JMIR Publications
The Leading eHealth Publisher

Factors influencing data quality in EHR systems in 50 health facilities in Rwanda and the role of clinical alerts: observational study

Hamish Scott Fraser, Michael Mugisha, Ian Bacher, Joseph Lune Ngenzi, Janise Richards, Xen Santas, Christopher Seebregts, Aline Umubyeye, Jeanine Condo

Submitted to: JMIR Public Health and Surveillance
on: May 19, 2023

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript **5**
Supplementary Files **25**
Figures **26**
Figure 1 ..... **27**
Figure 2 ..... **28**
Figure 3 ..... **29**
Figure 4 ..... **30**

Preprint JMIR Publications

Factors influencing data quality in EHR systems in 50 health facilities in Rwanda and the role of clinical alerts: observational study

Hamish Scott Fraser[1*] BSc, MBChB, MSc, MRCP; Michael Mugisha[2*] BSc, MPH; Ian Bacher[3] BA, PhD; Joseph Lune Ngenzi[2] BSc, MSc; Janise Richards[4] MSc, MPH, PhD; Xen Santas[5] BA; Christopher Seebregts[6, 7] BSc, PhD; Aline Umubyeye[2] MSc, MD, PhD; Jeanine Condo[8, 9] MD, MPH

[1]Brown Center for Biomedical Informatics Brown University East Greenwich, US
[2]School of Public Health, University of Rwanda Kigali RW
[3]Brown Center for Biomedical Informatics Brown University Providence US
[4]Centers for Disease Control and Prevention Dar Es Salaam TZ
[5]Centers for Disease Control and Prevention Atlanta US
[6]Jembi Health Systems Cape Town ZA
[7]University of Cape Town School of Public Health Cape Town ZA
[8]Center for Impact, Innovation and Capacity Building for Health and Nutrition (CIIC-HIN) Kigali RW
[9]Tulane University New Orleans US
[*]these authors contributed equally

Corresponding Author:
Hamish Scott Fraser BSc, MBChB, MSc, MRCP
Brown Center for Biomedical Informatics
Brown University
233 Richmond Street
East Greenwich,
US

Preprint JMIR Publications

Abstract

Background: Electronic Health Records (EHRs) play an increasingly important role in the delivery of HIV care in low-and-middle-income countries. The data collected is used for direct clinical care, quality improvement, program monitoring, public health interventions and research. Despite widespread EHR use for HIV care in many African countries, challenges remain especially in collecting high quality data.

Objective: To assess data completeness, quality and timeliness compared to paper-based records and factors influencing data quality in a large scale EHR deployment in Rwanda.

Methods: We randomly selected 50 health facilities (HF) using the OpenMRS EHR systems for HIV care in Rwanda and carried out a data quality evaluation. All HFs were part of a larger randomized controlled trial with 25 sites receiving an enhanced EHR with clinical decision support systems. Trained data collectors visited 50 HFs to collect 28 variables from both paper charts and the EHR system using the ODK app. We measured data completeness, timeliness, and the degree of matching of paper and EHR data, including use of concordance scores, for 24 required variables. Factors potentially effecting data quality were drawn from a previous survey of users in the 50 sites.

Results: We randomly selected 3,467 records, reviewing paper and EHR copies of each, (194,152 total data items). Data completeness was above the 85% threshold for all variables except viral load (VL) results, second and third line drug regimens variables. Matching scores for data values were close to or above 85%, but lower for dates particularly for drug pickups and VL. Mean data concordance was 10.2/15 (68%) for 15 variables. Site and user factors (years of EHR use, technology experience, EHR availability/uptime, intervention status) were tested for correlation with data quality measures. EHR system availability/uptime was positively correlated with concordance, and users' experiences of technology was negatively correlated. The alerts for missing viral load results implemented in 11 intervention sites showed clear evidence of improving timeliness and completeness of initially low matching of VL results in the EHR and paper records (11.9% to 26.7%, P<.001). Similar effects were seen on completeness of recording of medication pickups (18.7% to 32.6%, P<.001).

Conclusions: The EHR records in the 50 health facilities generally had high levels of completeness except for VL results. Matching results were close to or above 85% threshold for non-date variables. Higher EHR stability and uptime and alerts and

reminders for entering VL had a strong effect in improving data quality. Most data was considered fit for purpose, but more regular data quality assessments, training, and technical improvements in EHR forms, data reports and alerts are recommended. Application of quality improvement techniques described here should be beneficial across a wide range of sites and data uses for clinical care, public health and disease surveillance.

(JMIR Preprints 19/05/2023:49127)
DOI: https://doi.org/10.2196/preprints.49127

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Factors influencing data quality in EHR systems in 50 health facilities in Rwanda and the role of clinical alerts: observational study

Hamish Fraser[1]*, Michael Mugisha[2]*, Ian Bacher[1], Joseph Lune Ngenzi[2], Janise Richards[3], Xen Santas[3], Christopher Seebregts[4,5], Aline Umubyeyi[2], Jeanine Condo[6,7]

1) Brown Center for Biomedical Informatics, Brown University, 233 Richmond Street, Providence, RI, 02903, USA

Hamish_fraser@brown.edu. +1-401-863-1815

2) School of Public Health, University of Rwanda, Kigali, Rwanda

3) Center for Global Health, Centers for Disease Control and Prevention, Atlanta, USA

4) JEMBI Health systems, Cape Town, South Africa

5) University of Cape Town School of Public Health, Cape Town, South Africa

6) Center for Impact, Innovation and Capacity building for Health information and Nutrition (CIIC-HIN), Rwanda

7) Tulane University, USA

* Co-first authors.

Preprint JMIR Publications

Abstract

Background:
Electronic Health Records (EHRs) play an increasingly important role in the delivery of HIV care in low-and-middle-income countries. The data collected is used for direct clinical care, quality improvement, program monitoring, public health interventions and research. Despite widespread EHR use for HIV care in many African countries, challenges remain especially in collecting high quality data.

Objective:
To assess data completeness, quality and timeliness compared to paper-based records and factors influencing data quality in a large scale EHR deployment in Rwanda.

Methods:
We randomly selected 50 health facilities (HF) using the OpenMRS EHR systems for HIV care in Rwanda and carried out a data quality evaluation. All HFs were part of a larger randomized controlled trial with 25 sites receiving an enhanced EHR with clinical decision support systems. Trained data collectors visited 50 HFs to collect 28 variables from both paper charts and the EHR system using the ODK app. We measured data completeness, timeliness, and the degree of matching of paper and EHR data, including use of concordance scores, for 24 required variables. Factors potentially effecting data quality were drawn from a previous survey of users in the 50 sites.

Results:

We randomly selected 3,467 records, reviewing paper and EHR copies of each, (194,152 total data items). Data completeness was above the 85% threshold for all variables except viral load (VL) results, second and third line drug regimens variables. Matching scores for data values were close to or above 85%, but lower for dates particularly for drug pickups and VL. Mean data concordance was 10.2/15 (68%) for 15 variables. Site and user factors (years of EHR use, technology experience, EHR availability/uptime, intervention status) were tested for correlation with data quality measures. EHR system availability/uptime was positively correlated with concordance, and users' experiences of technology was negatively correlated. The alerts for missing viral load results implemented in 11 intervention sites showed clear evidence of improving timeliness and completeness of initially low matching of VL results in the EHR and paper records (11.9% to 26.7%, $P<.001$). Similar effects were seen on completeness of recording of medication pickups (18.7% to 32.6%, $P<.001$).

Conclusions:
The EHR records in the 50 health facilities generally had high levels of completeness except for VL results. Matching results were close to or above 85% threshold for non-date variables. Higher EHR stability and uptime and alerts and reminders for entering VL had a strong effect in improving data quality. Most data was considered fit for purpose, but more regular data quality assessments, training, and technical improvements in EHR forms, data reports and alerts are recommended. Application of quality improvement techniques described here should be beneficial across a wide range of sites and data uses for clinical care, public health and disease surveillance.

Key Words

Data quality, EHR, EMR, HIV, Rwanda

CDC Disclaimer:

This program was supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC). The findings and conclusions in this report do not necessarily represent the official position of the funding agencies.

Introduction

By 2021 over 67% of people living with HIV were from sub-Saharan Africa in 2018[1]. Rwanda along with other East African countries has made great progress in the care and treatment of patients living with HIV/AIDS including improvements in the uptake of prevention of mother-to-child transmission services, and reduction in the rate of loss to follow-up for patients receiving antiretroviral therapy (ART). In 2019, 84% of HIV+ adults in Rwanda knew their status, 98% of those knowing their status were on ART, and 90% of those on ART had suppressed viral loads[2]. HIV care has been available in Kigali (the capital) since about 2002. In 2005 partly through a collaboration between the Rwandan Ministry of Health (MOH) and the health care Non-Governmental Organization Partners In Health/Inshuti Mu Buzima (PIH/IMB) among other global partners including the Centers for Disease Controls and Prevention (CDC) and Global Fund, work started to scale up treatment to the whole population in need[3] [4]. This scale up was designed around a community-based model including routine care provision by nurses. It also included the training and deployment of community health care workers for home follow-up and support and directly observed therapy for some patients. Ensuring high quality care long term for chronic diseases like HIV also requires effective documentation of enrollment and follow up care, with reliable access to clinical data at the point of care, and for quality improvement processes and reporting[5] [6]. In 2005 (PIH/IMB) deployed a web based medical record system the HIV-EMR that had been developed for HIV care in Haiti[5]. These initial strategies were coupled with task shifting, scaling of diagnostic and treatment across Rwanda reducing barriers related to geographic and financial access to management of HIV/AIDS.

Digital Health Electronic Health Records

In the last 5 years, Rwanda has initiated digital health as one of the seven key pillars of the Smart Rwanda Master Plan, which calls for the digitalization of different processes of healthcare delivery and management. Recently, significant efforts were made to deploy information systems and numerous digital tools to support various aspects of service delivery in the health sector[2] in the aim of reaching “one citizen, one record” principle. The focus of building the national digital health enterprise architecture was developed to support seamless information exchange and need of the development of Rwanda Digital Health Interoperability Standards and Health Information Exchange (HIE), commonly known as OpenHIE.

Electronic Health Records (EHRs) are widely used for the collection, storage, viewing, analysis and sharing of clinical data. They address critical limitations in paper records that have been used for generations to support health care and are still relied on in many lower income countries. The potential benefits of EHRs over paper records include the following capabilities: unrestricted access to patients’ records for routine or emergency care often in multiple locations; storage of data securely with backups; search and retrieval of patient records and key clinical data; clinical decision support to improve quality of care; generating reports; sharing data between facilities; disease surveillance, and clinical research. The success of these activities is dependent on the design and cost of the software, the availability of suitable hardware and infrastructure particularly in low-income settings, and the ability and motivation of the staff to use the system[7, 8] [9]. Effective EHRs need to address important priorities of stakeholders (value of the system)[10] [11] and be usable by staff as part of their routine work, which requires attention to user interface, clinical workflow and training[12] [11] [13] (ease of use). Regardless of the specific system in use, one metric is critical to creating value from the system in the form of data for decision making processes: data quality[8, 14]. If an EHR implementation cannot achieve and sustain good data quality it is unlikely to create real value for stakeholders whether they are clinical staff caring for patients, health facility management, ministries of health carrying out disease surveillance, or researchers studying questions such as disease incidence or management.

What constitutes good data quality depends on the use of the data: the principle of “data which is fit for purpose”. Clinical care requires data that is as complete and accurate as possible, but often specific variables such as medications, past medical history or laboratory results are most important[15]. Missing data erodes the value of the EHR and trust in the system. That typically leads to staff questioning the value of using the systems and keeping data up to date which can lead to a downward cycle of deterioration, and sometimes abandoning the EHR[9]. Conversely a well-maintained system that is valuable for clinical tasks and reporting can incentivize better data entry. Many strategies have been used to improve data entry including mapping the workflow of the system to the typical task order of staff; providing warnings of inappropriate values for a form field; and reporting values inconsistent with previous entries - such as body weight or age. Training staff in system use and the importance of complete and up-to-date data is also a priority[16]. One essential strategy is monitoring data entry and completeness, and auditing data quality. As noted above data quality and completeness are typically linked. A fall off in data entry for example may be due to technical issue or lack of staff but can lead to gaps in the records and potentially falls in data quality. Conversely catching lack of data entry quickly can identify correctable technical issues and allow early intervention to maintain momentum and confidence[17].

Many of these issues are more challenging in Low- and Middle-Income Countries (LMICs). Additionally, the loss of reliable EHR data may cause a greater impact on patient care and health system performance in LMICs given the frequent lack of alternative records or strong health systems. Previous studies have investigated the benefits of different data entry and quality strategies, often in a small number of sites[6] [18]. However, many of these questions are most relevant in EHR systems deployed at scale and in longer term use rather than pilot projects and recent deployments. There are important unanswered questions on what strategies are most effective at improving data quality, and what factors related to health facilities and staff can facilitate or impede data quality[14]. Hedt et al studied the quality of data in paper HIV patient records in Malawi using a technique termed Lot Quality Assurance Sampling[19]. They showed high levels of completeness and

accuracy of these records, 100% of which were reviewed by a supervising team quarterly. They then studied how accurately overall data quality could be predicted from a random sample of 76 records in each of the 19 sites. That sample accurately predicted overall data quality when a threshold of 85% on quality metrics was used to indicate low quality.

EHR data use for clinical and public health.

Routine clinical data such as that collected in EHRs is an important component of public health systems in high income countries[20]. Moreover, the data in EHRs can help to understand the patterns of HIV, the progress of achieving national and global targets, oversight of the risks and patterns in specific population groups, and evolution of the infection over different circumstances. [21, 22] EHR data shared across health facilities may help to realize continuity of care as people move from one region to another. LMICs are increasingly adopting EHRs, with HIV care being in the vanguard in Eastern and Southern Africa, providing a test case for broader EHR adoption and use. The quality and timeliness of that data are a key factor in public health decision making and management of programs to improve population health outcomes. EHR data is used for public health surveillance, patient to provider communication, quality of care monitoring, linkage of primary care and public health system such as notifiable disease reporting [23], and strategic policy direction. [24]. Health Information Exchanges drawing from EHR and laboratory data for example have a growing role in population health and disease surveillance[25]. To maximized the benefits of EHRs for clinical and public health functions, they increasingly allow collection of data on behavioral, social, and environmental factors that may affect patient and population health[26]. EHR systems in LMICs such as OpenMRS are able to interoperate with pharmacy information systems, laboratory information systems and mHealth applications including through the HL7 FHIR standard, creating more comprehensive views of patient care, population and public health across a range of communities[27].

The OpenMRS EHR.

Recognizing the need for a general purpose HER system, PIH in collaboration with the AMPATH project in Eldoret, Kenya, the Regenstrief institute in Indiana, USA and the South African Medical Research Council deployed a new open source, general purpose EHR for HIV care in 2006[28]. This system, the OpenMRS EHR was subsequently scaled to 43 PIH/IMB supported health facilities in Rwanda. From 2011 onwards it was rolled out to over 300 MOH run health facilities. OpenMRS now supports comprehensive services in all departments of district hospitals. It also supports care for maternal-child health, non-communicable diseases, and oncology in health centers[7]. OpenMRS is used in more than 44 LMICs[7] including support for HIV care in more than 5000 health facilities primarily in Kenya, Uganda, Mozambique, Nigeria and Rwanda, and will soon be deployed throughout Haiti[29]. OpenMRS continues to be upgraded, with recent improvements in the user interface, data analysis and export, standardized clinical decision support systems and improved interoperability using HL7 FHIR[7, 27]. Previous studies of data quality in OpenMRS have been carried out[6] but only one published study that evaluated data quality at scale in more than 50 health facilities[14].

The Rwanda OpenMRS evaluation.

The study here was part of a larger process evaluation and randomized controlled trial of OpenMRS in Rwanda. The US Centers for Disease Control and Prevention (CDC) funded a process evaluation of the use of the OpenMRS EHR at scale in Rwanda. This included (1) a user survey, (2) key informant interviews, (3) monitoring of system use and technical stability, (4) a study of the costs of development and deployment of the enhanced EHR software, and (5) the data quality study described here.
There is also a cluster randomized controlled trial (RCT) of the impact of improving workflow and adding decision support tools on quality of HIV care currently being completed. Three types of alerts were implemented in the trial:

1. a warning if a patient newly diagnosed with HIV was not started on ART within 2 weeks (in the same

clinic).
2. a warning for a missing viral load result for patients (a) after 6 months of HIV care, and (b) annually there-after (with a 2-month window for test results to be returned and entered into the EHR)
3. a warning for abnormal viral load results (>1000) suggesting virologic failure.

112 sites were included in the RCT, half (56) were randomized to receive alert #1, 28 of those sites also received alert #2, and 14 also received alert #3. The alerts were triggered when clinicians opened the patient summary and were also presented as a report of all patients currently matching the rule. The end users were trained in interpreting the alerts in the patient summaries and reports. The larger study design has been described previously[4].

The aim of this study was to measure data completeness, matching, concordance, and timeliness, and to evaluate factors influencing data quality and strategies to improve it. The data studied was sourced from OpenMRS, a widely deployed EHR system supporting HIV care in Rwanda for more than 15 years [30] [7]. It builds on a clinical trial of the effects of clinical decision support tools on HIV care. It also includes data on the characteristics of the health facilities and the staff who work there, building on the results of an EHR user survey of the same group of facilities [4].

Methods

We describe here a cross sectional study based on the Lot Quality Assurance Sampling approach used by Hedt et al[19]. Fifty Health Facilities (HF) were selected from the 112 sites included in the clinical trial, chosen based on the power calculations for that study, selecting a random sample from both intervention and control sites and a representative mix of larger and smaller HFs but not hospitals. Trained research assistants from the Rwanda School of Public Health visited the 50 HFs and collected data using ODK, an open source mobile data collection software[31]. At each HF, depending on the patient volume, between 48-76 charts were randomly selected from the set of patients receiving HIV care based on Hedt et al[19]. Data was extracted from the paper records and the EHR records for each patient on 28 key data items on patient enrollment, ARV treatment, viral load, medication pickup, and retention in care. Data was entered into ODK from both sources and dates for each variable were also recorded. The variables collected are shown in table 1.

Deidentified data was uploaded to a study database housed at the University of Rwanda's School of Public Health. Data analyses were partly based on work of Hedt Gautier in Malawi[19] and Rwanda on public health reporting systems, and Muthee et al[14] on data quality analysis in the KenyaEMR OpenMRS based system.

Analyses included the following metrics:
- Comparison of the paper and EHR records for completeness of the data items
- Comparison of the paper and EHR records for percentage of matching of the data items
- Concordance scores comparing combined matches of 15 key variables.
- Comparison of HFs that were recorded as having high EHR uptime and availability (in the previous user survey[4]) to those with poor uptime.
- Comparison of HFs based on duration staff have worked with the EMR and their level of technology experience.
- Comparison of completeness of key variables in the EHR records in intervention HFs with control HFs, *for records completed after the intervention was implemented.*

Data processing and analysis

Firstly, the data were processed to identify any missing values. Values were regarded as missing if the field was blank, was coded with an answer indicating that the information was not available, and, for date variables, if the date reported was impossible (i.e., either too far in the future or dates related to HIV treatment that pre-date any HIV treatment occurring in Rwanda). This latter category helped address the fact that "missing"

dates were often recorded as 01-01-1980. For all 28 variables in the data set shown in table 1, we then counted the number of records that did not have missing data in either the paper file or the EMR and recorded that as the number filled or complete. This value was divided by the total number of records to give us the percentage complete. *Data were scored as "High Quality" if >= 85% of the records were complete* based on the LQAS protocol[19].

From the subset of records that were complete, we then analyzed how many of them matched in both the paper file and the EMR. Before we determined the number of matches, we carried out basic data standardization. Specifically, drug regimens were sometimes reported as multiple combinations (for example the medications could have been coded TENOFOVIR (300) + LAMIVUDINE (300) and EFEVIRENZ (600) instead of TENOFOVIR (300) + LAMIVUDINE (300) + EFEVIRENZ (600); for these cases, we standardized the drug list to one standard code. For these 22 data points, data were regarded as a match if the standardized value recorded in the paper file exactly matched the standardized value recorded in the EMR. Data quality for matches was regarded as "High Quality" if 85% of the records that were complete matched in both the paper file and the EMR.

For each patient, *we collected the last three viral loads recorded if available*. As seen in Table 2, there were few "matching" records for viral loads. Further analysis showed that this usually occurred when the paper file had one newer viral load than the EMR, so that the most recent viral load in the EMR exactly matched the second most recent viral load in the paper file, and so on. To account for this, we aligned the records based on date, so for example if the EMR had a record for 01-05-2019 this would be treated as being the same entry as a record from the paper file that was also dated 01-05-2019. Records with a date in the paper file that did not match the date of a recorded viral load in the EMR were counted as missing in the EMR and vice-versa. Because we only had the last 3 records from each file, in any case where the paper record had newer data than the EMR, there were always at least two missing records. To account for this, we aligned records in the EMR and paper file based on date and produced two derived variables, "All Viral Loads", which summarizes all viral load records, but counts matches based on date rather than position, and "At Least 1 Viral Load", which looks at whether the most recent value in the EMR matches one of the three values in the paper file. The data completeness for filled viral loads was regarded as "High Quality" if >= 85% of the records existed in both files.

After the date alignment was completed, we standardized values such that VL numbers which had a small difference within 1 copy / mL were counted as a match. This helped address cases where viral loads were reported with additional precision (20.6 vs 20 or 19.2 vs 20) noting that the EMR system only recorded absolute numbers (i.e 20 instead of 20.6). *For all filled viral loads (viral loads where date matched in both the paper file and the EMR), we counted them as a match if the normalized viral loads in both records match*. Viral load matches were regarded as "High Quality" if 85% of the records matched in both the paper file and the EMR.

We then performed the same analysis for the *data broken down by individual health care facilities* allowing analysis of potential factors affecting data quality. First, we looked for correlations between characteristics of the health facility and resulting data quality. Facility characteristics were taken from the user survey collected at the same time as this data quality study [4]. We hypothesized that relevant factors would be: (1) was the health facility an intervention or control site, (2) the average time (years) that the clinical users and data managers had been working with the EMR and (3) the relative technology experience of the users and (4) how available and stable the EMR was. We hypothesized that being an intervention site, length of time working with the EHR, EHR uptime, and technical experience would be positively correlated with better data quality.

To measure overall data quality, we developed a *concordance score* based on the work of Muthee et al[14]. We narrowed the number of variables down to 15 by considering only the most recent viral load result and drug pickup result, and by dropping the variables for health centre sector, district, and facility, since these were not patient-specific data (table 1). The dropped variables were either highly correlated or had little data entered in either paper or EHR records. The concordance score was calculated per record as the number of variables that exactly matched in both the paper file and the EMR after the values had been normalized.

Following the approach of Muthee et al, *this differed from our above analysis in that if the data were missing in both the paper file and the EMR, the records were counted as concordant.* The concordance score is therefore a number from 0-15 that is just the count of the number of concordant variables in a single record.

The randomization list of control and intervention arm sites was used to label sites in this study as intervention or control. For number of years' experience with the EMR we used the arithmetic mean of the number of years the clinical users and data managers at each site had been using the EMR (if there was more than 1 individual). Technology experience was determined based on answers to the survey about "outside of work" use of technology, which included how frequently they: (1) sent text messages; (2) used the internet on a mobile phone, (3) used a computer or (4) accessed the internet. For each question, they received one point for each time they said they used the technology either "most of the time" or "always", resulting in a score from 0-4.

Statistical analysis

Study metrics are summarized in table 1. We calculated the correlation between the concordance score and each site characteristic using the Pearson correlation coefficient with a Bonferroni correction applied to adjust for the multiple comparisons. Next, we sought evidence for the *hypothesis that: being an intervention site in the Randomized, Controlled Trial (RCT) would have a positive effect on the recording of viral loads or drug dispensing episodes in the EHR*. Viral loads and drug regimens were normalized as described above (due to different ways of recording the same data) and then counted as a match when the data were not missing and matched exactly. We could only compare data from after the intervention had been rolled out in July 2018. We then evaluated whether there was any relation between the variables using a simple Chi-squared test to determine whether or not there was an effect and Pearson's contingency coefficient to estimate the effect size. (For simple data like a 2x2 contingency square, Pearson's contingency is equivalent to both the phi coefficient and Cramers V). Data analyses were carried out using R.

Ethical considerations

The study was approved by three Institutional Review Boards: Rwanda National Ethics Committee, Kigali: #913/RNEC/2016, and the University of Leeds School of Medicine Research Ethics Committee, Leeds, UK: MREC16-176. This study was reviewed in accordance with the U.S. Centers for Disease Control and Prevention (CDC) human research protection procedures (approval #CGH HSR 2014-270a) and determined to be research. However, CDC investigators did not interact with human subjects or have access to identifiable data during this research.

Table 1: Metrics defined in the study

Metric	**Definition**
Data completeness	Number of fields filled divided by all fields.
Data matching	Number of fields in the EHR record that match the paper records divided by the number of filled fields in the paper records.
Concordance scores	Number of the 15 selected variables where the EHR and paper records match in each medical record (including empty fields in both record types).
Impact of site factors on concordance scores	Comparison of concordance scores in sites with and without the site characteristic (such as high EHR uptime).
** Data matches and concordance scores were based on data that had been normalized as described in Methods.*	

Results

A total of 3,467 records were reviewed in both the paper charts and the EHR in the 50 health facilities selected, with a mean of 69.3 records reviewed per site. A total of 194,152 data items were collected in the ODK based study tool. Three metrics of data quality were applied: (1) data completeness of EHR and paper records, (2) matching of the values of individual data types between EHR and paper, and (3) concordance scores of matches for 15 regularly collected data items between EHR and paper.

Table 2 shows the full 28 variables collected from the paper records and the EHR. It also shows the percentage completeness of each category. Few patients were on second line drugs and possibly one on third line (shown in EHR only). Importantly paper VL records were 20 – 30% more complete than EHR records.

	Variable name	Included in concordance score Yes (Y)/No (N)	Completeness in paper records N (%)	Completeness in EMR records N (%)
1	Date of Birth	Y	3465 (99.9%)	3466 (100%)
2	Gender	Y	3441 (99.3%)	3467 (100%)
3	Health Centre District	N	3467 (100%)	3467 (100%)
4	Health Centre Sector	N	3467 (100%)	3467 (100%)
5	Health Centre Name	N	3467 (100%)	3467 (100%)
6	Date of Last Visit	Y	3464 (99.9%)	3465 (99.9%)
7	Date of First HIV+ Test	Y	2711 (78.2%)	2273 (65.6%)
8	Date Enrolled in HIV Care	Y	3456 (99.7%)	3457 (99.7%)
9	Date First Prescribed ARTs	Y	3463 (99.9%)	3383 (97.6%)
10	Initial WHO Stage	Y	3433 (99%)	3267 (94.2%)
11	Regimen 1 Start Date	Y	3404 (98.2%)	3385 (97.6%)
12	Regimen 1 Drugs	Y	3411 (98.4%)	3311 (95.5%)
13	Regimen 2 Start Date*	Y	135 (3.9%)	86 (2.5%)
14	Regimen 2 Drugs*	Y	136 (3.9%)	86 (2.5%)
15	Regimen 3 Start Date*	N	0 (0%)	1 (0%)
16	Regimen 3 Drugs*	N	0 (0%)	1 (0%)
17	Drug Pickup 1 Date	Y	3466 (100%)	3465 (99.9%)
18	Drug Pickup 1 Drugs	Y	3466 (100%)	3312 (95.5%)
19	Drug Pickup 2 Date	N	3433 (99%)	3385 (97.6%)
20	Drug Pickup 2 Drugs	N	3434 (99%)	3233 (93.3%)
21	Drug Pickup 3 Date	N	3395 (97.9%)	3267 (94.2%)
22	Drug Pickup 3 Drugs	N	3397 (98%)	3123 (90.1%)
23	Viral Load 1 Date	Y	3280 (94.6%)	2645 (76.3%)
24	Viral Load 1 Value	Y	3284 (94.7%)	2649 (76.4%)
25	Viral Load 2 Date	N	2951 (85.1%)	1773 (51.1%)
26	Viral Load 2 Value	N	2955 (85.2%)	1775 (51.2%)

27	**Viral Load 3 Date**	N	2382 (68.7%)	1116 (32.2%)
28	**Viral Load 3 Value**	N	2386 (68.8%)	1118 (32.2%)

Table 2: Completeness of all 28 variables in paper records and EHR, and identification of variables used in the concordance score. N = 3467. All variables met the 85% threshold except "date of first HIV test", VL data, and second- and third-line drug regimen data.
* Very few participants had the records for these variables.

Table 3 shows the matching of variables between paper and the EHR. Most non-date variables had fairly high levels of matching, although 5 were between 82% and 85% just below the threshold set for high quality. Similar to data completeness results, viral load values had lower matching rates (59.5% for most recent value, 65.8% for the oldest). Few patients were prescribed second or third line drug regimens (and matching was low).

Additional analysis of date matches showed that for "one off" dates such as "date enrolled in HIV care" matches were fairly high, (60.2 – 75%) compared to repeated dates such as "date of last visit" or drug pickup date. For the "one off" dates most errors appeared to be mistakes in transcription (small date differences and/or single character errors). Detailed analysis of the degree of error in these date fields showed that many incorrect one-off dates such as "date enrolled in HIV care" were within one month of the correct value 34% - 38% of the time. Conversely, for regularly updated dates only 10-13% of errors were correct within one-month or 29-34% within three-months, suggesting that those items were not updated in the EHR frequently enough.

Concordance scores

The mean overall concordance score was 10.2/15 (SD 1.28, 68%) for all 50 sites as shown in figure 1. Only 2 sites were at or above the threshold of 85% concordance. However 8 of the 15 variables in the concordance score (not including DOB) were dates (see table 1 for included variables). The concordance score would be higher if dates that match within one month were included.

	# Complete in Paper and EMR	# Matches	% Matches
Date of Birth	3464	3242	93.6%
Gender	3441	3326	96.7%
District	3467	3452	99.6%
Sector	3467	3330	96.0%
Health Centre Name	3467	3467	100.0%
Date of Last Visit	3462	979	28.3%
Date of First HIV+ Test	1830	1102	60.2%
Date Enrolled in HIV Care	3446	2583	75.0%
Date Prescribed ARTs	3379	2443	72.3%
Initial WHO Stage	3236	2740	84.7%
Regimen 1 Start Date	3322	2382	71.7%
Regimen 1 Drugs	3256	2673	82.1%
Regimen 2 Start Date	78	42	53.8%
Regimen 2 Drugs	78	37	47.4%
Regimen 3 Start Date	0	0	0.0%
Regimen 3 Drugs	0	0	0.0%
Drug Pickup 1 Date	3464	986	28.5%
Drug Pickup 1 Drugs	3311	2772	83.7%

Drug Pickup 2 Date	3363	698	20.8%
Drug Pickup 2 Drugs	3212	2697	84.0%
Drug Pickup 3 Date	3220	541	16.8%
Drug Pickup 3 Drugs	3078	2583	83.9%
Viral Load 1 Date	2614	847	32.4%
Viral Load 1 Value	2622	1559	59.5%
Viral Load 2 Date	1719	547	31.8%
Viral Load 2 Value	1725	1095	63.5%
Viral Load 3 Date	1033	322	31.2%
Viral Load 3 Value	1038	683	65.8%

Table 3: matching each of the variables between paper and EHR. Several values were just below the 85% threshold. The only matching rate below 59% that was not a date was for Regimen 2 drugs which had very little data (n=78).

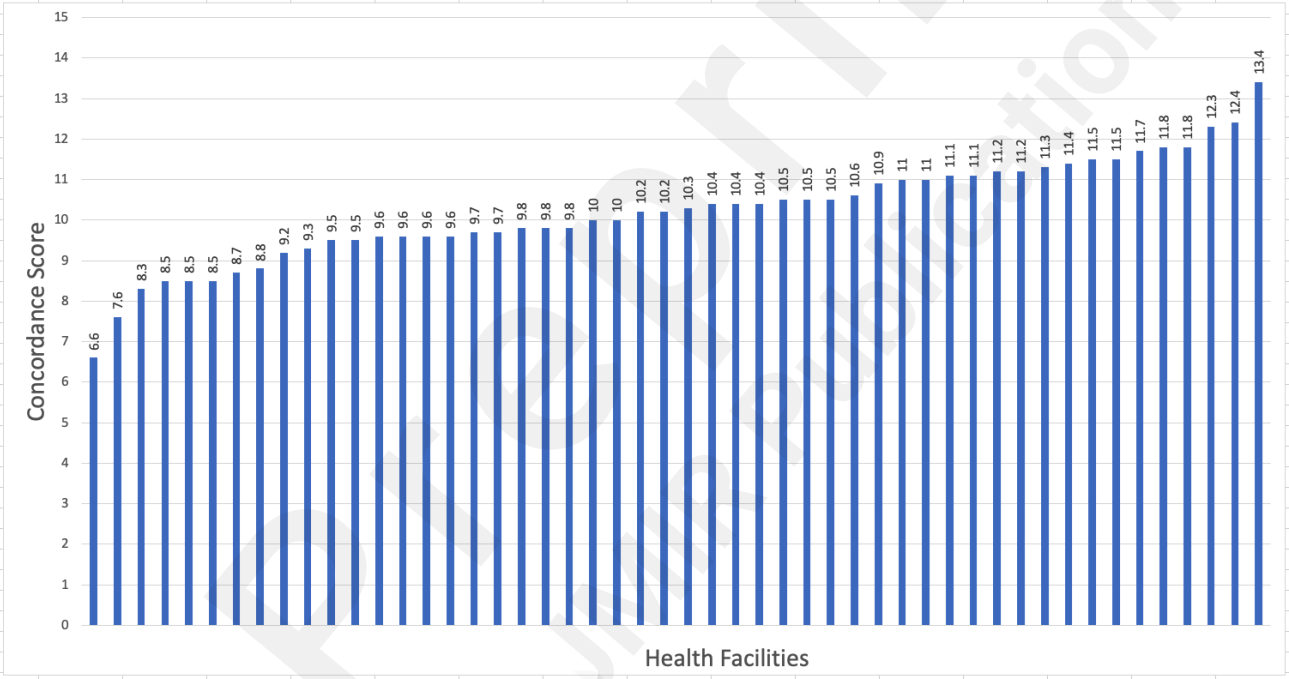


Figure 1 shows the mean concordance scores for all sites (maximum score 15).

Evaluation of factors associated with concordance scores in each site

Correlations between the concordance scores and most site characteristic were not significant. For the survey response that "the EHR was available always or nearly always" there was a positive correlation with *concordance score*, with r of 0.16, see figure 2. There was also a significant (negative) relationship between user responses suggesting they had higher *technology experience* and *concordance* score with r = -0.11. There was no significant correlation with users reported years of EHR use. See table 4.

Parameter-1	**Parameter-2**	**r**	**95% CI**	**t(3326)**	**p**
Concordance	EMR_available	0.16	[0.13, 0.19]	9.37	< .001***
Concordance	Technology experience	-0.11	[-0.15, -0.08]	-6.64	< .001***
Concordance	Years_EHR_use	-0.04	[-0.07, 0.00]	-2.10	=0.213

Table 4: Association of EMR performance and use characteristics with concordance scores

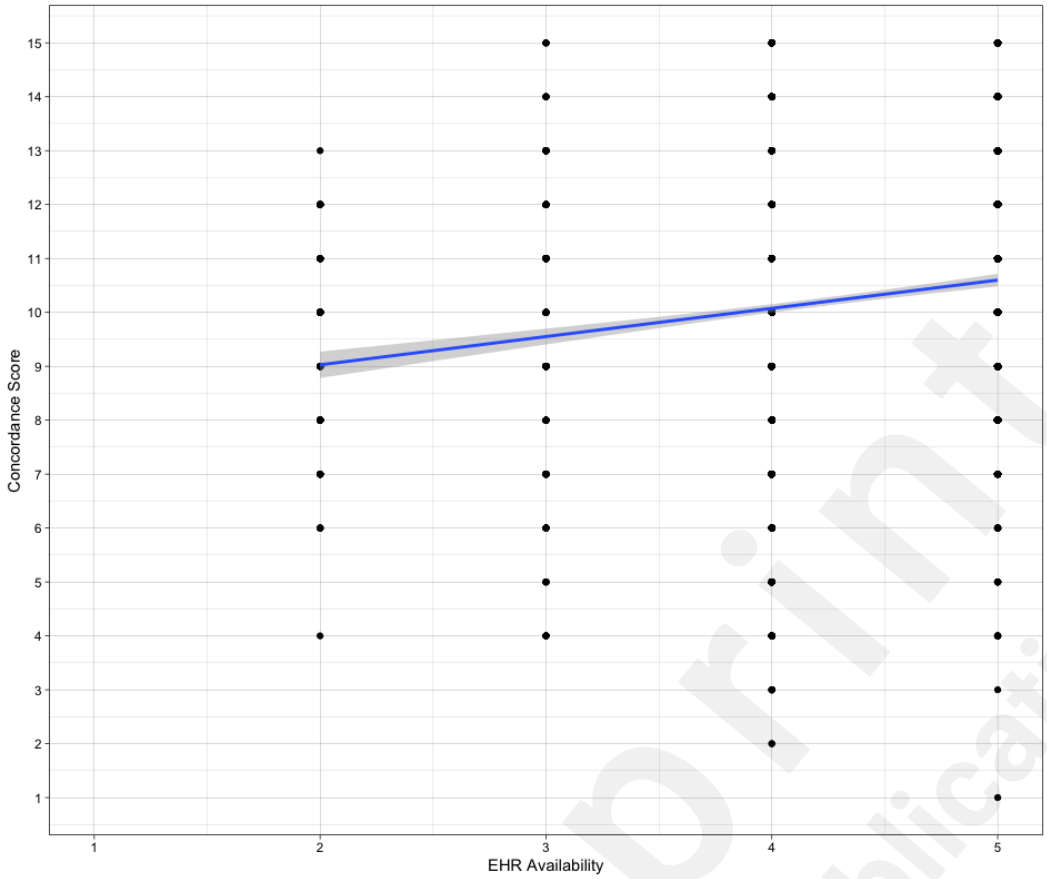


Figure 2: correlation between concordance scores, and the reported EHR availability or uptime (from the 1 – 5 Likert scale responses of EHR users).

Effects of alerts in the intervention sites on matching of key variables in EHR and paper

As the intervention was implemented in July 2018 and the data collection for the DQ study took place in November and December 2018, there was a 4–5-month window when the alerts were active. All intervention sites had alerts for *starting new treatment*, but few *new patients* were enrolled in the 5-month time frame. Therefore, there was insufficient data to study the effects of alert 1. However, there was sufficient data to compare data quality metrics in the 11 intervention sites which had *alerts for missing VL results* (alert 2) with the 39 sites without these alerts. Figure 3 shows the effect of the VL alerts on overall data matches between the EHR and paper records, ranked by percentage matching of variables between the two sources. The sites with the VL alerts (orange color) have a higher rank overall than the control sites.

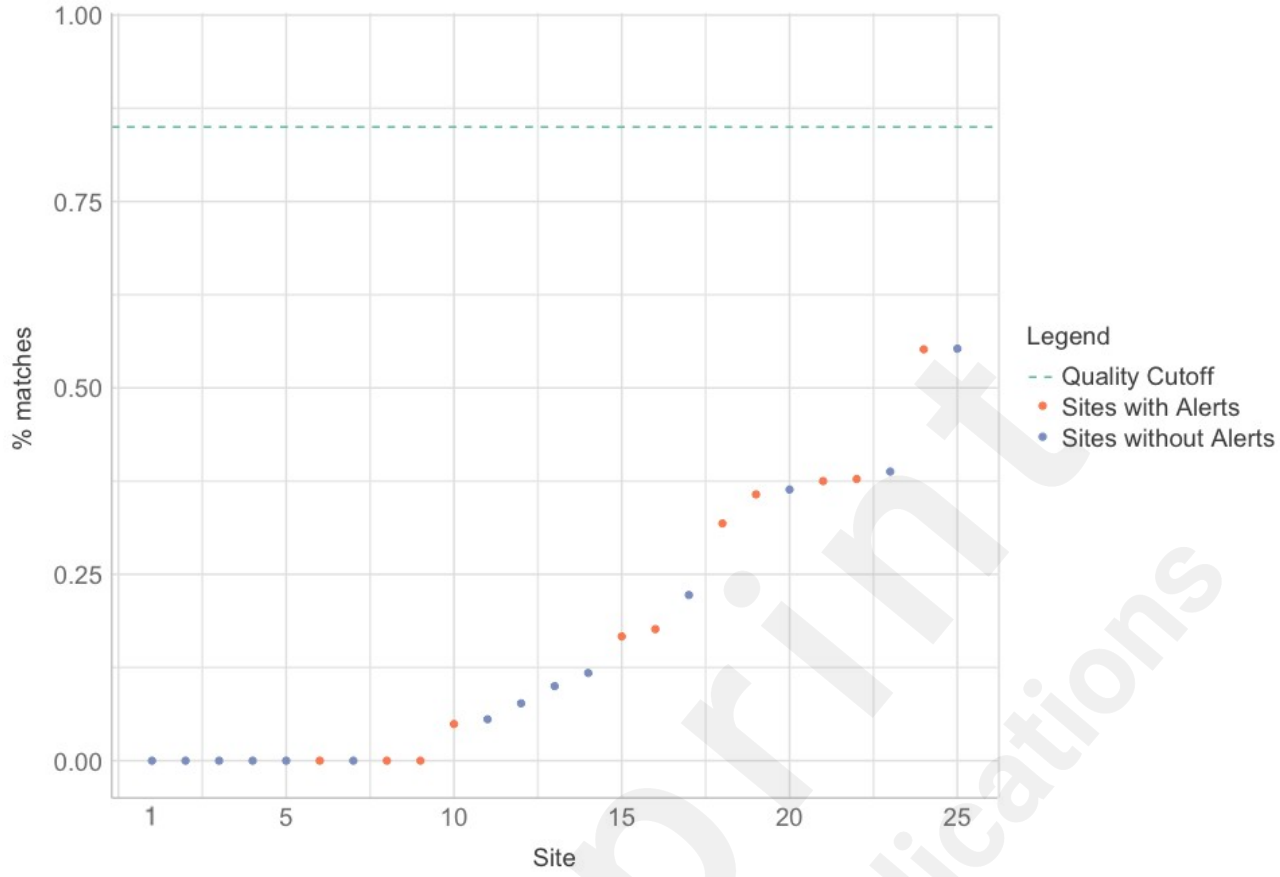


Figure 3: The graph shows percentage data matches for VL results between paper records and EHR for each of the 25 health care centers in the intervention arm, ranked by percentage of matches.

Table 5 shows the comparison of intervention sites with VL alerts to sites without such alerts. In sites with alerts 26.7% of records of VL in the EMR matched the paper records, but in sites without alerts only 11.9% matched (*P*<.001, Chi square)

Viral loads	Match	Non-match
Alerts	85 (26.7%)	233
No alerts	120 (11.9%)	888

Table 5: Number of VL results in the EHR that match the latest VL in the paper records, comparing the 11 sites that have viral load alerts versus the 39 sites without these alerts.

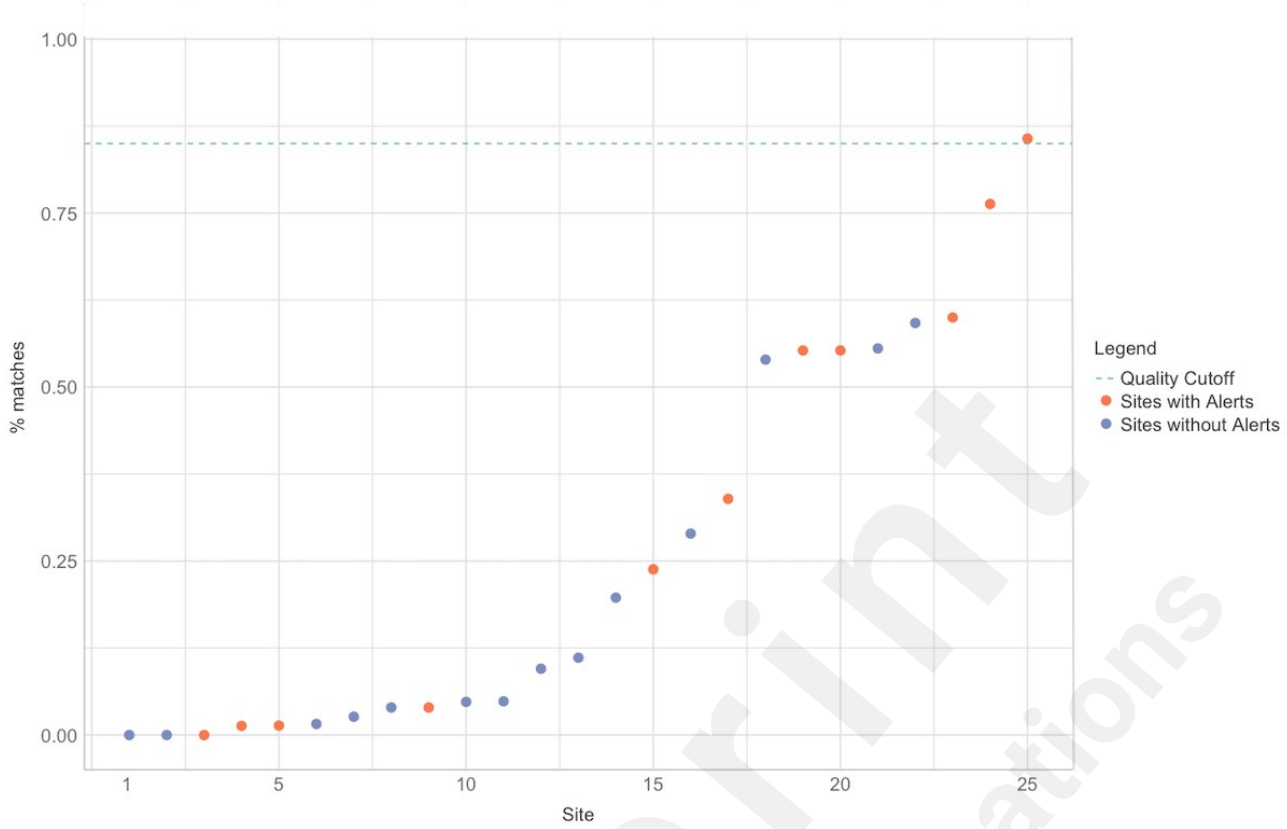


Figure 4: Graph showing percentage data matches for drug pick-ups between paper and EHR records among the same 25 intervention sites as figure 4. Scores were higher overall than in Figure 3 but only one (intervention) site was over the data quality threshold.

We also *compared the number of drug pickup records in the EHR that exactly matched the paper file*, again comparing the intervention sites with VL alerts versus sites without such alerts. Table 6 shows that in sites with alerts 36.2% of records of drug pickups matched but in sites without alerts only 18.7% matched (*P*<.001). In a second analysis using Chi squared and Pearson's contingency coefficient in the 25 sites, we show (X^2 = 67.5, *P*<.001) indicating a strong effect (ϕ = 0.20, 95% CI [0.15, 0.24]). Despite the significant trend for higher data quality in the intervention sites, not all of them showed this effect.

	Match	Non-match
Alerts	280 (32.6%)	494
No alerts	183 (18.7%)	798

Table 6: impact of the presence of VL alerts on recording of drug pickups in the EHR for the 25 intervention sites, comparing the 11 sites that have viral load alerts versus the 39 sites without these alerts.

Discussion

The results showed that the completeness of the paper records and EHR records were generally well above the 85% threshold for good quality, with exception of viral load data collected in the two year prior to the study data collection (VL results 2 and 3). Completeness varied across the data items with the best results (>99%) for EHR data seen in demographic data, date of enrollment in ARV care, date of being prescribed ARVs, and date of last visit. Lowest completeness was seen in in second- and third-line drug regimens, but very few patients were prescribed these regimens. There was clear evidence that the electronic records were not being updated in a complete and timely manner for some data items in most sites (based on concordance scores) with only 2 sites meeting the 85% threshold for good data quality. There was also evidence that certain sites were more effective in the data entry and management. Using data from the previously published user experience survey[4] allowed comparison of measures of user experience with the EHR in the same HFs as

were included in this study. Facilities reported to have more consistent uptime and availability of the EHR showed significantly better data quality based on the concordance scores. However, sites with higher "technology experience" scores had lower concordance scores, the reason for which is unclear.

A key scientific question for the overarching study was to determine whether the enhanced EHR in the intervention sites improved both the usability and use of the EHR and the completeness and overall quality of key data items required to support effective HIV care. The second decision support alert (see background section) was designed to warn if a VL result had not been uploaded to the system by 6 + 2 months after starting ARV treatment or 12 + 2 months after the last VL (allowing for 2 months transport and data entry delays). There was a 4-5 month period with the intervention in place before the data collection for this data quality study took place, allowing assessment of whether VL data was more complete in the EHR records from the 11 sites that received this alert. While data completeness and timeliness for VL data was overall well below 85 threshold, with only 11.9% completeness in the 39 control sites lacking the VL alert, it was significantly higher (26.7%) in the 11 intervention sites that had the VL alerts. A similar effect was seen for drug pickup data suggesting that the presence of VL alerts impacted overall data entry processes/behavior not just the specific data type in the alert. It should be noted that VL testing was infrequent prior to the start of the study as the laboratory capacity was still being established, which may have affected data management and data entry in the EMR, especially for older values.

The metrics for overall data quality and completeness may have been higher if there had been regular data quality improvement visits in Rwandan MOH sites, an effect seen with the study by Muthee in Kenya[14]. Another strategy for improving data quality and user experience is the implementation of EHR systems allowing direct point of care data entry rather than entering data on paper and in the EHR as seen for viral load results in Kenya[32]. The findings here also indicate that data quality and timeliness would have likely been improved more generally using automated tools to alert staff of missing data for key variables such as VL results, and potentially form field validation to detect impossible values, and comparison of date values with other data. The analysis of the degree of error in date fields showed that many incorrect one-off dates such as "date enrolled in HIV care" were within one month of the correct value about a third of the time and therefore may have been adequate for the care of HIV patients seen every 2-3 months. Incorrect repeated dates such as "last clinic visit", "drug pickup" and "viral load" were within one month of the correct date 12% or less of the time and therefore less fit for use in patient care. Overall most of the data was considered fit for purpose as confirmed by the user survey[4] which showed a reliance on the EHR data for clinical care and reporting. Regular monitoring of data quality metrics and intervention in poorly performing sites would likely help address major deficiencies efficiently.

The implications of these results are much broader than just HIV care which provided the environment for this study. Primary care EHR systems remain rare in LMICs but recent progress with facility based EHR systems like OpenMRS in Haiti, Sierra Leone, parts of India and in the Philippines are laying the foundations for wider use[33]. Mobile and tablet-based systems that support care in communities and some health facilities should provide a much wider range of data for clinical and public health uses. Open Smart Register Project (OpenSRP[34]) led by the WHO now reaches 150 million patients, the majority in Bangladesh. Other mHealth systems such as CommCare[35] and Open Data Kit[31] are used in many LMICs[36]. All of these systems have the ability to share data with EHRs like OpenMRS. They increasing support interoperability standards like HL7 FHIR[27] for interoperability with a range of other health information systems including lab and pharmacy systems, radiology information systems, and the District Health Information System (DHIS2)[37]. FHIR based APIs and toolkits are also being built into Apple OS and Android operating systems[36]. Projects using data science techniques and machine learning are increasing in LMICs and require high quality data sets that are representative of all communities and groups to avoid biases or poor performance in vulnerable groups[38].

Comparison to previous work

Muthee and colleagues studied the data quality in 53 health facilities running the KenyaEMR OpenMRS

based system[14]. They analyzed data from a data quality assessment (DQA) which had been collected by MOH lead teams that visited the health facilities. They found one or more missing values in 735 (31%) paper forms and 747 (32%) of Kenya EMR records, a higher missing rate than in this study. Concordance scores were calculated the same way as the current Rwanda study, with a mean concordance of 11.9/20 at baseline. They were also able to compare DQA results to a second DQA assessment to 27 health facilities. They showed that the health facilities with previous DQA had substantially better data quality than the first visit with missing data falling to 13% and Concordance scores increasing to a mean of 13.6. This result suggests that the process of measuring data quality and giving feedback to staff can have long term benefits. Mean concordance scores at the second DQA 13.6/20 (68%) were equivalent to those in Rwanda 10.2/15(68%). Abiy et al studied data quality in an HIV treatment program in Ethiopia[39]. They compared paper and EHR versions of 250 records. Data completeness was 78% (95% CI 70.8% - 85.1%) for paper records and 76% (95% CI 67.8% - 83.2%) for the EHR. They used the Kappa statistic to compare variables in both records, with results ranging from demographics 0.93 to WHO stage 0.86, and "general appearance" 0.83. Ngugi et al also studied clinical use, and the completeness of data entry in KenyaEMR at 219 health facilities in Kenya[40]. They showed a wide variation in data entry per month in different health facilities and noted that this was likely affected by "patients' volume, frequency of patients' visits (encounters), EHRs mode of use and active usage of the system during care". Haskew et al studied the quality of data collection and the effects of clinical alerts on HIV patient care in Western Kenya before and after the cloud-based implementation of an OpenMRS EHR. They showed that clinical alerts reduced missing data and improved quality of care. Missing data was significantly reduced for key variables including "patient source" (375 – 69 patients), "first CD4 count" (826 – 354) and "first WHO stage" (2,258 – 479), all $P<.001$. The number of patients eligible for ART (based on CD4 count and/or WHO stage) but not yet started, fell from 1346 (29.6%) to 270 (6.2%), $P<.001$ [41].

A study of Health Management Information system in 16 health centers, 28 health posts and one hospital in Ethiopia, showed that training staff to fill out the data forms was significantly associated with improved data quality OR = 2.253, 95% CI [1.082, 4.692]), and measures of effective supervision and leadership were also beneficial[42]. A study of HIV clinical data collection systems in 21 sites (18 in in Africa) showed a reduction in missing data associated with training in data management (odds ratio, OR: 0.58; 95% confidence interval, CI: 0.37–0.90) and weekly hours spent by a data clerk working on the data collection system (OR: 0.95; 95% CI: 0.90–0.99)[16]. Previous studies suggest that poor completion and/or accuracy of key data items required for clinical care is likely to limit the benefit of EHR systems in day-to-day patient management, for example the functioning of CDSS in the study here, and that reported by Haskew et al above. It also impacts quality of care metrics (such as for in cancer care in the US[43]), and in efficient reporting of data for clinical teams, MOH and funders[16] [42]

Limitations

The study was limited to a subset of the MOH run OpenMRS EHR systems in use for HIV care. The inclusion criteria for the 112 sites in the larger study favored health facilities with better IT hardware and successful implementation of a server monitoring tool[4]. While the intervention sites appeared to show improved data quality and completeness for certain variables this effect might change over time. Although the user survey showed that staff in intervention and control sites received similar amounts of training and were equally positive about their training, it is possible that differences in the type of training in the intervention sites and emphasis on the importance of VL contributed to better data quality. More robust and longitudinal studies are needed to understand the effect of different technologies on the use of EHR, including variations within and across sites and individuals.

Conclusions

The study found that there was generally high data completeness in both EMR and paper records in Rwandan MOH supported health facilities. The percentage matching and the concordance scores for the EHR and paper records was lower, but this mainly affected recording of dates. The considerable potential benefits of EHR systems for patient care for HIV and other diseases will only be realized with sufficient support, training and monitoring of data quality, and adequate technical support and infrastructure to ensure reliable systems. Use of

"point of care" EHR systems rather than transcription of data from paper records can also improve user experience and data completeness. The results here show that automated alerts regarding data quality, completeness and timeliness can significantly improve these metrics in remote health facilities in a low-income country. To be fully effective however these alerts should be combined other strategies to ensure high system uptime and a range of data quality improvement strategies including training, regular supervision, and feedback. These improvements will likely help drive much wider use of EHR data in LMICs for clinical care, reporting, data science, and machine learning. The data collected will also support population and public health uses including syndromic surveillance, notifiable disease reporting, real time monitoring of disease burden, and forecasting resource requirements to meet care needs.

Acknowledgements

We thank clinic staff at the study sites in Rwanda for their participation in the data quality study, the staff of the Rwanda Biomedical Center and the MOH, and the informatics team at Partners in Health in Rwanda, for technical assistance and advice. We thank Dr Eric Green (Duke University) for assistance with the RCT design and randomization. We thank Henry "Roo" Fraser for assistance with creation of diagrams.

Conflicts of interest

Dr Fraser and Dr Seebregts are co-founders of the OpenMRS EHR project. Dr Richards and Mr Santas worked for the CDC Division of Global HIV and TB which funded the study. None of these individuals received compensation for their time from the CDC implementation science grant. Dr Fraser received travel and accommodation support for study meetings and workshops from the grant. Jembi health systems, of which Dr Seebegts is the director, received funds from the grant for the software development work on OpenMRS. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agencies.

Funding

This research has been supported by the President's Emergency Plan for AIDS Relief (PEPFAR) through the Centers for Disease Control and Prevention (CDC) under the terms of program grant number U01 GH000782-01 awarded to the National University of Rwanda. Additional funding and in-kind support was provided by Brown University, Providence, RI, USA. HF was a Marie Skłodowska-Curie Fellow 2015 - 2017, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 661289, "Global eHealth."

Data sharing

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

References

1. HIV/AIDS: WHO African Region; 2023 [cited 2023 february 15]. Available from: https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update
2. Rwanda Population-based HIV Impact Assessment. ICAP University of Columbia. : Columbia University; 2019 [Available from:

https://phia.icap.columbia.edu/wp-content/uploads/2019/10/RPHIA-Summary-Sheet_Oct-2019.pdf.
3. Rich ML, Miller AC, Niyigena P, Franke MF, Niyonzima JB, Socci A, et al. Excellent clinical outcomes and high retention in care among adults in a community-based HIV treatment program in rural Rwanda. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2012;59(3):e35-e42.
4. Fraser HSF, Mugisha M, Remera E, Ngenzi JL, Richards J, Santas X, et al. User Perceptions and Use of an Enhanced Electronic Health Record in Rwanda With and Without Clinical Alerts: Cross-sectional Survey. JMIR Med Inform. 2022;10(5):e32305.
5. Fraser HS, Jazayeri D, Nevil P, Karacaoglu Y, Farmer PE, Lyon E, et al. An information system and medical record to support HIV treatment in rural Haiti. BMJ. 2004;329(7475):1142-6.
6. Amoroso CL, Akimana B, Wise B, Fraser HS. Using electronic medical records for HIV care in rural Rwanda. MEDINFO 2010: IOS Press; 2010. p. 337-41.
7. Verma N, Mamlin B, Flowers J, Acharya S, Labrique A, Cullen T. OpenMRS as a global good: Impact, opportunities, challenges, and lessons learned from fifteen years of implementation. International Journal of Medical Informatics. 2021;149:104405.
8. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc. 2010;2010:232-6.
9. Fritz F, Tilahun B, Dugas M. Success criteria for electronic medical record implementations in low-resource settings: a systematic review. Journal of the American Medical Informatics Association. 2015;22(2):479-88.
10. Were MC, Nyandiko WM, Huang KT, Slaven JE, Shen C, Tierney WM, et al. Computer-generated reminders and quality of pediatric HIV care in a resource-limited setting. Pediatrics. 2013;131(3):e789-96.
11. Yehualashet DE, Seboka BT, Tesfa GA, Demeke AD, Amede ES. Barriers to the Adoption of Electronic Medical Record System in Ethiopia: A Systematic Review. J Multidiscip Healthc. 2021;14:2597-603.
12. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, et al. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc. 2013;20(e1):e2-8.
13. Aldredge NH, Rodriguez D, Gonzalez J, Burt DR. A Case Study of a Point-of-Care Electronic Medical Record [SABER] in Totonicapan, Guatemala: Benefits, Challenges, and Future Directions. Ann Glob Health. 2020;86(1):122.
14. Muthee V, Bochner AF, Osterman A, Liku N, Akhwale W, Kwach J, et al. The impact of routine data quality assessments on electronic medical record data quality in Kenya. PLoS One. 2018;13(4):e0195362.
15. Blaya JA, Shin SS, Yagui M, Contreras C, Cegielski P, Yale G, et al. Reducing communication delays and improving quality of care with a tuberculosis laboratory information system in resource poor environments: a cluster randomized controlled trial. PLoS One. 2014;9(4):e90110.
16. Forster M, Bailey C, Brinkhof MW, Graber C, Boulle A, Spohr M, et al. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limited settings. Bulletin of the World Health Organization. 2008;86:939-47.
17. Fraser HSF, Mugisha M, Richards J, Chelliah W, Ngenzi JL, Gaju E, et al. A System for Monitoring Stability, Usage and Data Completeness in OpenMRS Electronic Health Records in Rural Rwanda. AMIA Fall Symposium AMIA; 2017.
18. Blaya JA, Shin SS, Yale G, Suarez C, Asencios L, Contreras C, et al. Electronic laboratory system reduces errors in National Tuberculosis Program: a cluster randomized controlled trial. Int J Tuberc Lung Dis. 2010;14(8):1009-15.

19. Hedt-Gauthier BL, Tenthani L, Mitchell S, Chimbwandira FM, Makombe S, Chirwa Z, et al. Improving data quality and supervision of antiretroviral therapy sites in Malawi: an application of Lot Quality Assurance Sampling. BMC Health Services Research. 2012;12(1):1-8.
20. Birkhead GS, Klompas M, Shah NR. Uses of electronic health records for public health surveillance to advance public health. Annu Rev Public Health. 2015;36:345-59.
21. Dornan L, Pinyopornpanish K, Jiraporncharoen W, Hashmi A, Dejkriengkraikul N, Angkurawaranon C. Utilisation of electronic health records for public health in Asia: a review of success factors and potential challenges. BioMed research international. 2019;2019.
22. Adedimeji AA, Hoover DR, Shi Q, Kim HY, Brazier E, Ross J, et al. Trends in demographic and clinical characteristics and initiation of antiretroviral therapy among adult patients enrolling in HIV care in the Central Africa International epidemiology Database to Evaluate AIDS (CA-IeDEA) 2004 to 2018. Journal of the International AIDS Society. 2021;24(6):e25672.
23. Klompas M, McVetta J, Lazarus R, Eggleston E, Haney G, Kruskal BA, et al. Integrating clinical practice and public health surveillance using electronic medical record systems. Am J Public Health. 2012;102 Suppl 3(Suppl 3):S325-32.
24. Tomines A, Readhead H, Readhead A, Teutsch S. Applications of electronic health information in public health: uses, opportunities & barriers. EGEMS (Wash DC). 2013;1(2):1019.
25. Herwehe J, Wilbright W, Abrams A, Bergson S, Foxhood J, Kaiser M, et al. Implementation of an innovative, integrated electronic medical record (EMR) and public health information exchange for HIV/AIDS. J Am Med Inform Assoc. 2012;19(3):448-52.
26. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. Health Policy and Technology. 2017;6(1):20-5.
27. Bacher I, Mankowski P, White C, Flowers J, Fraser. HS. A New FHIR-based API for OpenMRS. AMIA Clinical Informatics Conference; May 20212021.
28. Wolfe BA, Mamlin BW, Biondich PG, Fraser HS, Jazayeri D, Allen C, et al., editors. The OpenMRS system: collaborating toward an open source EMR for developing countries. AMIA Annual Symposium Proceedings; 2006: American Medical Informatics Association.
29. OpenMRS Annual Report. Indianapolis, Indiana, USA: OpenMRS Inc; 2021.
30. Allen C, Jazayeri D, Miranda J, Biondich PG, Mamlin BW, Wolfe BA, et al. Experience in implementing the OpenMRS medical record system to support HIV treatment in Rwanda. Studies in health technology and informatics. 2007;129(1):382.
31. Open Data Kit 2022 [Available from: https://getodk.org.
32. Masibo WS. Automated Clinical Decision Support in HIV management: a comparative study of Point-of-Care and Retrospective Data Entry outcomes Nairobi, Kenya: Strathmore University; 2021.
33. Fraser HS, Adedeji T, Amendola P. “The importance of primary care records in low and middle income settings for care, resource management and disease surveillance”. ICT in Health 2021, Survey on the Use of Information and Communication Technologies in Brazilian Healthcare Facilities
. Sao Paulo, Brazil: Brazilian Ministry of Health; 2021. p. 273 – 95.
34. OpenSRP – Open-source smart register platform (SRP): Opensrp; 2023 [Available from: https://opensrp.io/.
35. CommCare: Dimagi Inc; 2021 [Available from:
https://www.dimagi.com
.
36. McCool J, Dobson R, Whittaker R, Paton C. Mobile Health (mHealth) in Low- and Middle-Income Countries. Annu Rev Public Health. 2022;43:525-39.

37. Integration & Interoperability with DHIS2 2021 [Available from: https://dhis2.org/integration/.
38. Parikh RB, Teeple S, Navathe AS. Addressing Bias in Artificial Intelligence in Health Care. JAMA. 2019;322(24):2377-8.
39. Abiy R, Gashu K, Asemaw T, Mitiku M, Fekadie B, Abebaw Z, et al. A Comparison of Electronic Medical Record Data to Paper Records in Antiretroviral Therapy Clinic in Ethiopia: What is affecting the Quality of the Data? Online J Public Health Inform. 2018;10(2):e212.
40. Ngugi P, Babic A, Were MC. A multivariate statistical evaluation of actual use of electronic health record systems implementations in Kenya. PloS one. 2021;16(9):e0256799.
41. Haskew J, Ro G, Turner K, Kimanga D, Sirengo M, Sharif S. Implementation of a Cloud-Based Electronic Medical Record to Reduce Gaps in the HIV Treatment Continuum in Rural Kenya. PLoS One. 2015;10(8):e0135361.
42. Teklegiorgis K, Tadesse K, Terefe W, Mirutse G. Level of data quality from Health Management Information Systems in a resources limited setting and its associated factors, eastern Ethiopia. . South African Journal of Information Management. 2016;18(1):1-8.
43. Schorer AE, Moldwin R, Koskimaki J, Bernstam EV, Venepalli NK, Miller RS, et al. Chasm Between Cancer Quality Measures and Electronic Health Record Data Quality. JCO Clin Cancer Inform. 2022;6:e2100128.

Preprint JMIR Publications

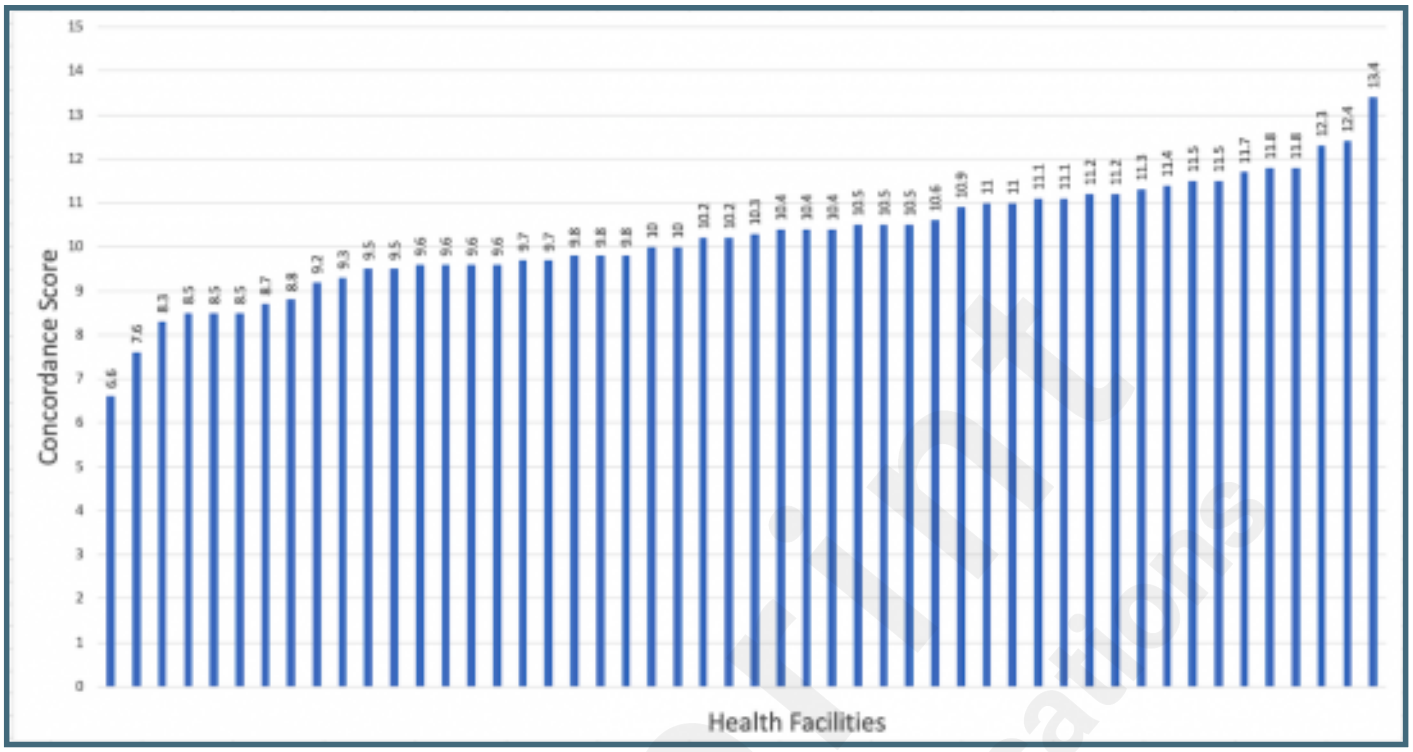
Supplementary Files

Preprint JMIR Publications

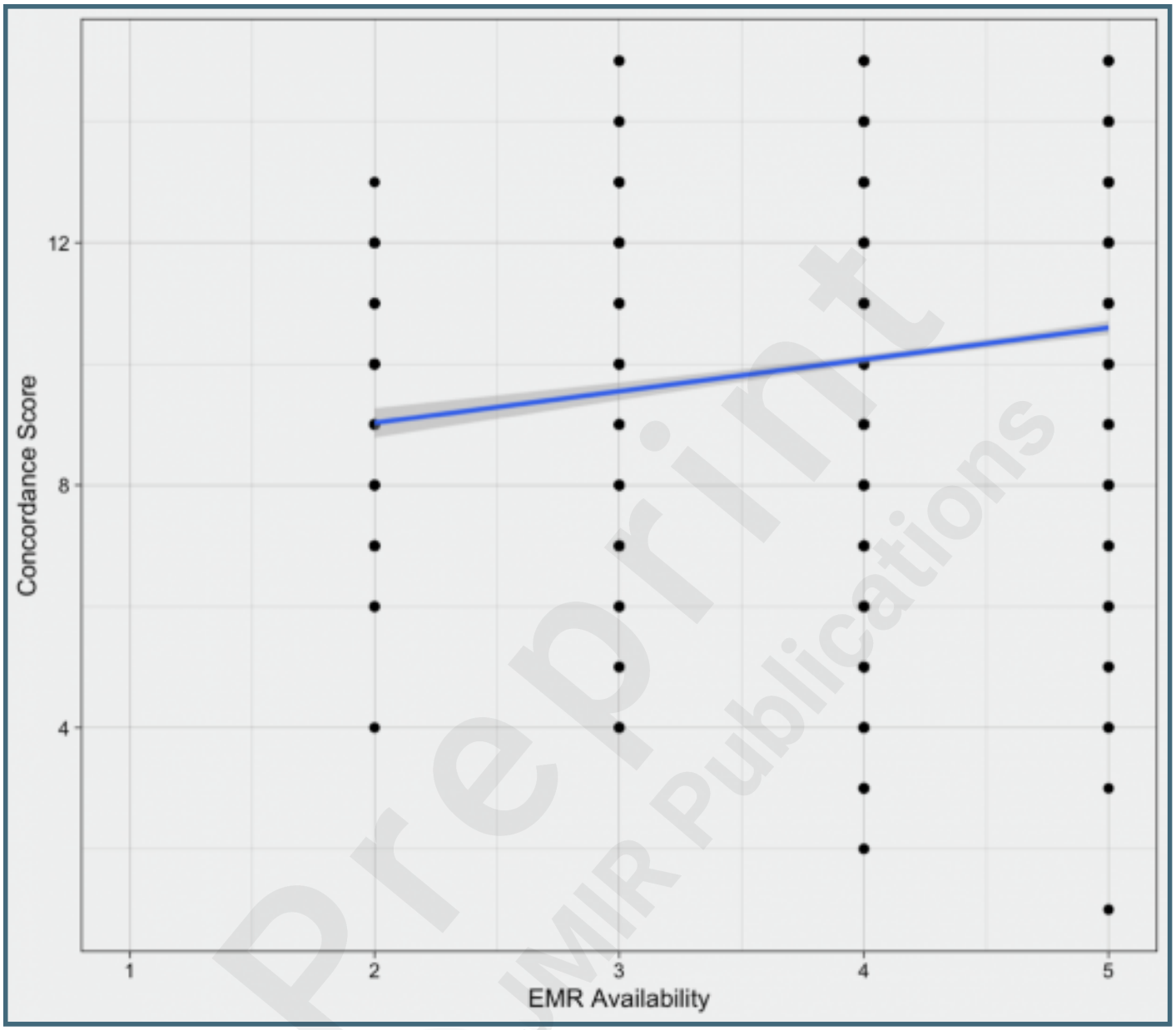
Figures

Preprint JMIR Publications

The mean concordance scores for all sites (maximum score 15).

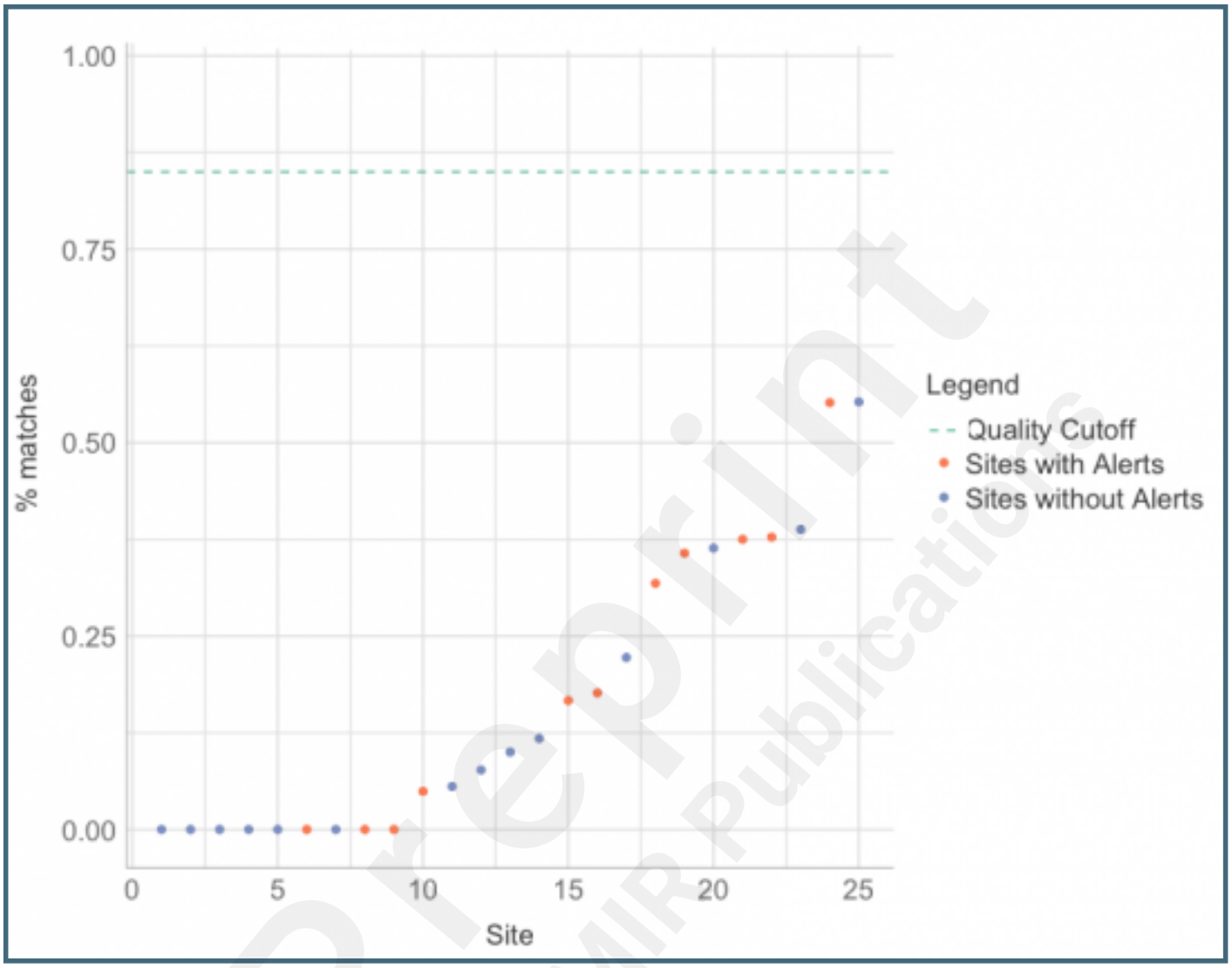


Correlation between concordance score and reported EHR availability (uptime).



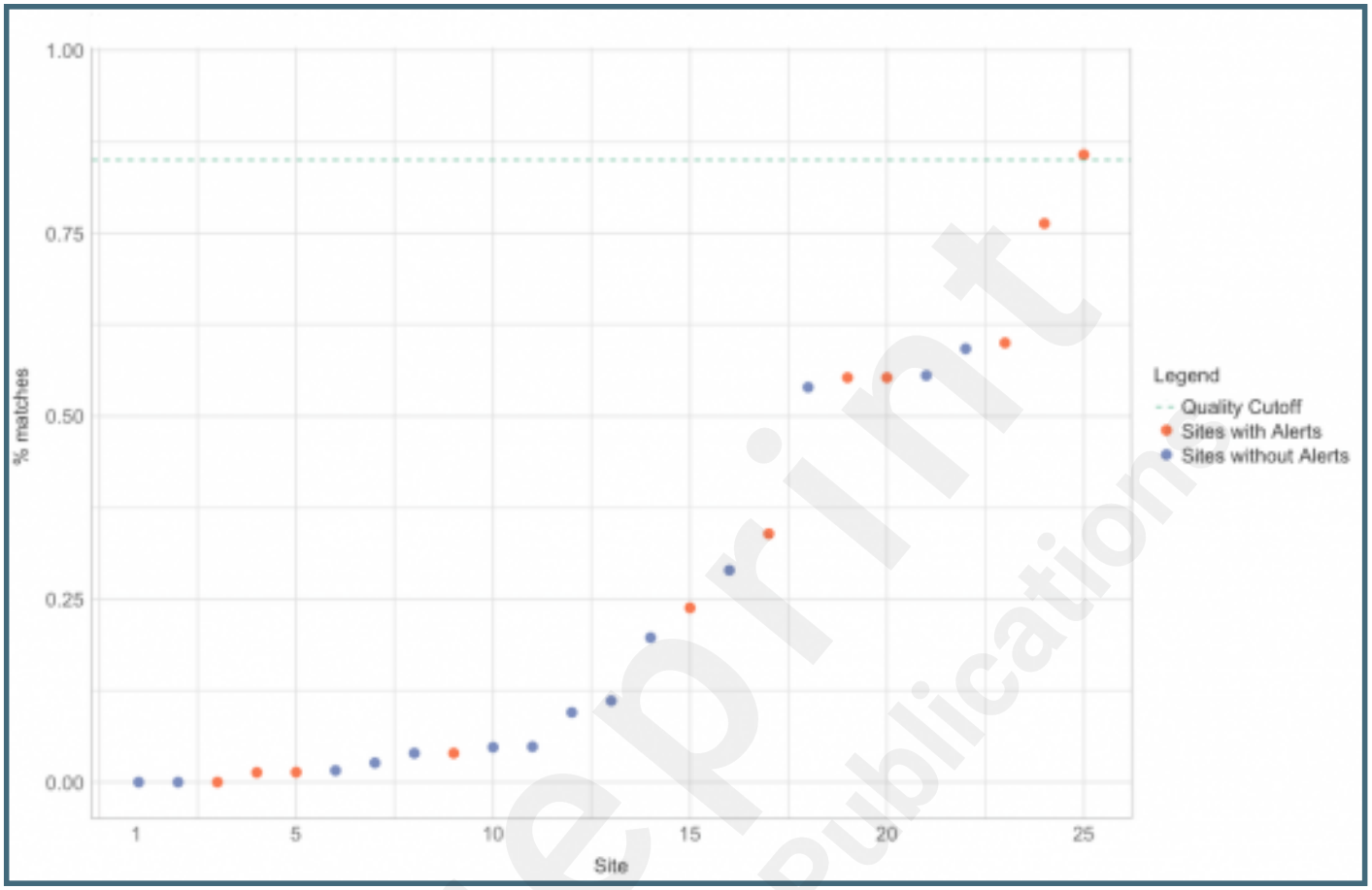
Preprint JMIR Publications

Graph of percentage data matches for VL results between paper records and EHR for each of the 25 health care centers in the intervention arm, ranked by percentage of matches.



Preprint
JMIR Publications

Graph of percentage data matches for drug pick-ups between paper and EHR records among the same 25 intervention sites as figure 4. Scores were higher overall than in Figure 3 but only one (intervention) site was over the data quality threshold.



Preprint
JMIR Publications