received ethical approval from the CHUB Research Ethics Committee (REC/UTHB/158/2024, approved August 26, 2024). The requirement for individual informed consent was waived by the Research Ethics Committee because this study was a retrospective, quality-improvement evaluation of routinely collected administrative HMIS records, with no patient contact, no change to clinical care, and no collection of new clinical data. All analyses were conducted on deidentified, aggregate episode- and transaction-level records (billing, invoicing, and payment time stamps; payment channel; and Mobile Money transaction status); no patient names, national identification numbers, or clinical content were extracted or analyzed, and access to the extracted datasets was restricted to the study team under CHUB institutional data governance. No participants were directly recruited or contacted for this study, and no participant compensation was applicable.

Data Sources

Two complementary datasets were extracted from the OpenClinic GA database. The first provided payment cycle time data, obtained by linking the patient billing records with the payment credit records and the Mobile Money transaction log, filtered to closed invoices with billing and payment time stamps falling within the same calendar day as the respective date fields. This ensured clock consistency within included records. The dataset captured the following information for each invoiced episode, including bill creation time, invoice generation time, payment confirmation time, payment channel (Mobile

XSL•FO
RenderX

Money via Irembo or cash desk), and whether a Mobile Money payment was triggered, regardless of the final payment channel.

The second dataset provided transaction-level performance data, obtained from the Mobile Money transaction log joined with a status-history audit table. This yielded, for each Mobile Money transaction, the unique transaction identifier, final payment status, mobile network operator, payee message, and the complete status history capturing each workflow step and any associated error messages.

Billing Segment Scope: Prepaid Services and Cycle Time Measurement

A critical methodological consideration concerns the reliability of time stamps used to compute payment cycle times. The HMIS records 2 categories of billing events with different temporal characteristics.

Laboratory investigations, diagnostic imaging, and pharmacy prescriptions operate on a prepaid basis within the CHUB workflow, whereby the HMIS generates a bill at the time the clinician places the order, and the ordered service is only executed or dispensed after that bill has been paid. Each such bill is billed individually and can, in principle, be settled on its own; in practice, a patient may accumulate more than one outstanding bill during a visit, including a mix of prepaid (laboratory, imaging, and pharmacy) and postpaid (consultation, ward, and procedure) items billed at different times, and OpenClinic GA can settle several outstanding bills together within a single payment invoice. This prepaid model still ensures a direct, unambiguous causal relationship between each bill and the payment that settles it because the patient must pay before receiving the specific prepaid service, and the interval between that bill's creation and the confirmation of the invoice that settles it represents the patient's actual payment waiting time for that item. Because both the billing time stamp and the payment confirmation time stamp are recorded automatically by the application server at the moment the respective events occur, these categories also provide the temporal precision required for reliable cycle time measurement. These 3 billing categories constitute the analytical scope of this study; the unit of analysis throughout is the individual prepaid bill, not the aggregate payment invoice (see the Statistical Analysis section).

Other billing categories at CHUB, including consultation fees, ward charges, and procedure fees, operate on a postpaid basis. The service is delivered first, and the patient settles the corresponding invoice at a later point, which may be hours or days after the service was rendered. For example, an inpatient procedure fee may be raised at the time of admission and paid only at discharge several days later. In this model, the interval between billing and payment no longer represents a payment waiting time in the operational sense relevant to queue elimination; it represents the length of the care episode. Including these categories in a payment cycle time analysis would therefore conflate clinical episode duration with payment behavior, which would confound the comparison between Mobile Money and cash payment channels. These categories are accordingly outside the scope of this analysis.

Variable Definitions and Transaction Classification

Payment cycle time was defined as the interval in minutes between bill creation time and payment confirmation time. The payment channel was classified as Mobile Money (via Irembo gateway) or cash (traditional cashier). An additional variable identified episodes where Mobile Money payment was triggered but ultimately completed at the cashier, enabling separate analysis of cash fallback rates. Both channels are ultimately settled via Mobile Money, as genuine cash (banknotes and coins) had already been phased out of routine use at CHUB before the study period, chiefly for infection-control reasons following the COVID-19 pandemic and recurring difficulties reconciling exchange/change balances. The label "cash channel" is retained here for continuity with existing hospital operational terminology and denotes payment processed manually at the cashier desk, with the cashier entering the transaction via the cashier's own device and the payer's Mobile Money code, rather than payment by physical currency; the substantive distinction between channels is therefore whether payment was triggered remotely by the payer (or a delegate) or processed in person by a cashier, not whether Mobile Money was used as the underlying payment instrument.

Transactions were classified into 3 operationally meaningful segments: *quick* (30 minutes or less), reflecting prompt payment processing; *normal* (31 to 1440 minutes), representing same-operational-day payments; and *extended* (beyond 1440 minutes), representing transactions completed beyond the HMIS-configured 24-hour Irembo invoice expiry window. Extended transactions are attributable to next-day batch data entry, deferred payment reconciliation, or other administrative processes rather than real-time patient waiting, and are excluded from performance comparisons but reported separately.

Mobile Money Transaction Performance Analysis

The final payment status was used to determine transaction outcome. Payment confirmed was classified as successful, failed invoice creation or failed payment initiation was classified as failed, and payment initiated without subsequent confirmation was classified as pending. The 4-step transaction workflow was reconstructed from the status history audit log: (1) local invoice created in the HMIS, (2) invoice creation request sent to the Irembo gateway (success or failure), (3) payment initiation request sent to the patient's handset (success or failure), (4) payment confirmed by the gateway. Failure events were classified by root cause from the associated error messages.

Statistical Analysis

Descriptive statistics are presented as median (IQR), mean (SD), and 90th percentile for cycle time data, and as counts and percentages for categorical variables. Two source extractions are used, at 2 different units of analysis, and they are not restricted to the same population. The episode-level extraction (n=146,203 total invoiced episodes, or the Mobile Money–triggered subset, n=22,073) is restricted to the prepaid (laboratory, imaging, and pharmacy) billing cohort defined in the Billing Segment Scope: Prepaid Services and Cycle Time Measurement section, and its unit of analysis is the individual prepaid bill; all cycle time, adoption, and channel-comparison

XSL•FO
RenderX

results in this paper use this extraction. The transaction-level extraction (n=44,216 total Mobile Money transaction attempts) is the Mobile Money gateway's own transaction log, generated at the level of the payment invoice rather than the individual bill, and spans all CHUB billing categories in which Mobile Money payment is offered, including postpaid categories outside the prepaid cohort. The principal reason these 2 counts cannot be reconciled one-to-one is structural rather than simply a missing join key. A single payment invoice, and therefore a single Mobile Money transaction, can settle more than one outstanding bill, potentially mixing prepaid and postpaid items billed at different times during the same visit, so the transaction count is not expected to equal, or to be a simple multiple of, the count of billed items. OpenClinic GA maintains a linking key between individual bills and the invoice that settles them, but joining it was outside the objectives of this study (see Introduction) and was not attempted. Cycle time, adoption, and channel-comparison results are established entirely from the episode-level extraction using each bill's own time stamps, and the transaction-system reliability and failure-rate results are established entirely from the transaction log using each transaction's own status history and final gateway message. Neither depends on linking a bill to the invoice that settled it, so a reconciled count of bills settled per Mobile Money invoice is not needed to support any finding reported in this paper, and its absence does not affect the reliability of either analysis; the transaction log additionally does not carry a billing-category field, so transactions cannot be restricted to the prepaid cohort after the fact regardless. Duplicate payment or double-crediting of the same bill is precluded independently of this reconciliation. In the OpenClinic GA schema, a bill can be attached to only one invoice, so the same bill cannot subsequently be attached to, and paid through, a second invoice, and the credit-entry logic that settles a paid invoice runs atomically and at most once per invoice (see the System Architecture Overview section; System Data Model and Credit Entry Mechanics section in Multimedia Appendix 1). Consequently, the transaction-level metrics (completion rate, failure rate, and root-cause distribution) describe hospital-wide Mobile Money transaction-system performance rather than the performance of Mobile Money transactions specifically within the prepaid analytic cohort. For the failure-rate analysis specifically, a Mobile Money transaction attempt (one row in the transaction log, identified by a unique transaction identifier) is defined as being created on each trigger or retrigger of an invoice; failure rates are reported per attempt (failed events divided by all attempts at that workflow step) and, separately, per transaction (permanently failed transactions divided by all transactions), rather than dividing all logged failure events—most of which are transient retries on transactions that eventually succeeded—by the much smaller count of permanently failed transactions. Root-cause distributions are reported separately for the 3167 transient retry-failure events (event-level, describing the pattern of recoverable errors encountered during retries) and for the 329 permanently failed transactions specifically (transaction-level, using each transaction's own final gateway message), because these are different populations. The vast majority of retry events (n=3160, 99.8%) occurred on transactions that did not ultimately fail (see Failure Analysis: Root Causes and Workflow Steps section).

Bootstrap 95% CIs for median cycle times (3000 resamples, percentile method) are reported given the large sample size. To test whether the association between payment channel and cycle time was confounded by calendar month, a median (quantile, τ=0.5) regression of cycle time on payment channel and calendar-month fixed effects was fitted on the within-24-hour analytic sample. To test whether the 1440-minute analytic cutoff itself affected the channel comparison, 2 sensitivity analyses were performed on the full, untruncated duration distribution (all 146,203 episodes). The primary sensitivity analysis was the simple, assumption-free comparison of raw medians without truncation (see Results). As a secondary, exploratory check, time-to-payment was modeled with a Kaplan-Meier estimator (time zero=bill creation; event=payment confirmation) and compared between channels by log-rank test; because every episode in the extracted dataset had a recorded payment confirmation, no observations were censored in this analysis. Department (laboratory, imaging, and pharmacy) and insurance status were not available as record-level fields in the extracted analytic dataset and could therefore not be entered as covariates; this is noted as a limitation. Regarding missing data, the episode-level analytic dataset (n=146,203) had zero missing values across all fields used in this analysis (bill, invoice, and payment time stamps, payment channel, and trigger status), so no imputation or missing-data test was required for the primary cycle time analysis. The transaction-level status-history audit log used for the failure-root-cause analysis was, by contrast, incomplete for terminal state transitions; this data-quality limitation is reported separately and does not affect the primary cycle time comparison, which relies on the complete episode-level dataset. All analyses were conducted in Python 3.12 (pandas, NumPy, and SciPy for descriptive and bootstrap analyses; statsmodels for quantile regression; and lifelines for the Kaplan-Meier and log-rank analyses).

Results

Study Population and Transaction Overview

The dataset comprised 146,203 invoiced episodes from October 1, 2025, to May 31, 2026, restricted to laboratory, imaging, and pharmacy billing lines with server-generated time stamps; this episode-level dataset had no missing values in any field used in this analysis (see the Statistical Analysis section). Of these, 131,037 (89.6%) were within-24-hour transactions included in the primary payment cycle time analysis. A total of 15,166 (10.4%) were classified as extended (beyond 24 hours), with a higher proportion in the cash channel (n=14,876, 11.7% of 127,268 cash episodes vs n=290, 1.5% of 18,935 Mobile Money episodes). Mobile Money payments accounted for 18,935 (12.9%) of 146,203 episodes; a further 22,073 (15.1%) episodes had Mobile Money triggered, of which 3138 (14.2% of triggered episodes) fell back to cash payment (Table 1). Monthly invoiced episode volumes ranged from 15,425 (May 2026) to 20,499 (March 2026).

XSL•FO
RenderX

Table 1. Mobile Money transaction system performance, University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), Rwanda (October 2025-May 2026). Sections A and B describe different populations, not the same population at different units of counting (see Statistical Analysis section).[a]

Metric	Transactions/episodes, n (%)	Triggered episodes, n (%)	Comment
Section A. Hospital-wide transaction-level metrics (n=44,216 unique Mobile Money transactions)[b]			
Total Mobile Money transactions	44,216 (100)	N/A[c]	October 2025 to May 2026
Completed (payment confirmed)	43,278 (97.9)	N/A	Payment confirmed by the Irembo gateway
Pending at data extraction	609 (1.4)	N/A	574 awaiting gateway confirmation after successful initiation + 35 awaiting initiation after invoice creation
Permanently failed	329 (0.7)	N/A	Invoice creation or payment initiation failed after all retries
Section B. Episode-level metrics (N=146,203 invoiced episodes)[d]			
Mobile Money triggered	22,073 (15.1)	22,073 (100)	Section A is not restricted to this prepaid cohort, and the 2 extractions do not share a persistent invoice-level identifier
Completed via Mobile Money	18,935 (12.9)	18,935 (85.8)	Episode ultimately paid via Mobile Money
Cash fallback (triggered, paid cash)	3138 (2.1)	3138 (14.2)	Mobile Money initiated but completed at the cashier
Never triggered (pure cash)	124,130 (84.9)	N/A	Traditional cashier payment
Extended beyond 24 hours (excluded from cycle time comparison)	15,166 (10.4)	N/A	14,876/127,268, 11.7% of cash episodes vs 290/18,935, 1.5% of Mobile Money episodes

[a]A transaction (section A) is 1 Mobile Money payment attempt; an episode (section B) is 1 invoiced episode in the prepaid analytic cohort. These are different, nonnested populations (see Statistical Analysis section); no ratio between section A and section B counts should be read as a within-cohort attempt or retrigger rate.

[b]Section A presents hospital-wide transaction-level outcomes across all CHUB billing categories. Percentages use the total number of unique Mobile Money transactions (n=44,216) as the denominator. This analysis is not restricted to the prepaid analytic cohort in section B.

[c]N/A: not applicable.

[d]Section B presents episode-level outcomes for the prepaid analytic cohort (laboratory, imaging, and pharmacy). The primary percentages use all invoiced episodes (n=146,203) as the denominator, whereas the secondary percentages use Mobile Money-triggered episodes (n=22,073) as the denominator. The 2 data extractions do not share a persistent invoice-level identifier and therefore cannot be directly linked.

Payment Cycle Time: Mobile Money vs Cash

Among within-24-hour transactions (n=131,037), the median payment cycle time was 5 (IQR 2-17; 95% CI 5-5) minutes for Mobile Money and 72 (IQR 30-165; 95% CI 71-72) minutes for cash, a difference of 67 (95% CI 66-67) minutes associated with payment channel (Table 2; Figure 2). This association was materially unchanged after adjustment for calendar month by quantile regression (adjusted coefficient –67.0, 95% CI –67.7 to –66.3 minutes; $P<.001$; unadjusted coefficient –67.0, 95% CI –67.7 to –66.3 minutes), indicating that calendar month was not an important confounder of the channel association. The mean cycle time was 49.8 (SD 171.9) minutes for Mobile Money vs 162.4 (SD 256.1) minutes for cash, a difference associated with a 69.3% lower mean for Mobile Money. The pronounced divergence between the Mobile Money mean (49.8, SD 171.9 minutes) and median (5, IQR 2-17 minutes) reflects strong right skew within the 24-hour analytic window. A minority of Mobile Money payments completed many hours after billing (up to the 1440-minute cutoff) and disproportionately raised the mean, whereas the median, which is robust to this skew, better represents the typical patient experience. The speed ratio (cash median divided by Mobile Money median) was 14.4, indicating that cash patients waited, at the median, 14.4 times longer to complete payment. The 90th percentile cycle time was 82 minutes for Mobile Money vs 378 minutes for cash. The IQR for cash (135 minutes) was 9.0 times wider than for Mobile Money (15 minutes), indicating substantially greater variability in the cash payment experience.

XSL•FO
RenderX

Table 2. Payment cycle time comparison, Mobile Money vs cash, retrospective cohort of prepaid (laboratory, imaging, and pharmacy) billing episodes at University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), Rwanda, October 2025-May 2026. Rows use different populations as stated in the Population column.

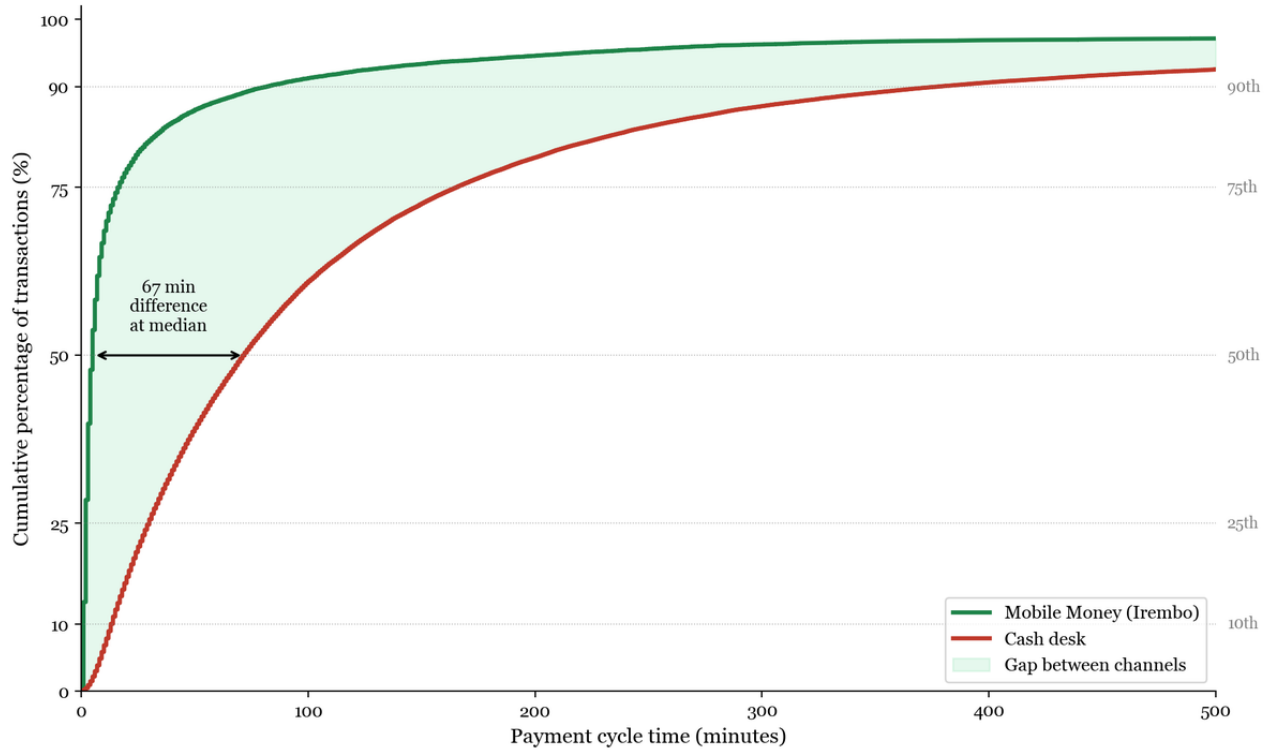
Metric	Mobile Money	Cash desk	Difference	Change
Cycle time analysis (episodes completed within 24 hours, n=131,037)				
Cycle time (minutes), median (95% CI)[a]	5 (5-5)	72 (71-72)	67 (66-67)	93.1% lower
Cycle time (minutes), mean (SD)	49.8 (171.9)	162.4 (256.1)	112.6	69.3% lower
Speed ratio (cash/Mobile Money)	N/A[b]	N/A	14.4×	N/A
Cycle time (minutes), IQR	2-17	30-165	N/A	9.0× wider (cash)
Cycle time (minutes), 90th percentile	82	378	296	78.3% lower
Additional payment cycle time metrics (all episodes, n=146,203)				
Quick payments within 30 minutes, n (%)	15,290 (80.7)	28,783 (22.6)	58.1[c]	N/A
Extended payments beyond 24 hours, n (%)	290 (1.5)	14,876 (11.7)	10.2[c]	N/A
Cycle time (untruncated sensitivity; minutes), median (IQR)	5 (2-18)	88 (34-263)	83	94.3% lower

[a]CIs for medians are bootstrap 95% CIs (3000 resamples). The untruncated-median row is the primary sensitivity analysis addressing the asymmetric 24-hour exclusion (see the Statistical Analysis section).

[b]N/A: not applicable.

[c]Percentage points.

Figure 2. Cumulative distribution of payment cycle time by payment channel, within-24-hour prepaid (laboratory, imaging, and pharmacy) billing episodes at University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), a public tertiary hospital in Huye, Rwanda, October 2025-May 2026 (n=131,037). The horizontal gap at the 50th percentile is the median difference reported in Table 2.



Because the 1440-minute analytic cutoff excluded a larger share of cash than Mobile Money episodes (n=14,876, 11.7% vs n=290, 1.5%), the primary sensitivity analysis repeated the raw median comparison on the full, untruncated duration distribution (all 146,203 episodes, no exclusion). The median was 5 (IQR 2-18) minutes for Mobile Money vs 88 (IQR 34-263) minutes for cash, a 94.3% lower median for Mobile Money—if anything slightly larger than the 93.1% figure obtained after truncation, indicating that the 24-hour cutoff does not inflate the apparent channel difference. As a secondary, exploratory check, a Kaplan-Meier estimate of time-to-payment on the full duration distribution gave the same medians and differed between

XSL•FO
RenderX

channels by log-rank test (χ^2_1=31,522; P<.001), confirming the difference nonparametrically.

The proportion of quick payments (30 minutes or less, out of all 146,203 episodes) was 80.7% (n=15,290) for Mobile Money vs 22.6% (n=28,783) for cash, a difference of 58.1 percentage points. Extended payments (beyond 24 hours) accounted for 1.5% (n=290) of Mobile Money episodes vs 11.7% (n=14,876) of cash episodes.

Monthly Evolution of Adoption and Performance

Table 3 and Figure 3 present the monthly evolution of Mobile Money adoption and cycle time metrics, using the bill (order) date to assign each episode to a calendar month, so that every percentage in Table 3 uses a single, clearly stated denominator. Completion via Mobile Money grew from 3.8% (n=724) in October 2025 to 20.9% (n=3228) in May 2026. The Mobile Money median cycle time remained consistently low (4-6 minutes) across all months. Cash desk median cycle times rose from 68 (IQR 29-162) minutes (October 2025) to a peak of 83 (IQR 34-186) minutes (April 2026), before declining to 71 (IQR 30-165) minutes in May 2026; this upward drift over most of the study period is discussed as a possible compositional effect in the Limitations section, since it is consistent with faster, less complex payers progressively migrating to Mobile Money and leaving relatively more complex cash episodes behind. Cash desk 90th percentile cycle times were highest in October 2025 (478 minutes) and ranged from 317 to 392 minutes thereafter; the Mobile Money 90th percentile rose from 41 to 100 minutes over the same period, so the cash 90th percentile remained 3.8 to 11.7 times the Mobile Money value in every month (Table 3). The proportion of Mobile Money-triggered episodes converting to cash fallback rose gradually, from 0.4% (n=73) of all episodes in October 2025 to 3.3% (n=506) in May 2026 (Table 3), consistent with a growing pool of patients experiencing an unsuccessful Mobile Money attempt (most plausibly insufficient balance; see the Failure Analysis: Root Causes and Workflow Steps section) before paying at the cashier rather than a declining share. The relatively modest absolute completion level (n=3228, 20.9% by month 8) is not primarily attributable to patient reluctance or technical barriers; it reflects the structural constraint that a large proportion of patients covered by health insurance or mutual schemes are required to physically present ink-signed diagnostic orders, prescriptions, appointment slips, and other paper documents to the cashier as a precondition of payment clearance, irrespective of the payment channel chosen. This insurer-driven paper documentation requirement is discussed further in the Discussion; its precise prevalence was not directly measured in this study (Limitations).

Table 3. Monthly evolution of Mobile Money completion and payment cycle time metrics, prepaid billing episodes at University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), Rwanda, October 2025-May 2026 (n=146,203 episodes; n=131,037 within-24-hour episodes for the median and 90th percentile columns).

Month	Total episodes[a], n	Triggered[b], n (%)	Completed via MM[c,d], n (%)	Cash fallback[e], n (%)	MM cycle time (minutes), median (IQR)	Cash cycle time (minutes), median (IQR)	MM cycle time (minutes), 90th percentile	Cash cycle time (minutes), 90th percentile
October 2025	19,189	797 (4.2)	724 (3.8)	73 (0.4)	6 (3-17)	68 (29-162)	41	478
November 2025	17,916	992 (5.5)	825 (4.6)	167 (0.9)	5 (2-12)	63 (27-145)	42	372
December 2025	17,840	2067 (11.6)	1790 (10)	277 (1.6)	5 (2-14)	68 (29-149)	64	317
January 2026	18,710	2526 (13.5)	2130 (11.4)	396 (2.1)	5 (2-18)	70 (28-164)	86	363
February 2026	17,798	3755 (21.1)	3180 (17.9)	575 (3.2)	5 (2-15)	73 (30-169)	86	373
March 2026	20,499	3965 (19.3)	3384 (16.5)	581 (2.8)	4.5 (2-18)	82 (34-183.5)	94	391
April 2026	18,826	4237 (22.5)	3674 (19.5)	563 (3)	5 (2-18)	83 (34-186)	100	392
May 2026	15,425	3734 (24.2)	3228 (20.9)	506 (3.3)	5 (2-18)	71 (30-165)	97	368

[a]All invoiced episodes billed that month (denominator for the Triggered, Completed, and Cash fallback columns).

[b]Mobile Money payment initiated; denominator=total episodes that month (includes episodes that fell back to cash).

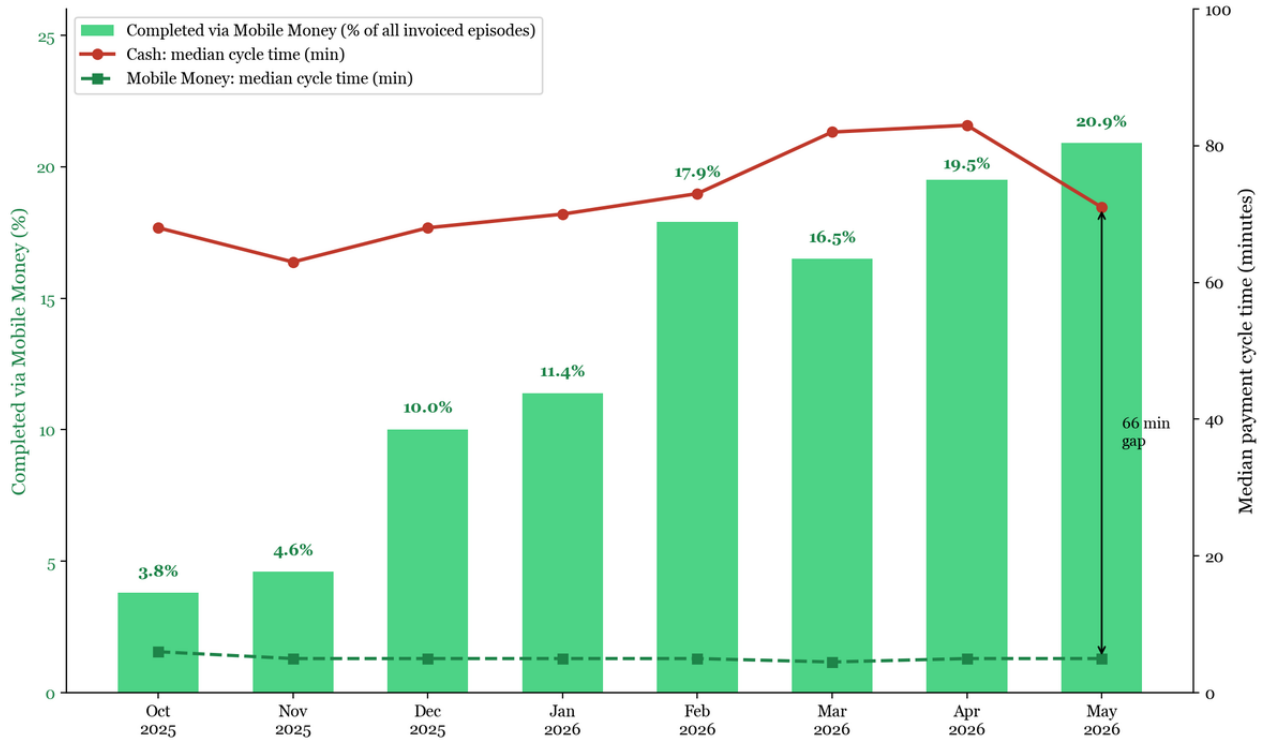
[c]Payment ultimately completed via Mobile Money, denominator=total episodes that month.

[d]MM: Mobile Money.

[e]Mobile Money triggered but completed at the cashier; denominator=total episodes that month (Triggered % = Completed % + Cash fallback %, by construction).

XSL•FO
RenderX

Figure 3. Monthly completion of payment via Mobile Money (bars, % of all invoiced episodes) and median payment cycle time by channel (lines), prepaid billing episodes at University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), Rwanda, October 2025-May 2026 (n=146,203 episodes; Table 3).

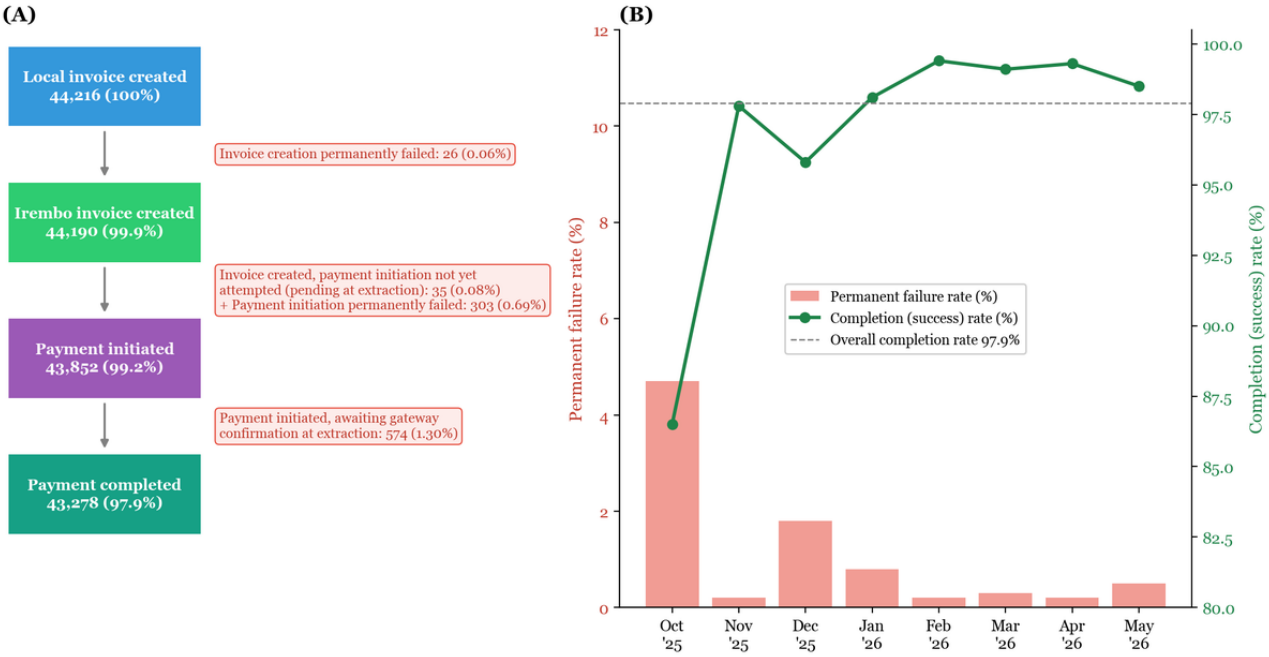


Mobile Money Transaction System Performance

The hospital-wide Mobile Money transaction log encompassed 44,216 unique transaction attempts (Table 1, section A). This log spans all CHUB billing categories in which Mobile Money payment is offered (prepaid laboratory, imaging, and pharmacy bills together with postpaid consultation, ward, and procedure fees), because it is generated by the Mobile Money integration itself rather than by the prepaid-billing extraction used for cycle time analysis; it is therefore not restricted to the 146,203-episode prepaid analytic cohort used elsewhere in this paper (see the Statistical Analysis section). The overall transaction completion rate was 97.9% (43,278/44,216); 609 (1.4%) transactions remained pending at the time of data extraction, and 329 (0.7%) failed permanently (Figure 4A). At the episode level, within the prepaid cohort specifically (Table 1, section B), of 22,073 Mobile Money-triggered episodes, 18,935 (85.8%) ultimately completed via Mobile Money and 3138 (14.2%) fell back to cash payment. Because the transaction log (Table 1, section A) and the prepaid episode cohort (Table 1, section B) are different, nonnested populations rather than the same population counted at 2 different units, a reconciled count of transaction attempts per prepaid-cohort episode cannot be computed from the available extraction. The transaction log is generated per payment invoice, and a single invoice can settle more than one prepaid bill and, in some cases, postpaid charges from the same visit (see the Statistical Analysis section); OpenClinic GA holds a linking key between bills and their settling invoice; joining it was outside this study's objectives (see the Statistical Analysis section) and was therefore not attempted, and the transaction log separately does not carry a billing-category field, so transactions cannot be restricted to the prepaid cohort after the fact.

XSL•FO
RenderX

Figure 4. Mobile Money transaction system performance, hospital-wide transaction log spanning all University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare) billing categories, October 2025-May 2026 (n=44,216 unique transactions; see the Statistical Analysis section). (A) Reconciled transaction-outcome funnel, with permanently-failed and still-pending transactions shown as branch-offs at each step. (B) Monthly completion and permanent-failure rate, by month of transaction creation.



The monthly completion rate improved from 86.5% (n=1853) in October 2025 to a sustained plateau of 95.8% (n=4840) to 99.4% (n=7240) from November 2025 onward, consistent with stabilization of the integration following initial deployment (Figure 4B). MTN Mobile Money accounted for 37,321 (84.4%) of 44,216 transactions with a 97.5% (n=36,396) success rate. Airtel Money represented only 37 (0.08%) transactions with a nominal success rate of 64.9% (n=24); given this very small sample, the Airtel-specific success rate should be treated as uninterpretable rather than a reliable estimate of Airtel channel reliability, and is reported for completeness only.

Failure Analysis: Root Causes and Workflow Steps

The status-history audit log recorded 3167 transient retry-failure events across the 44,216 transactions, of which 2833 (89.5%) occurred at step 3 (the payment initiation request) and 334 (10.5%) occurred at step 2 (the gateway invoice creation; Table 4, section A). These events should not be divided by the 329 permanently failed transactions to obtain a failure rate, because the 2 counts describe different, largely nonoverlapping populations. Of the 3167 logged events, only 7 (0.2%) occurred on transactions whose final outcome was permanent failure, while the remaining 3160 (99.8%) occurred on transactions that were subsequently retried successfully or remain pending—transient retries absorbed by the retry mechanism, not permanent failures. The cleanest, reproducible failure-rate statistics are therefore reported per attempt, with 334 (0.8%) of 44,037 step 2 attempts and 2833 (6.6%) of 43,232 step 3 attempts failed, for a combined per-attempt failure rate of 3.6% (3167/87,269); separately, 329 (0.7%) of 44,216 transactions failed permanently.

XSL•FO
RenderX

Table 4. Mobile Money transaction failure analysis, University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare), Rwanda, October 2025-May 2026. Sections A and B describe different populations (see the Statistical Analysis section).[a]

Root cause	Events/transactions, n (%)	Step	Notes
Section A. Transient retry-failure events (n=3167 events)[b]			
Insufficient patient mobile money balance	2820 (89)	Step 3 (n=2810); step 2 (n=10)	Inferred cause (not independently verified); Irembo returned a generic HTTP 500 instead of a specific error
SSL or certificate error	159 (5)	Step 2	CHUB firewall blocking outbound HTTPS to Irembo end points; resolved by firewall reconfiguration
Network or API unreachable	94 (3)	Step 2	CHUB firewall blocking HTTPS/SSL to Irembo end points; resolved by firewall reconfiguration
DNS resolution failure	73 (2.3)	Step 2	CHUB firewall blocking DNS resolution of Irembo end points; resolved by firewall reconfiguration
Invalid request or missing invoice reference	18 (0.6)	Step 3	Race condition: payment initiated before Irembo invoice creation confirmed; resolved by software patch
Timeout	3 (0.1)	Steps 2 and 3	Transient network timeout (2 at step 2 and 1 at step 3); resolved on retry
Total retry-failure events	3167 (100)	—[c]	Per-attempt failure rate 3.6% (3167/87,269 Step 2 + Step 3 attempts). Only 7 (0.2%) of these events occurred on transactions that were ultimately permanently failed (section B); the remaining 3160 (99.8%) occurred on transactions that succeeded or remain pending.
Section B. Root cause of permanently failed transactions (n=329 transactions)[d]			
Generic gateway server error (500/SERVER_ERROR)	302 (91.8)	Step 3	Most plausibly reflects insufficient balance; inferred, not independently verified. Gateway-response limitation remains unresolved
SSL/certificate handshake failure	14 (4.3)	Step 2	Verified CHUB-firewall cause; resolved
Network unreachable or DNS resolution failure	13 (4)	Steps 2 and 3	12 at step 2 (7 DNS resolution and 5 unreachable) and 1 at step 3; verified CHUB-firewall cause; resolved
Total permanently failed transactions	329 (100)	—	Only 7 (2.1%) of these 329 transactions have a matching row in section A's event log; the remaining 322 reflect underrecording of terminal state transitions in this audit trail, not an absence of cause (Statistical Analysis). No permanently failed transaction was attributable to the step 3 invalid-request race condition, which was always resolved on retry.

[a]Step 2: gateway invoice creation (44,037 attempts). Step 3: payment initiation (43,232 attempts). Section A (event level) should not be divided by section B's 329 permanently failed transactions: 99.8% (n=3160) of section A's events occurred on transactions that succeeded or remain pending (Statistical Analysis). The firewall and race-condition causes were verified and resolved; the insufficient-balance attribution remains inferential and unresolved (Discussion).

[b]Event-level analysis of transient retry failures (n=3167). Overall, 99.8% (n=3160) of events occurred on transactions that were not ultimately permanently failed.

[c]Not available.

[d]Transaction-level analysis based on each permanently failed transaction's final gateway message. This classification is independent of the event-level analysis in section A.

Because only 7 of the 329 permanently failed transactions have a matching row in the event-level audit log above, the root cause of the 329 permanent failures specifically was instead determined directly from each transaction's own final gateway-message field, which is complete for all 329 (Table 4, section B). Of these, 302 (91.8%) carried a generic step 3 server-error message, 14 (4.3%) a step 2 SSL/certificate handshake failure, and 13 (4%) a network failure (12 at step 2, of which 7 were domain name system [DNS] resolution failures, and 1 at step 3). No permanently failed transaction was attributable to the step 3 invalid-request race condition, indicating that retry fully absorbed that category during the study period. The gap between the 3167-event log and this 329-transaction accounting confirms that the status-history audit table underrecords terminal state transitions broadly, not only for permanent failures—only 1902 (4.4%) of the 43,278 completed transactions have a matching completion-event row either—a data-quality limitation of this particular audit trail that does not affect the primary outcome fields used elsewhere in this paper (the final transaction status and the episode-level cycle time dataset), both of which were complete for all records analyzed (see the Statistical Analysis section).

XSL•FO
RenderX

For the dominant step 3 category, postimplementation review found that the generic server-error message was most plausibly associated with the patient handset's mobile money balance being below the invoice amount, based on the pattern of payee messages and the timing of failures relative to retry success; this attribution is inferential and was not independently verified against Irembo-side account balance data, which was not available to the study team. This finding was documented and formally communicated to Irembo Ltd with a recommendation to implement a context-specific response code for insufficient balance; as of the end of the study period, this gateway-side limitation remained unresolved. The built-in 3-attempt retry mechanism absorbed the vast majority of these transient events, consistent with the low permanent-failure rate of 0.7% (n=329; Table 4; Figure 4).

Step 2 event-level failures (n=334, 10.5% of retry events; Table 4, section A) were dominated by CHUB router configuration, with the hospital network firewall blocking outbound traffic to the Irembo payment end points, causing SSL handshake failures, network unreachability, DNS resolution failures, and timeouts. Consistent with this, 26 of the 329 permanently failed transactions (14 SSL/certificate, 12 network-unreachable, including 7 DNS resolution failures; Table 4, section B) were step 2 failures; the remainder of step 2 retry events were transient and resolved on retry. These were not gateway-side issues but local infrastructure barriers that prevented the integration server from reaching the payment gateway, and were resolved by updating the CHUB firewall rules to explicitly permit outbound access to the required Irembo end points. The concentration of these events in the first 3 months (October-December 2025) reflects the period between initial deployment and completion of the firewall reconfiguration. A further 18 (0.6% of all retry events) failure events at step 3 were caused by a software bug in which a payment initiation request was occasionally dispatched before confirmation that the corresponding Irembo invoice had been successfully created, resulting in an invalid request; none of the 329 permanently failed transactions were attributable to this cause, indicating it was always resolved on retry before the fix. This race condition was identified, reproduced, and corrected by a targeted software patch deployed during the study period, with no recurrence thereafter.

Discussion

Principal Findings

This study addressed the objectives set out in the Introduction by describing a reproducible asynchronous Mobile Money-HMIS integration architecture and by evaluating, over 8 months of routine operation, payment cycle time by channel, the growth of Mobile Money completion, the technical reliability of the transaction system, and the structural constraints on further adoption. Mobile Money payment through the Irembo gateway was associated with a 93.1% lower median patient payment cycle time than cash (5 vs 72 minutes; 67-minute difference, 95% CI 66-67), an association that was materially unchanged after adjustment for calendar month and in a sensitivity analysis using the full, untruncated duration distribution (5 vs 88 minutes, a 94.3% lower median; log-rank $P<.001$). These gains coexisted with growing completion, from 3.8% (n=724) to 20.9% (n=3228) of all invoiced episodes over 8 months, achieved without external funding as part of the hospital's routine quality improvement program. The hospital-wide Mobile Money transaction system demonstrated high technical reliability, with monthly completion rates stabilizing above 95.8% from November 2025 onward. Because the Mobile Money payment channel was self-selected by patients rather than randomly or historically assigned, and because no direct measure of cashier queue length or a preimplementation baseline was available, these findings support an association between payment channel and cycle time under routine operating conditions, not a causal claim that the integration eliminated queues; this distinction, and its implications, are developed further below. The findings also point to a broader conceptual conclusion developed throughout this Discussion. The principal constraint on further Mobile Money adoption at CHUB is institutional (insurer paper-documentation workflow) rather than technical (payment-rail reliability), a distinction with direct implications for where implementing institutions should focus improvement effort.

Comparison With Prior Work

The magnitude of the cycle time association observed here is broadly consistent with other evaluations of mobile money–based payment in health care in the region. Franke et al [21] reported that a mobile money–based conditional cash transfer intervention improved health care utilization in Southern Madagascar, illustrating that mobile money rails can materially change how patients interact with the point of payment in resource-limited hospital settings. Zakumumpa et al [8], examining mobile money–based health insurance premium payment for people living with HIV and with hypertension or diabetes in Uganda, likewise found that mobile money channels were perceived as reducing the practical burden of health-related payment relative to cash-based alternatives, while also identifying persistent structural barriers to full adoption analogous to the insurer paper-documentation constraint identified in this study. Schuetze et al's [22] qualitative account of a digital health-financing intervention (the Insurance Management Information System) at scale in Tanzania similarly found that technical performance alone did not determine adoption, as human, organizational, and health-system-ecosystem factors were decisive, echoing our finding that the principal ceiling on Mobile Money completion at CHUB is a paper-based insurer documentation requirement rather than a technical limitation of the integration itself. More broadly, Ogundaini et al's [13] qualitative study of mHealth information and communication technology integration into hospital clinical settings across sub-Saharan Africa found that integration opportunities are shaped by the specific inefficiencies experienced by health workers at each point of care, a framing consistent with this study's emphasis on the dual trigger pathway design as a response to heterogeneous patient and staff digital literacy. Architecturally, the decoupled, asynchronous broker pattern used here parallels design choices made in other mHealth systems built for resource-constrained settings, such as the

XSL•FO
RenderX

adaptable psychoeducational mHealth platform described by Balderas-Díaz et al [18], in which a similarly decoupled communication layer separates patient-facing triggers from backend clinical logic. Xie and Cao's [2] historically controlled study of mobile payment and outpatient waiting time in China is the closest prior quantitative comparator to the present cycle time analysis; this study extends this line of evidence to a sub-Saharan African tertiary hospital, a Mobile Money rather than card-based mobile-payment rail, and an asynchronous rather than point-of-service payment model.

A reduction in median payment cycle time from 72 to 5 minutes represents a clinically relevant change in the patient journey because, at CHUB, payment clearance is a prerequisite for dispensing medications, releasing laboratory results, or proceeding with diagnostic procedures. The effect is particularly pronounced for the highest-burden patients, with the 90th percentile cash cycle time reaching 378 minutes (over 6 hours), compared with 82 minutes for Mobile Money, indicating that Mobile Money completion is associated with disproportionate benefit for patients who would otherwise face the longest waits. The 9.0-fold wider IQR for cash payments (135 vs 15 minutes for Mobile Money) is consistent with the intrinsic unpredictability of queue-based payment systems, where service time is sensitive to staffing levels, shift transitions, and demand peaks; Mobile Money payment, being asynchronous and not location-dependent, is associated with a more consistent payment experience.

Methodological Considerations

A central methodological strength of this study is the restriction of cycle time analysis to prepaid billing categories (laboratory, imaging, and pharmacy), where payment is required before service delivery. In this prepaid model, the interval between bill creation and payment confirmation directly and exclusively reflects the patient's payment waiting time, free from confounding by clinical episode duration. This distinguishes the present analysis from any approach that would include postpaid categories such as consultation fees, ward charges, or procedure fees, where services are delivered before invoices are settled and cycle times would reflect the length of care episodes rather than queuing behavior. The prepaid categories also carry server-generated time stamps recorded automatically at the point of order entry and payment confirmation, providing the temporal precision that supports reliable quantitative analysis.

The 1440-minute (24-hour) threshold used to define the primary analytic sample is operationally grounded in the configured Irembo invoice expiry window; because it excluded a larger share of cash than Mobile Money episodes, its influence on the channel comparison was tested directly via a sensitivity analysis (see the Results and Comparison With Prior Work sections) rather than assumed to be innocuous, and the channel association was not attenuated by removing the cutoff.

The 14.2% (n=3138) cash fallback rate among Mobile Money-triggered episodes merits attention. This rate remained stable across months rather than declining with system maturation. The root-cause investigation into transaction failures revealed that insufficient mobile money balance is a confirmed trigger for payment failure, with the Irembo gateway returning a nonspecific server error in this scenario. It is plausible that a subset of the cash fallback episodes also represents patients who were directed to the cash desk following a failed payment initiation resulting from insufficient balance. This population is structural rather than technical in nature, as it reflects patients who have mobile money accounts but insufficient funds at the time of the visit. Targeted investigation to characterize this group would inform complementary strategies, such as patient previsit balance awareness messaging or instalment payment pathways, to further reduce cash desk dependence.

Transaction Failure Analysis

The failure analysis reveals a technically robust integration, with a 0.7% (n=329) permanent-failure rate and a 3.6% (n=3167) per-attempt failure rate (see the Statistical Analysis and Results sections). The 2 infrastructure-level failure categories (CHUB firewall configuration; the step 3 race condition) had independently verifiable root causes and were confirmed resolved. Neither recurred after their respective fixes, and neither was attributable to any of the 329 permanently failed transactions after the firewall fix (see Results). The dominant category, the generic step 3 gateway server error, remains an inferred rather than a verified explanation, and its underlying gateway-side limitation remains unresolved, as detailed below.

The dominant failure category (n=2820, 89% of all retry events, almost all at step 3; n=302, 91.8% of the 329 permanently failed transactions specifically) was reported by the Irembo gateway as a generic server error. Postimplementation review found this pattern most plausibly associated with a payment initiation request being sent to a patient whose mobile money account balance was insufficient to cover the invoice amount, based on the pattern of payee messages and the timing of failures relative to retry success; this attribution is inferential and has not been independently verified against Irembo-side balance data, which was not available to the study team. The Irembo gateway returned the same generic server-error response in this scenario as it would for an unrelated backend fault, rather than a dedicated, semantically appropriate response code indicating insufficient balance. This is a significant gateway design limitation because it prevents the integration layer from distinguishing a genuine server failure from a patient-side payment constraint, precluding a useful, actionable message to the patient or cashier at the point of failure. The finding was documented and formally communicated to Irembo Ltd as a recommendation to implement a specific response code and descriptive message for the insufficient-balance scenario; this remains an open, unresolved limitation on the gateway side as of the end of the study period. Addressing it would enable the system to guide patients proactively, for example by suggesting they top up their mobile money account or proceed to the cash desk, rather than silently retrying a transaction that cannot succeed. The built-in 3-attempt retry mechanism absorbed most of these events without resulting in permanent failure, but a dedicated response code would improve both system efficiency and patient experience.

All step 2 failures (n=334, 10.5% of events, covering certificate errors, network unreachability, DNS failures, and timeouts)

XSL•FO
RenderX

were attributable to a single root cause: the CHUB network router was configured in a way that blocked outbound traffic to the Irembo payment gateway end points. This was a local infrastructure barrier unrelated to the gateway itself. Once identified, the firewall rules were updated to explicitly permit the required outbound connections, and step 2 failures ceased. The concentration of these events in the first 3 months of the study reflects the gap between integration deployment and completion of the network reconfiguration. This experience highlights the importance of including a dedicated network access review as a predeployment step when integrating external payment services in hospital environments where strict firewall policies are standard.

The remaining 0.6% (n=18) of failure events were caused by a software bug in which a payment initiation request was occasionally sent before the corresponding Irembo invoice had been confirmed as successfully created, resulting in a request referencing a nonexistent invoice. This race condition was identified through failure log analysis, reproduced in a test environment, corrected by a targeted software patch, and did not recur after the fix was deployed.

Design Rationale

Several architectural choices were driven by the same underlying design goal, which was to avoid any addition to the latency of the triggering clinical workflow and accommodate heterogeneity in patient and staff digital literacy. Payment initiation runs as a background broker task rather than a synchronous web service call, so a clinician entering a prescription or a nurse reviewing a patient record experiences no perceptible delay; the fixed-delay scheduling pattern prevents task execution from accumulating under load, and multiple pending invoices within a scheduler cycle are processed in parallel using Java asynchronous futures.

The 2 trigger pathways address different points on the digital-literacy spectrum. The SMS pathway requires the patient to know their OpenClinic GA patient identifier, met for established patients but not first-time attendees, while the HMIS staff pathway requires no patient action and can be initiated at the point of order entry. In practice, the staff-initiated pathway dominated during the study period, consistent with an early adoption phase.

Payment confirmation supports both webhook (near-real-time, but requiring an externally accessible HTTPS end point) and polling (more robust to network topology constraints, at the cost of up to one broker cycle of latency, 20 seconds by default) mechanisms, selectable via configuration; CHUB runs in polling mode. The explicit retrigger pathway in the HMIS user interface allows a clinician or cashier to retry a previously failed invoice, most commonly after the patient tops up an insufficient balance, without creating a duplicate invoice.

The patient-initiated pathway uses pseudo-Unstructured Supplementary Service Data (USSD; a structured SMS command) rather than a registered USSD short code, because obtaining regulatory approval for a short code was outside the project's scope at implementation; the message-parsing pipeline is agnostic to the physical channel, so migrating to a registered USSD short code in the future would require only a change to the inbound message-reception layer, with no change to invoice creation, payment initiation, or crediting.

Generalizability and Implications

CHUB's experience offers a replicable model for asynchronous mobile money payment integration in sub-Saharan African hospital settings. Two features of the implementation are particularly transferable: (1) the use of an existing open-source HMIS (OpenClinic GA) as the integration host, reducing proprietary system dependencies and (2) the framing of the integration within routine quality improvement rather than as a funded research intervention, demonstrating that impactful digital health transformation is achievable within normal operational budgets.

The dual trigger pathway design (patient-initiated SMS and clinician-initiated HMIS form) is an important transferable feature because it accommodates the heterogeneity of patient digital literacy while enabling staff to initiate payment on behalf of patients with limited familiarity with the SMS pathway, a pragmatic design that is likely to accelerate adoption in comparable settings.

Recommendations

Based on the findings of this study, 3 recommendations are made for implementing institutions. First, decouple insurer paper-document submission from the payment transaction, and pursue its progressive digitalization. This is the single highest-leverage action identified. Patients should be permitted to pay via Mobile Money at the time of service, with insurer documentation submitted to a dedicated administrative window at a separate time, reducing mandatory cashier attendance to administrative rather than financial transactions. Institutions should engage health insurers toward electronic clinician signature of diagnostic orders and prescriptions within the HMIS, a structured pilot of paperless claim submission, and eventually a direct HMIS-to-insurer interface for automated preauthorization. Until such reforms are implemented, the cashier desk will retain a role that limits Mobile Money adoption regardless of the technical quality of the payment integration; this recommendation targets the institutional, not the technical, constraint on adoption.

Second, request a specific gateway response code for insufficient mobile money balance. Mobile money gateways should be formally requested to return a dedicated, semantically unambiguous response code when a payment initiation fails due to insufficient patient balance, distinct from the generic server error currently returned. This would let the integration provide a context-specific patient-facing message (eg, prompting a top-up or redirecting to the cash desk) instead of silently retrying a transaction that cannot succeed.

Third, include a predeployment network access review in the standard implementation checklist. Hospitals deploying mobile money integrations should explicitly test outbound connectivity to all payment gateway end points from the integration server before go-live. The CHUB experience shows that standard hospital firewall policies may block required outbound traffic by default, producing failures not immediately attributable to the integration itself.

XSL•FO
RenderX

Limitations

This study has several limitations. First, the Mobile Money payment channel was self-selected by patients rather than randomly or historically assigned. Because both channels are in practice settled via Mobile Money (see Methods), channel choice does not simply track differential phone ownership or comfort with mobile technology; the more relevant distinction is whether the payer (or a delegate) triggered payment remotely, or the transaction was instead processed in person by a cashier, a choice most plausibly driven by whether the patient's bill required insurer paper documentation, which mandates cashier attendance, and by patient preference for self-service over assisted payment. Because the cashier-processed pathway definitionally requires the payer or a delegate to be physically present at the cash desk, part of the observed cycle time gap reflects this structural difference in how the 2 pathways are delivered, rather than being attributable solely to compositional self-selection; however, in the absence of direct queue-length measurement we cannot partition the gap between this structural component and residual case-mix differences, such as more complex or insured cases remaining concentrated in the cash channel, and the cash-vs-Mobile Money comparison should therefore still be interpreted as an association under routine self-selected use, not as the causal effect of offering Mobile Money to a given patient. Second, the study has no physical queue-length measurement, no cashier staffing data, and no preimplementation (cash only) baseline period, because the extracted dataset begins in October 2025, the month of go-live; the design therefore cannot separate the effect of the integration from secular trends over the study period. Consistent with a possible compositional-drift explanation, the cash median cycle time rose over most of the study period (from 68, IQR 29-162 minutes in October 2025 to a peak of 83, IQR 34-186 minutes in April 2026, before declining to 71, IQR 30-165 minutes in May 2026), a pattern compatible with faster or less complex payers progressively migrating to Mobile Money and leaving a relatively more complex case-mix in the cash channel, rather than cash service itself becoming slower. Third, department (laboratory, imaging, and pharmacy) and insurance status were not available as record-level fields in the extracted analytic dataset and could not be entered as covariates in the quantile regression; only calendar month was available for adjustment, and adjustment for month did not materially change the estimated channel association. Fourth, patient-level data on waiting experience, satisfaction, or clinical outcomes are not available, limiting the ability to quantify downstream clinical impact. Fifth, the proportion of episodes subject to mandatory cashier attendance for insurer paper-document submission was not directly measured; the adoption-ceiling interpretation is therefore inferential, as is the attribution of extended cash transactions to batch data entry and the attribution of the dominant gateway server-error category to insufficient balance. Sixth, the reported Airtel Money success rate (n=24, 64.9%) is based on only 37 transactions and should be treated as uninterpretable rather than a reliable estimate of Airtel-specific reliability. Seventh, as a single-institution study, the findings may not generalize directly to hospitals with different HMIS architectures, patient demographics, insurer mix, or mobile money infrastructure maturity. Eighth, the integration source code is not currently deposited in a public repository and is available from the corresponding author on reasonable request subject to institutional intellectual-property review; reproducibility is therefore supported through the architectural description in this paper, which is intended to be sufficiently detailed to guide an independent reimplementation, rather than through direct code reuse, and the "reproducible reference model" framing used in this paper should be read in that sense. Ninth, the patient-initiated trigger pathway currently relies on pseudo-USSD via SMS rather than a registered USSD short code, requiring payers to know the gateway phone number. Tenth, the episode-level billing dataset and the transaction-level Mobile Money gateway log were obtained as independent database extractions, and neither included the bill-to-invoice linking key held in OpenClinic GA; individual patient billing episodes therefore cannot be traced to the specific gateway transaction attempts that settled them, and the 2 datasets cannot be cross-validated at the record level (Statistical Analysis).

Conclusions

Integration of asynchronous Mobile Money payment through the Irembo gateway with the OpenClinic GA HMIS at CHUB was associated with a 93.1% lower median patient payment cycle time (72 vs 5 minutes) and growing completion from 3.8% (n=724) to 20.9% (n=3228) of all invoiced episodes over 8 months, with a sustained transaction completion rate of 95.8% or above from November 2025 onward. These associations were robust to adjustment for calendar month and to a sensitivity analysis removing the 24-hour analytic cutoff, and were achieved through a routine quality improvement initiative without external funding. Because payment channel was self-selected and no preimplementation baseline was available, these findings should be read as evidence of a strong, consistent association between asynchronous Mobile Money payment and shorter billing-to-payment cycle time under routine operating conditions, rather than as a causal demonstration of queue elimination.

The central practical finding of this study is that CHUB's adoption ceiling is institutional rather than technical. The transaction system itself is reliable (n=43,278, 97.9% completion; n=329, 0.7% permanent failure), while the requirement for patients to present ink-signed insurer documentation at the cashier desk, independently of payment channel, is the principal identified constraint on further Mobile Money completion. This reframes the priority for institutions considering similar integrations. Payment technology is necessary but not sufficient, and the higher-leverage intervention is decoupling paper documentation workflows from the payment transaction itself.

The integration architecture described here provides a reproducible reference architecture and a quantitative evidence base with explicit uncertainty estimates for asynchronous mobile money payment integration in sub-Saharan African hospital settings. Future work should pursue a controlled or prospective design capable of directly measuring cashier queue length and isolating the institutional-workflow constraint identified here.

XSL•FO
RenderX

Acknowledgments

The authors acknowledge the contributions of the University Teaching Hospital of Butare (CHUB) cashier, nursing, and registration teams for their engagement with the Mobile Money payment workflow, and the CHUB information technology department for infrastructure support. The authors also thank the Irembo Ltd technical team for their support during integration development and testing. Authors acknowledge that tasks were delegated to GenAI under full human supervision according to the GAIDeT taxonomy (2025) [23]. A section-by-section account of GenAI involvement is provided in Multimedia Appendix 3.

Funding

This study received no external funding. The integration was implemented as part of the University Teaching Hospital of Butare's routine continuous quality improvement program.

Data Availability

The datasets analyzed in this study are derived from routine hospital management information system records at University Teaching Hospital of Butare (CHUB). Deidentified aggregate data are available from the corresponding author upon reasonable request, subject to institutional data governance approval. The integration source code described in this paper is part of a customized OpenClinic GA installation maintained by CHUB; the integration source files are available from the corresponding author upon reasonable request, subject to institutional data governance and intellectual property review.

Generative AI Use

The authors declare the use of Claude (Anthropic; models Claude Opus 4.5 and Claude Sonnet 4.5, accessed via Claude Code, June-August 2026) as a generative AI (GenAI) tool in the manuscript writing and data-analysis process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI under full human supervision: data analysis coding assistance (including the Python/matplotlib code used to produce the data-derived charts), proofreading and editing, summarizing text, and reformatting. Responsibility for the final manuscript, including verification of all GenAI-assisted outputs against the underlying source code and datasets, lies entirely with the authors, and GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Authors' Contributions

Conceptualization: DM, OB, IB, VN, CN
Data curation: DM, JDG
Formal analysis: DM, JDG
Investigation: DM, AH, IB, PK, OB, JNN, LI
Methodology: DM, JDG
Project administration: OB, JNN, LI
Resources: VN, CN, TT
Software: YM, JDG, FN, LI, DM
Supervision: FM, TT, JNM, VN, CN
Validation: JDG, CK, EN, AH, IB
Visualization: DM, JDG
Writing – original draft: DM
Writing – review & editing: DM, JDG, YM, OB, JNN, FN, CK, EN, AH, PK, LI, IB, FM, TT, JNM, VN, CN

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full implementation detail, including the technology stack and configuration parameters, the 2 payment-trigger pathways, the background task scheduler, the 4-step Irembo API transaction workflow, retry and error-handling logic, and the extended OpenClinic GA data model.
[PDF File (Adobe PDF File), 1095 KB-Multimedia Appendix 1]

Multimedia Appendix 2

The completed SQUIRE 2.0 reporting checklist.
[PDF File (Adobe PDF File), 27 KB-Multimedia Appendix 2]

XSL•FO
RenderX

Multimedia Appendix 3

A section-by-section account of what was delegated to the AI tool and how output was checked before use.
[PDF File (Adobe PDF File), 5 KB-Multimedia Appendix 3]

References

1. Baballe M, Muhammad A, Abdullahi J, Abubakar M, Yau A, Giwa I, et al. The impact of hospital queue management systems. Glob J Res Med Sci. 2022;2(05):88-91. [doi: 10.5281/zenodo.7117651]
2. Xie W, Cao X. The effect of mobile payment on payment waiting time for outpatients with medical insurance: historically controlled study. JMIR Form Res. 2023;7:e43167. [FREE Full text] [doi: 10.2196/43167] [Medline: 36696970]
3. Kyarisiima I, Nzayirambaho M, Nkurunziza A, Twagirayezu I. Assessment of patient waiting time in primary health care settings in Rwanda: a mixed-method study. Rwanda J Med Health Sci. 2024;7(1):6-21. [doi: 10.4314/rjmhs.v7i1.1] [Medline: 40568310]
4. Seif S, Shah J, Chandani A, Ali SK. The perception of waiting times on patient satisfaction and patient care: a cross-sectional study at a tertiary health care institution in Kenya. PLoS One. 2025;20(5):e0322015. [FREE Full text] [doi: 10.1371/journal.pone.0322015] [Medline: 40315218]
5. Makua SR, Khunou SH. Nurse managers' views regarding patients' long waiting time at community health centers in Gauteng province, South Africa. Belitung Nurs J. 2022;8(4):325-332. [FREE Full text] [doi: 10.33546/bnj.2096] [Medline: 37546493]
6. Mugabo F, Njenga G. Strategic alliances and competitive advantage in telecommunication industry: a case of MTN Rwanda PLC. J Res Innov Implic Educ. 2026. [doi: 10.59765/xdfx438]
7. Gerard N, Njenga G. Internal core competencies and competitive advantage in telecommunication industry. A case of mobile telecommunication network Rwanda. Int J Appl Behav Sci. 2025;3(3):39-50. [doi: 10.59857/ijabs1232]
8. Zakumumpa H, Ssempala R, Kiplagat J, Kwiringira J, Quentin W, Struckmann V. Examining the potential of mobile money-based health insurance for people living with HIV and hypertension or diabetes in Uganda. Front Health Serv. 2026;6:1779532. [FREE Full text] [doi: 10.3389/frhs.2026.1779532] [Medline: 41924716]
9. Ggombe MK, Kwizeraand S, Akampumuza P. Development of the digital financial ecosystem in Rwanda: patterns, trends, and way forward. IDEAS. URL: https://ideas.repec.org/p/aer/wpaper/319299ff-f4f7-4716-a582-12b056c2bb85.html [accessed 2026-05-27]
10. Joseph ES, Chegere MJ, Mdadila K. Financial inclusion in sub-Saharan Africa: does the mobile money adoption matter? Sci Afr. 2026;32:e03282. [doi: 10.1016/j.sciaf.2026.e03282]
11. Polasik M, Górka JG, Wilczewski G, Kunkowski J, Przenajkowska K, Tetkowska N. Time efficiency of point-of-sale payment methods: empirical results for cash, cards and mobile payments. Lect Notes Bus Inf Process. 2013:320. [doi: 10.2139/ssrn.1769922]
12. Chandnani S. Leveraging asynchronous frameworks to scale payment systems: a technical analysis. Glob J Eng Technol Adv. 2025;23(2):037-044. [doi: 10.30574/gjeta.2025.23.2.0143]
13. Ogundaini OO, de la Harpe R, McLean N. Integration of mHealth information and communication technologies into the clinical settings of hospitals in sub-Saharan Africa: qualitative study. JMIR Mhealth Uhealth. 2021;9(10):e26358. [FREE Full text] [doi: 10.2196/26358] [Medline: 34643540]
14. Gupta YK, Jeswani G, Pinto O. M-commerce offline payment. SN Comput Sci. 2022;3(1):100. [FREE Full text] [doi: 10.1007/s42979-021-00978-x] [Medline: 34927084]
15. CHUB. URL: https://www.chub.rw/about/background [accessed 2026-05-27]
16. OpenClinic GA. SourceForge. 2025. URL: https://sourceforge.net/projects/open-clinic/ [accessed 2025-08-28]
17. Karara G, Verbeke F, Ndabaniwe E, Mugisho E, Nyssen M. OpenClinic GA open source hospital information system enabled universal health coverage monitoring and evaluation in Burundian hospitals. Stud Health Technol Inform. 2017;245:738-742. [Medline: 29295196]
18. Balderas-Díaz S, Rodríguez-Fórtiz MJ, Garrido JL, Bellido-González M, Guerrero-Contreras G. Design of an adaptable mHealth system supporting a psycho-educational program for pregnant women with SGA foetuses. In: Adv Concept Model. Cham. Springer International Publishing; 2021:125-135.
19. Adireddy SNK. Asynchronous processing in payments software backend system. Int Res J Eng Technol. 2024;11(02):468-477. [FREE Full text]
20. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. BMJ Qual Saf. 2016;25(12):986-992. [FREE Full text] [doi: 10.1136/bmjqs-2015-004411] [Medline: 26369893]
21. Franke MA, Neumann A, Nordmann K, Suleymanova D, Ravololohanitra OG, Emmrich JV, et al. Impact of a mobile money-based conditional cash transfer intervention on health care utilization in Southern Madagascar: mixed-methods study. JMIR Mhealth Uhealth. 2025;13:e60811. [FREE Full text] [doi: 10.2196/60811] [Medline: 40029333]
22. Schuetze L, Srivastava S, Missenye AM, Rwezaula EJ, Stoermer M, De Allegri M. Factors affecting the successful implementation of a digital intervention for health financing in a low-resource setting at scale: semistructured interview

XSL•FO
RenderX

study with health care workers and management staff. J Med Internet Res. 2023;25:e38818. [FREE Full text] [doi: 10.2196/38818] [Medline: 36607708]
23. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res. 2026;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

CHUB: University Teaching Hospital of Butare (Centre Hospitalier Universitaire de Butare)
DNS: domain name system
HMIS: hospital management information system
SQUIRE 2.0: Standards for Quality Improvement Reporting Excellence
USSD: Unstructured Supplementary Service Data

Edited by J Sarvestan; submitted 21.Jul.2026; peer-reviewed by K Ofosu-Ampong, G Guerrero-Contreras; comments to author 28.Jul.2026; revised version received 28.Oct.2026; accepted 29.Oct.2026; published 23.Nov.2026

Please cite as:
Mbarushimana D, Gatera JD, Muganga Y, Benimana O, Ntawurushimana JN, Niyomugabo F, Kayitasire C, Nsengiyumva E, Hitimana A, Kabanguka P, Ingabire L, Bagirinshuti I, Mukamana F, Twagirumugabe T, Ndoli Minega J, Nsanganiye V, Ngarambe C
Asynchronous Mobile Money Integration for Bypassing the Cash Desk for Patient Payments and Evaluating Payment Cycle Time Over 8 Months at a Rwandan Tertiary Hospital: Retrospective Observational Study
JMIR Form Res 2026;10:e107597
URL: https://formative.jmir.org/2026/1/e107597
doi: 10.2196/107597
PMID:

©Djibril Mbarushimana, Jean Damascene Gatera, Yaqub Muganga, Oswald Benimana, Jean Nepomuscene Ntawurushimana, Felix Niyomugabo, Clarisse Kayitasire, Emmanuel Nsengiyumva, Adolphe Hitimana, Prosper Kabanguka, Liliane Ingabire, Issa Bagirinshuti, Felicite Mukamana, Theogene Twagirumugabe, Jules Ndoli Minega, Vincent Nsanganiye, Christian Ngarambe. Originally published in JMIR Formative Research (https://formative.jmir.org), 23.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on https://formative.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX