Describing Documentation of Electronic Health Records (EHR) in Anti-Retroviral Therapy (ART) Clinics to Improve Data Quality for Healthcare Processes in Malawi

by

Oliver Jintha Gadabu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Infrastructures and Learning Systems)
in the University of Michigan
2021

Doctoral Committee:

Professor Anne Sales, Chair
Associate Professor Marcelline Harris
Assistant Professor Zach Landis-Lewis
Assistant Professor Karandeep Singh

Oliver J. Gadabu

ojgadabu@umich.edu

ORCID iD: 0000-0001-7960-5376

Dedication

This dissertation is dedicated to those that go beyond believing in the power of health infrastructure and are not shy to make bold decisions to implement even the most unpopular interventions. This work is dedicated to the individuals that have made multiple personal sacrifices that have enabled me to immerse myself into health infrastructures and learning systems – yes and I mean you mate, Maggie. Your encouragement and endurance throughout the process that started some time back cannot be expressed easily in words.

I also dedicate this dissertation to my parents, my siblings, nephews and my extended family. Being the youngest in the family, you have thoroughly provided an environment that has helped understand the concept of "Ubuntu", "Umunthu" or having a deep appreciation of any person regardless of who they are. Together we have gone through many ups and downs, but by the nature of my current location of my studies, I miss our sister Olivia deeply. She is the one who drove me to the interviews for this Ph.D. journey and the joy she had in her eyes on the prospect of moving close to where she lived emanated from deep in her heart. I miss you and I dedicate this work to you too my dearest sister. I also dedicate this work to my dearest Aunt Nazia, whom I always looked forward to silent conversation. Lastly, but not least, I dedicate this work to my professional mentors. There are so many of you and you have all encouraged me in many ways. Gerald P. Douglas, you created the foundation for health informatics in Malawi by introducing point of care EHRs and that have changed the health situations of many people. I also dedicate this work to you.

Acknowledgements

“Ask, and it will be given to you; seek, and you will find; knock, and it will be opened to you” –
Matthew 7:7

This work would have never been possible without the guidance, patience and wisdom of my committee. I would like to thank all of you dearly. I would also in particular like to thank my committee chair, Prof. Anne Sales, for your patience and demand for high standards of work and generous support all the time. I shall forever be grateful.

I would also like to acknowledge the support I received from the Malawi Ministry of Health (MoH) through the department of HIV/AIDS that enabled me to get research approval for this work. Similarly, I would also like to thank the leadership and staff of Baobab Health Trust for your technical support in extracting the data as well as in the research process via coordinating with health facilities and other partners.

I would also like to acknowledge Mwatha Bwanali, Deliwe Nkhoma, DavieTembo, Timothy Mtonga, Chimango Munthali and Hannock Twea for your support whilst collecting my data, identifying the appropriate data fields and validating some of the initial data queries and arranging meetings.

Lastly but not least, I would like to acknowledge the learning experience of my classmates in the HILS program throughout the journey. It was not only great to learn from your ideas, but also a good cultural exchange for me.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

In healthcare, documentation deals with recording of significant patient encounters within a healthcare information system. Documentation has evolved from paper-based tools to electronic means such as Electronic Health Record (EHR) systems. The EHR may optimize healthcare processes by facilitating communication, offering decision support, and allocating and distributing resources to health facilities for patient care. Faced with the Human Immunodeficiency Viruses that leads to the Acquired ImmunoDeficiency Syndrome (HIV/AIDS) pandemic, Malawi has embraced an EHR for its Anti-Retroviral Therapy (ART) clinics since 2006. The 90-90-90 HIV/AIDS eradication strategy which Malawi adopted with the leadership of the World Health Organization (WHO) United Nations against AIDS (UNAIDS) may benefit from this EHR. In this initiative, identification and enrollment of HIV positive clients into ART clinics is essential for viral load suppression. The EHR benefits rely on good data quality, which has been lacking in Low-Income Countries (LICs) including Malawi. Although data quality has many definitions, a widely accepted concept is "data quality is fit for purpose." This concept has not been used in many data quality assessments, particularly in LICs. I asked the following research question:

How can the varying purposes of EHR data use for different stakeholders inform interventions for EHR data quality improvement?

I had the following research aims:

- To understand the importance of data quality characteristics in the context of EHR stakeholders' purposes of data use
- To conduct a data quality assessment and understand the drivers of the observed data quality

I conducted this research at primary, secondary and tertiary level ART clinics that manage over 200 patients a day in Malawi.

My study used an observational study design with mixed methods. I conducted semi-structured interviews with 34 stakeholders comprised of nurses and clinicians as well as public health officials and donors who support the EHR. Additionally, I assessed 160,647 patient records that had 549,826 visits across 10 different health facilities, extracted from existing EHR data.

My work established that stakeholders have clinical facing purposes or administrative purposes of data use. Stakeholders with clinical facing purposes of data use expected their data to be plausible. On the other hand, stakeholders with administrative purposes of data use expected completeness first. I found variation in the observed proportions of the data quality characteristics of completeness and plausibility. Completeness ranged from 5% to 99% while plausibility ranged from 40% to 99% across variables essential to the 90-90-90 initiative. After integrating my qualitative and quantitative results, some results indicated agreement or convergence between different sources of information about data quality.

Data quality was high for the preferred characteristic of plausibility for patient health tracking. Conversely, there were divergent results between findings in my first and second aims for data management and use. For this purpose of data use, I found that plausibility was highly prioritized by participants, but in the data assessment, I found that only 40% of the records were plausible.

The recommendations that follow from my work regarding interventions to improve data quality include:

- Consider priorities based on importance of specific data quality characteristics for different stakeholder purposes of data use.
- Assess observed data quality proportions after conducting a quantitative data quality assessment to identify where interventions should first be implemented in relation with expected stakeholder data characteristics and purposes of data use.

Introduction

Many countries and health systems in the world continue to adopt electronic health records (EHR) as tools for documentation and facilitating patient assessment, patient care, and public health programming (1,2). EHRs have been implemented in high income countries (HICs) and in low- and middle-income countries (LMICs) (3). The implementation of EHRs is based on the hope of reaping some primary benefits of use that computerized systems promise such as better storage of information and easier computations, to secondary uses in healthcare such as facilitating better decisions by providing computerized clinical decision support for health care providers and hence better patient outcomes (4,5). The implementation of EHRs may also lead to secondary benefits of data use such as the better management of health resources such as medical drugs and other medical commodities by providing automated reporting (4,5). These benefits may result in the support of the design of health interventions that reduce patient morbidity and mortality.

These benefits, however, have hardly been realized despite the implementation of EHRs due to various challenges.

Some challenges reported concern EHR infrastructure due to limited power supply that leads to power failures, limited communication technology that leads to inability to transfer data across health facilities, and inadequate capacity of human resources that may lead to improper use of the EHRs (1).

Another challenge in the developing world is the source and amount of funding for implementing EHRs (10). The sources of funding usually have specific strategic goals and limited amounts of resources targeting their strategic goals that are usually specific diseases. This leads to fragmented EHR infrastructure from the beginning (7,11). This focus on disease specific EHRs may be observed from funding opportunity announcements of major donor funding organizations and programs such as the Centers for Disease Control and Prevention with programs such as the President’s Emergency Program for AIDS Response and the President’s Malaria Initiative. The Gates Foundation has similar programs that focus either on Malaria or Maternal, Neonatal and

Child Health Interventions. The disease-focused EHRs usually lead to data silos and missing critical data for healthcare providers to use in their various roles. It may also lead to redundant documentation of patient encounters with the healthcare system, and thus lead to longer waiting times for patients at facilities.

Nevertheless, the developing world needs EHR investments to effectively tackle some of the disease burdens, which in most cases are global emergencies, such as the HIV/AIDS pandemic (3). Consequently, most of the EHRs developed and implemented in the developing world, particularly sub-Saharan Africa, have focused on HIV/AIDS program monitoring and evaluation. In health facilities that have implemented EHRs for some time in the developing world, there have been additional reported challenges to the ones mentioned above, such as poor data quality (12). Data quality may have different definitions and interpretations, depending on the data quality framework chosen to measure it (13). One perspective is that data quality should be defined with consideration of the purpose of use of any data whose quality is being examined (14). This perspective brings in the view that data quality is fit for purpose. Poor data quality directly defeats the benefits of EHRs at the patient and population levels. For example, poor data quality is a barrier to decision support. Poor data quality is also a barrier for national program strategic goals such as the proposed 90-90-90 HIV/AIDS eradication goals (78).

This strategic goal first proposed by the UNAIDS and the WHO has been adopted by many governments including in the developing world. The 90-90-90 HIV/AIDS goals include (78):

- Identifying 90% of a country's population that is HIV positive via HIV diagnostic tests;
- Making sure that 90% of the identified HIV positive population is enrolled in care and retained in care accessing anti-retroviral (ARV) therapy treatment; and lastly,
- Making sure that 90% of the maintained population achieve viral load suppression. The recommended viral load for patients to remain healthy and reduce morbidity due to HIV opportunistic infections, is below 1000 copies/ml (77) and this is considered viral load suppression.

In the case of EHRs being implemented for HIV/AIDS management, poor data quality has severe implications for monitoring progress toward the 90-90-90 targets (15).

My study addresses the problem of poor EHR data quality in Anti-retroviral Therapy (ART) clinics for HIV/AIDS, as reported in Malawi (9). It builds on the rich body of literature of a mixed-methods data quality study conducted in ART clinics for HIV in Haiti (16). A previous study conducted in Malawi identified missing data as a barrier for providing audit and feedback reports of specific clinical practice guideline recommendations to healthcare providers (9). Similarly, a study in Haiti assessed data quality in selected health facilities in Haiti and found poor data quality based on the characteristic of completeness. The completeness for variables such as weight, height, and ART pregnancy status was measured in the range of 20% to 40%. In a follow-up study also in Haiti, the researchers presented a process for identifying variables to prioritize and implement with an intervention of automated audits (16). In their process, they suggested audit and feedback as an intervention for improving data quality prior to considering data quality characteristics in relation to stakeholder purposes of data use and the priority they would attach to a data quality characteristic.

My work examines electronic data quality to identify factors to consider when selecting interventions to improve it. These factors have the potential to increase the benefits of EHR use in meeting national priorities such as the 90-90-90 targets for HIV/AIDS eradication in Malawi. To begin investigating the challenge of data quality, I present "Background and review of the literature," in Chapter 1, where I provide definitions used in my study and discuss the state of EHR implementations in low-income countries (LICs). With my health infrastructure and learning systems background, I also discuss how my research fits into concepts of learning health systems and infrastructure. I introduce my research question and aims as well in this chapter. I also explain the data quality frameworks used in my study.

In Chapter 2, I describe my study methods. I used mixed methods to compare results for qualitative aims to my quantitative aims with an appropriate integration strategy. The methods include the data collection and data analysis plan for my research. The perspective I adopted for my research is that data quality is "fit for purpose."

Chapter 3 presents the qualitative findings and the quantitative results of my study, where I establish that different stakeholders have different purposes of data use and attach different priorities when it comes to data quality characteristics that they expect in the data that they use. This chapter also presents my quantitative data quality assessment results, where I found poor data quality for some of the 90-90-90 variables, such as patient physical address completeness, which is necessary for patient health tracking, and patient drug adherence completeness data quality for monitoring how well patients are taking their drugs which is essential to achieve viral load suppression. On the other hand, I also present results with high data quality for monitoring patient body mass index using weight and height and these are essential for routine patient care. A continuously declining BMI is a sign of treatment failure and may affect retaining these patients in care.

In Chapter 4, I discuss how the differing priorities for data quality characteristics may inform interventions for improving data quality, including a procedure for identifying interventions for data quality improvement in low-income settings.

The chapters above, together with the frameworks I chose for my study, help me answer the research question I present in Chapter 1:

“How do stakeholders’ purposes of EHR data use inform interventions for EHR data quality improvement?”

Chapter 1 Background and Review of Literature

Background

1.0 What is Documentation?

Broadly, documentation is defined as the process of producing any proof in support of a fact(1). In healthcare, documentation deals with recording any details of a patient's encounters with a healthcare system. These processes may include the patient's history, clinical observations, diagnostic studies, medical history, clinical course or management, and any care-related outcomes and decision making(2). In the medical field, documentation has a long history that can be traced back to the Egyptian empire, where physicians recorded details of surgical procedures on papyrus(3,4). Such accounts of documentation almost 4,000 years ago emphasized the structure of the clinical documentation. This structure consisted of 4 main parts starting with an examination of the patient, moving onto diagnosis, then providing a suggested treatment, and lastly, a prognosis for the patient. Nowadays, a similar structure in documenting a patient encounter may be found with the addition of recommendations, such as uniquely identifying patients to allow for continuity of patient care, as advanced by global health organizations such as the World Health Organization (WHO)(5).

The WHO has also recommended documentation mechanisms, such as the use of specific forms for particular healthcare services in Low and Middle-Income Countries (LMICs) to ensure quality in documentation(6). Recommended documents such as these have been adopted and adapted for different health organizations to enhance the quality of documentation. In addition to these forms, most national monitoring and evaluation programs have adopted indicators for measuring disease burden from the WHO. Specific variables need to be collected in order to report on these indicators. Examples of such indicators are those for measuring HIV/AIDS disease burden. These include indicators for prevalence or incidence of HIV/AIDS disease in different countries. Most countries measure these indicators by counting cases of HIV infected individuals and clinical or diagnostic reasons for patients being enrolled into HIV/AIDS programs plus the primary and secondary disease outcomes of the patients. The primary outcomes describe drug regimens that patients are on and respond to indicators such as health facilities that had stock outs (7). Secondary outcomes describe adverse outcomes such as death and patients that have been lost to treatment. Techniques such as cohort analysis are used to achieve these measurements(8). The forms proposed by the WHO and associated indicators have been adapted to come up with tools suitable for their different LMIC settings(9). These forms have been the basis for defining minimum data fields in EHR systems used in HIV/AIDS clinics. Given that the EHR may be used beyond documentation as infrastructure that may routinely help improve healthcare systems, I considered the role of the EHR in the context of a learning health system which could be a model for improving health processes and outcomes.

1.1 The Learning Health System

A learning health system is defined as a structure in which "science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the delivery process and new knowledge captured as an integral by-product of the delivery experience" (10). Although such systems are nascent in the developing world, discussions have begun around the improvements in health processes that learning health systems may bring in the area of health research(11). The discussions have focused on pooling data from several research institutions as a step towards enabling learning health systems. A learning health system may exist at any level of health interaction (12) and is enabled by a learning health cycle that requires an appropriate infrastructure. The underlying infrastructure is the support or base for the learning health cycle. The learning health cycle describes the transition of data in health settings into knowledge and the knowledge into performance for use in a learning health system (13).

Support is generally defined as an underlying base or foundation for an organization or a network and is an essential component for a learning health system. Learning health systems may function independently but synergistically to deliver services to a community of stakeholders (14). For example, at a global level, each country may have an independent electronic health system as infrastructure. Still, these standalone systems may follow data exchange standards to synergistically respond to global needs via coordinating bodies such as the WHO. Examples of standards to allow this coordination are the indicator standards that different countries adopt for HIV/AIDS reporting (7). This would enable learning at a technical level of data. Additionally, there would be a need to have common practices and guidelines to enable any discovered knowledge identified to transition into improved performance at a people or society level. This would meet some socio requirements of infrastructure as healthcare providers would need to be conversant with these practice guideline recommendations. A common socio-technical infrastructure that would have technical infrastructure via standards and socio infrastructure via clinical practice guidelines and recommendations may enable a learning health system. The HIV/AIDS EHR in Malawi is an example of such an infrastructure having embedded measures that contribute to national indicators and clinical practice guideline recommendations (15). I illustrate this concept of a learning health cycle with a common socio-technical infrastructure in the figure below.

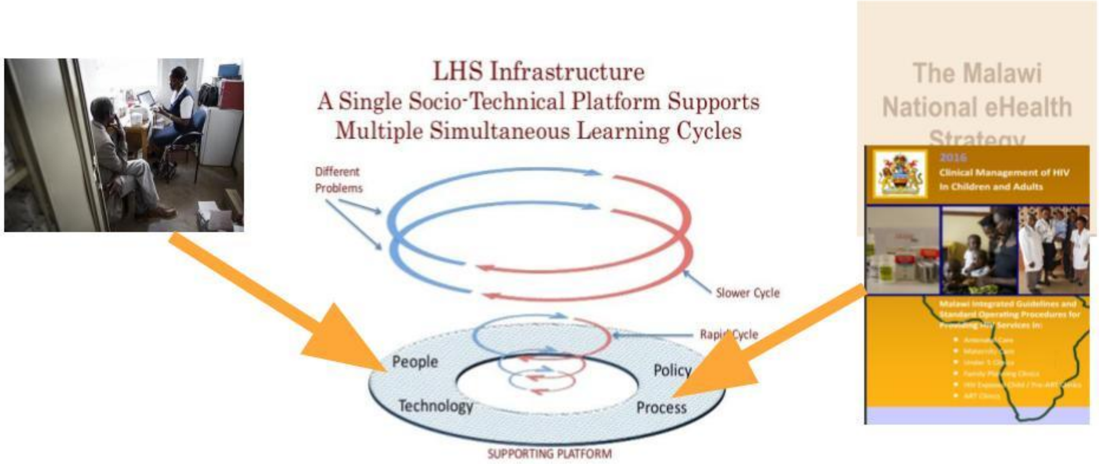


Figure 1.1 Common Components of Socio-Technical Infrastructure in Malawi

I discuss health documentation in the context of learning health cycle and health infrastructure in the next sections. Health infrastructure in the context of my study is represented by EHRs in ART clinics.

1.1.1 Documentation in the Context of the Learning Health Cycle

Documentation commences with the recording of facts, observations or impressions among other things and serving these in support of a phenomenon. Figure 1 illustrates my mapping of documentation processes in the context of my study to the learning health cycle. As can be seen, I map EHR documentation of health processes in Anti-Retroviral Therapy (ART) clinics for HIV in the Data to Knowledge section of the learning health cycle. This position is taken because my study assesses existing patient data in the ART clinics to identify knowledge that would be used to inform interventions for improving data quality in the ART clinics. Furthermore, my qualitative study was again conducted to understand other factors that may influence data quality and therefore gathering data for further analysis into knowledge.

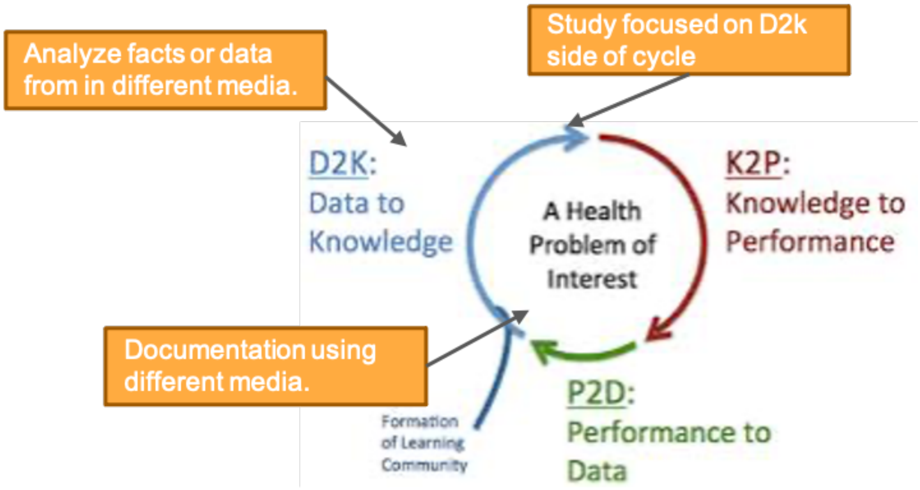


Figure 1.2 Mapping of Study to the Learning Health Cycle

1.2 What is the purpose of healthcare documentation?

Documentation in healthcare promotes communication and coordination among health care providers, provision of evidence for patient assessments, facilitation of continuity of patient care and facilitation of the development of public health programs (16).

Promoting communication and coordination has been highlighted in several studies where proper documentation has been shown to influence the quality of care delivered to patients in different settings. For example, one study in a maternity clinic in Senegal showed that documentation of postpartum hemorrhage at admission facilitated communication between health care providers and thus reduced adverse outcomes compared to cases where postpartum hemorrhage was not documented (17).

Additionally, documentation provides evidence and context that health care providers may use to make decisions (16). In chronic illnesses such as HIV/AIDS, documented observations of continued weight loss over a long period may suggest the need to reconsider treatment options as this may be an early sign of treatment failure (18). In this case, documentation provides a basis for clinical reasoning and decision making (19).

Continuity of patient care is another essential application for documentation with observations that documentation determines the tests that are ordered for patients transferred for care at other facilities (20). Proper documentation, in this case, may influence treatment outcomes as the patient may receive quicker treatment when accurate information is available.

Documentation is also essential for facilitating public health programming via analysis of clinical records that are a result of documentation. Proper documentation can result in the adequate allocation of resources and, eventually, the implementation of public health programs. This use of documentation was particularly crucial for the Malawi national HIV program (21).

Documentation in health is also traditionally used for medico-legal functions as a signed and dated medical record may be used for legal purposes. This purpose of data use has been emphasized by communities of practice, and an example is the American Speech-Language-Hearing Association (ASLH) that has developed explicit guidelines so that the resultant documents may stand as evidence in a court of law when need be (22).

Beyond the benefits of documentation mentioned above, additional interest for promoting documentation in health settings has been the desire to improve health care processes and improve patient outcomes. Documentation provides an opportunity to collect data that may be used to respond to specific indicators of quality of healthcare delivery. Thus, documentation may also be seen as a way to support primary and secondary data use in the context of ART clinics for HIV.

1.3 Operational Definition of Documentation for this Study

For purposes of this study, documentation is:

A process of recording facts or use of recorded facts in an Electronic Health Record system (EHRs) about patients' health during a patient and healthcare provider encounter by healthcare providers in ART clinics for HIV in Low-Income Countries (LICs).

1.4 How is healthcare documentation conducted in different parts of the world?

Healthcare data capture may take several forms ranging from handwritten notes on patient charts to the use of electronic media such as EHRs or automatically generated laboratory test results (19). Paper is the most basic form of documentation, requires little training and technology (23), and has been the form of documentation in healthcare settings for a long time. Unlike its digital counterparts, which can be limited by downtime due to technology failures, paper-based documentation tools have been considered reliable from an availability perspective (23).

Despite the convenience of using paper-based tools for documentation in healthcare, advances in technology have presented opportunities for solving challenges of using paper-based documentation. The problems include: the patient record being unavailable to multiple health care providers at the same time; data loss due to paper damage; legibility of hand-written patient records; passiveness of paper records which suffer from not being computable and hence not being able to incorporate automated reminders or alerts; privacy risks due to non-privileged access once in the open; and difficulty enforcing accountability due to lack of full chain of custody of the patient record(23). Technology has been developed to address some of these challenges, beginning with the development of electronic health records. These EHRs include building blocks such as Laboratory Information Management Systems (LIMS), Computerized Physician Order Entry systems, Picture Archiving and Communication Systems , Electronic Health Record systems (EHRs), mobile health applications (mHealth) and Civil Registration and Vital Statistics Systems (24).

The applications above facilitate stakeholders' primary data use. This use leads to data by-products or secondary data uses, such as reporting for quality improvement programs and public health programming. These are examples of EHR benefits that go beyond the low-level benefits realized from technology, such as storage of patient records.

The low-level benefits such as larger storage capacities for patient records and the prevention from damage due to natural events such as weather have been widely accepted and reported (24). However, high-level benefits, such as improving the quality of patient care processes and improved patient outcomes through techniques such as decision support, remain a work in progress (25). For example, an assessment of EHR deployments in the US observed that the use of EHRs to improve patient care processes and outcomes and reduce healthcare costs have not yet been reached (25). In attempting to solve this issue, high income countries (HICs) have exerted efforts to introduce process indicators that may help in measuring the benefits of EHRs. There have also been efforts to promote primary and secondary level use of data through establishing continuous patient care through new approaches such as learning health systems (13).

Unlike in the U.S., EHRs in LMICs have mostly been implemented through donor support (26). Most of these EHRs have been program focused, being implemented for specific diseases such as malaria, tuberculosis, and HIV/AIDS (22). Additionally, the type of EHRs in LMICs is influenced by the overall cost of implementing technology in these settings. Therefore, most programs opt for low-cost technology such as mHealth (27). A recent survey of Health

Information Technologies (HITs) in 16 LMICs showed that mHealth technology projects comprised 71% of the technology implementations, while 29% were computer-based (27). Public health reporting and programming also drive most of the EHRs in these LMICs, with the data being collected influenced by major local government bilateral partners (28). For example, in 2007, the Ministry of Health (MoH) and its partners made the recommendation to deploy an EHR in Malawi for the HIV/AIDS sector after they observed data quality challenges while manually compiling routine reports (15,21).

I identified themes such as the scale of implementation, policy for application, integration among implemented EHRs in different settings, funding mechanisms, and business models that enabled implementation of the different EHRs and different designs of the HITs as characteristics to describe the implementation of EHRs in HICs and LMICs (15,25,26,29–31). I summarized the characteristics across the various economic settings in Table 1.1 below.

Table 1.1 Comparison of Implementation of HITs in HICs and LMICs

HIT Characteristic	High-Income Countries (HIC)	Low Middle-Income Countries (LMIC)
Scale	Usually hospital-wide and national	Disease-specific and program-oriented
Legislation	Often through act of parliament and other policy documents	Missing, or eHealth strategies or policy initiatives through government departments
Integration	Multiple vendors, health information exchanges	Multiple implementers, some have a Master Patient Index (MPI) but missing in most countries.
Funding	Government-funded mostly	Donor funded mostly
Design	Off the shelf mostly (Pre-built but customizable to health facilities)	Custom-built using open source tools. Usually, regional focus.
Data Use	Combination of primary and secondary data use.	Mostly for secondary data use such as program reports

Despite varying levels of EHR and developments such as the challenges in different healthcare settings, the discussion above shows that EHRs are a mechanism for documentation and a tool for improving healthcare delivery is available in LMICs.

1.5 Examples of EHR deployments as Health Infrastructure in the LICs

I conducted this study with a focus on EHRs in Anti-retroviral Therapy (ART) clinics in Malawi. I therefore reviewed EHR implementations in similar settings. Seventeen out of 34 LICs have reported the existence of some sort of Health Information Technologies (HIT) to the WHO as of 2018(32). Some examples of LICs that have adopted HIT in addition to Malawi are Haiti, Rwanda, and Kenya. I reviewed the implementations in these countries through publications that have described how the EHRs are implemented. I found that several non-profit organizations led disease-focused EHR implementations. HIV/AIDS is the leading disease around which the EHRs have been developed (26).

Other LICs such as Haiti have also implemented EHRs in ART clinics to manage patients with HIV (33). The implementation in Haiti has evolved from collecting patient data on paper forms and retrospectively entering these into the EHR by data entry clerks, to a real-time point of care EHR where documentation of the patient encounter is done directly by healthcare providers (33). Despite managing over 150,000 HIV positive patients, the MoH in Haiti and other stakeholders have expressed concern over the data quality in the EHR that led to data quality assessment efforts being initiated in the country (34). In these efforts, stakeholders were asked to prioritize variables that were selected for automated data quality assessments in the Haiti EHR that was being used in ART clinics for HIV (35).

Although not a LIC, but rather a Lower Middle-Income Country, Kenya has also implemented an EHR in ART clinics for managing HIV/AIDS patients called KenyaEMR (36). I discuss this implementation because of the similarity in the motivation for implementing the EHR in ART clinics for HIV in Kenya as in other LICs. This motivation is to improve patient health and program monitoring and evaluation (22). Although the EHR in Kenya was implemented in over 340 health facilities at the time of this study, there were reports of data quality challenges. For instance, a study that was conducted to assess completeness and concordance of selected patient observations before and after implementing the EHR found that health facilities that had implemented the EMR for more than 12-15 months were 1.31 times more likely to have missing data than those that had the EMR implemented for 5-8 months. Thus, there was a decline in data quality over time of the implementation in the Kenya setting. The study observed that poor data quality is a threat towards achieving benefits of the EHR, such as public health reporting and thus recommended routine data quality audits (37).

The Malawi EHR has mainly been implemented to manage patients enrolled in ART clinics to improve documentation and the care process (15) by providing guideline-recommended reminders and alerts at the point of care. The EHR implementation in the country manages 50% of all patients alive and on treatment.

Despite the successes reported in Malawi, such as engineering solutions to power outages (15) to address infrastructure challenges reported in similar settings, poor data quality is a challenge that was identified as a barrier to implementing some interventions that may improve patient care via the EHR (38). Another study directly assessed the agreement of paper-based patient register records versus EHR records in Malawi (39). The study found that out of 31,763 patients registered in the EHR, only 28,359 (89%) were found in the paper-based registers. The study also found significant differences in patient death outcomes at two of the five health facilities. These two findings confirm the challenges in data quality in the ART clinics. The study did not, however, attempt to understand how the observed data quality related to any purposes of data use of stakeholders in the study sites. Instead, it was a measure of intrinsic data quality alone (40). Intrinsic data quality only measures the proportions of selected data quality characteristics without taking into consideration any contextual factors that may have affected the observed numerical data quality measures.

I therefore aimed to understand data quality in Malawi and the perspectives of stakeholders and how these correlates to the actual quality attributes observed in the data. From this information, I believe future interventions can be developed to continue scale up of the EHR.

1.6 The EHR Health Infrastructure in Malawi

1.6.1 About Malawi

Malawi is a landlocked country with a high population growth, which grew from 13 million, as reported in the country's 2008 census, to 17.5 million in 2018(41). The country's population is young and energetic to work towards the development of the country. The country is a biodiversity hotspot, home to Lake Malawi, the 3rd largest freshwater lake in Africa(42). The country's economic backbone is agriculture. The farmers are hardworking, but productivity is hampered due to the use of primitive farming methods and lack of access to competitive markets for their produce. The main cash crop of the country is tobacco, which is facing a decline on a global scale due to smoking cessation campaigns (43). The country has not identified a viable alternative cash crop. These factors contribute to reduced social and economic indicators with about 70% of the country living below the poverty level, which is less than $1.90 per day (44,45). Despite growing urbanization, the country's majority of the population has remained in rural areas where poverty is the highest (41). Malawi has made progress in the education sector with the introduction of free primary education (46) but has experienced stagnated growth in economic levels through poor infrastructure and limited industries to spur economic growth for the nation.

Figure 1.3 Map of Malawi

The country faced a debilitating HIV/AIDS epidemic which began in 1985 when the first HIV/AIDS case was diagnosed (47). When HIV/AIDS prevalence in pregnant women rose to 30% in 1992, the government established national strategic plans to tackle the epidemic. Actions taken resulted in the development of a national AIDS commission responsible for educating the country about HIV/AIDS and coordinating other programs to minimize transmission of the virus, such as the distribution of condoms (47). The HIV/AIDS epidemic exerted a significant burden on the health system, with the pressure created on human resources of all sectors, including the health sector (48). Additionally, to address the epidemic in its early stages, the country set up a public health response by adopting a simplified HIV/AIDS treatment approach so that many people in need of medication could access treatment (49).

Government actions have resulted in many people accessing therapy, with 80% of the population living with HIV being on treatment (50). Holistic approaches to voluntary testing and treatment have also resulted in a current HIV prevalence of 9.2% (51). Although this is still high, it represents tremendous progress, considering that the prevalence was around 15% at the beginning of the scaling of HIV treatment in 2004(52). To support efforts of fighting the HIV/AIDS scourge, the country has also received development assistance from foreign governments and non-governmental organizations, leading to further improvements in tackling HIV through the adoption of new policies such as "test and treat" (53).

The country set up a monitoring and evaluation system using a cohort analysis method and implemented via quarterly health facility supervision visits that count the number of patients that are alive and require treatment. This activity takes place in over 800 health facilities in the country (8).

The 800 health facilities are part of a public health system which is organized into three main levels:

- Primary care health levels which translate into health posts, rural health centers, and urban health centers;
- Secondary health levels that comprise community hospitals, rural hospitals, and district hospitals;
- Tertiary hospitals that comprise central hospitals.

Patients are referred up the three levels above, depending on the severity of the presenting condition (54).

The health system also has missionary hospitals that are organized by religious denominations, the largest being the Christian Health Association of Malawi (CHAM). The country has some private health facilities; however, the most significant health provider remains the government health system, which is free for users.

The monitoring and evaluation requirements increased with the growing number of patients and a recommendation on implementing an EHR (8). The EHR has been implemented in the country starting in 2007 (15). Over 750,000 patients were on treatment in these 800 health facilities at the end of 2017. 50% of whom are managed using the EHR. The EHR is thus an important part of national HIV/AIDS strategic goals such as the goal to eradicate HIV/AIDS by 2030 using the 90-90-90 strategy which Malawi adopted with the leadership of the World Health Organization (WHO) United Nations against AIDS (UNAIDS). In this initiative, identification and enrollment of HIV positive clients into ART clinics is essential for viral load suppression. The EHR benefits rely on the availability of data of good quality, which has been lacking in some LICs including Malawi.

1.6.2 Malawi EHR Infrastructure

The EHR is a web-based application that was designed and developed using open-source technologies on behalf of the Malawi MoH. The application is available online on GitHub to help diffuse the software to other parts of the region (55). The EHR also has hardware innovation that takes into consideration constraints such as limited power supply from the national grid supplier; and the low computer literacy among health care providers in the country. To address these challenges, the resultant technologies have been low power consumption technology whose devices are powered by solar batteries using direct current. Touchscreen user interfaces are used to enhance the usability of the EHR across healthcare providers who have usually not interacted with EHRs before (15).

The EHR in Malawi was implemented for scale in 2006, and each patient record is first documented in the EHR. The visit details are summarized and printed on a sticky label that is pasted on specially designed patient charts as a backup of the electronic record (15).

Driven by a desire to impact the care delivery process, the MoH insisted that the EHR developed for the country should embed guideline recommendations so that while documentation is being conducted through the EHR, reminders or alerts would appear to health care providers to improve adherence to the clinical practice guideline recommendations (15). The system has thus recognized clinical practice guideline recommendations although the knowledge has been hardcoded in the EHR (56). The national guideline recommendations in the EHR were shown to be auditable, where it was found that a performance measure could be defined for the recommendations using EHR data (57). An example of an auditable performance measure would be computing a proportion of patients that have a drug-induced side effect as a denominator with the numerator as patients that had their regimen changed reduce side effects, based on the guideline recommendation.

The observation that the EHR data are auditable means that the EHR may be used to support other interventions for patients such as meeting milestones for viral load monitoring and enforcing adherence to clinical practice guideline recommendations for other conditions. These guideline recommendations include giving counseling to patients based on how adherent they are to taking medications. This guideline recommendation implemented via the EHR is an example of primary data use via the EHR.

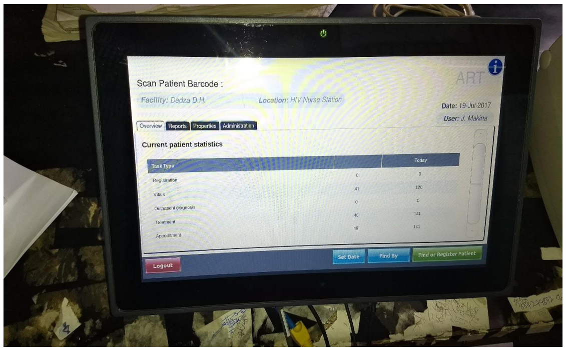
The EHR data may also be used for secondary data uses via facilitating programs such as “back to care” that trace and find patients using patient address data and bring them back to the health facility when they miss scheduled appointments. These applications fall in the country’s broader effort adopted from the WHO and the UNAIDS that focuses on identifying 90% of the population that is HIV positive, getting 90% of this population into care, retaining them under care, and making sure that 90% of this retained population has viral load suppression (58). Table 1.2 below is an example of how the EHR in Malawi maps to the 90-90-90 eradication goals:

Table 1.2 Mapping EHR modules to the 90-90-90 HIV eradication targets

90-90-90 Goal	Mapping to EHR	Description
First 90 – Identification of all people that are HIV positive	HIV testing services module	This EHR module guides healthcare providers into the processes of HIV testing and counseling following national clinical practice guideline recommendations. It assigns unique patient identifiers to all clients regardless of the HIV status after trial. Generates national reports for public health planning.
Second 90 – Enrollment of all HIV positive patients into ART clinics and maintaining them into care	ART EHR module	Assigns unique IDs to HIV positive patients that do not have unique IDs yet; Maintains the ID for those that are already positive; guides healthcare providers to manage the patients according to national clinical practice guidelines. Recommendations such as drug adherence, viral load monitoring, among others. Generates national reports for public health planning.
Third 90 – Viral load tracking and reporting	National Laboratory Information Management System	Tracks all diagnostic requests for patients enrolled in ART clinics by accepting lab orders from ART clinics and guiding laboratory technicians and managers through the testing process. Generates national reports for public health planning.
	ART EHR module	Gives alerts on viral load milestone and facilitates sample ordering via the EHR Receipt of viral load results from the National Laboratory Information Management System.

1.6.3 How is the EHR implemented and used in Malawi?

The EHR in Malawi was designed to be used at the point of care, meaning that a workstation is present at each location that a patient interacts with a health care provider. The EHR uses a touch screen input enabled computer system, similar to that of a point-of-sale device used in supermarkets. The system follows a wizard-like design to simplify navigation of guideline recommendations through the system. Figure 1.4 presents the use of an example of the clinical workstation (59). The points of care at a health facility are mapped to the patient flow with figure 1.5, representing an example of a typical workflow and various points of care where an EHR would be deployed in an ART clinic.



Scan Patient Barcode :

Facility: Dedza D.H.

Location: HIV Nurse Station

Date: 19-Jul-2017

User: J. Makina

Current patient statistics

Task Type		Today
Registration	0	0
Vitals	41	120
Outpatient diagnosis	0	0
Treatment	46	141
Appointment	46	141

Figure 1.4 Clinical Workstation Touchscreen

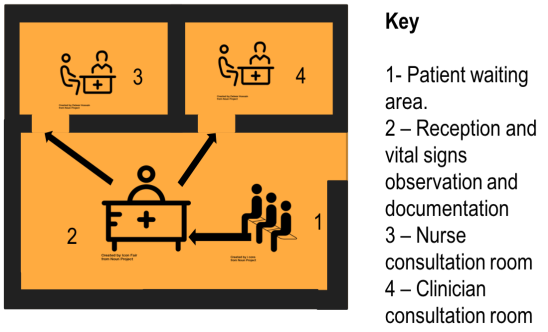


Figure 1.5 Typical ART Clinic Points of Care and Workflow

In deploying the EHR, different stakeholders, directly and indirectly, interact with it. A stakeholder is broadly defined as a person with an interest or concern in something (60). In this study, I defined stakeholders as anyone with interest with the EHR in Malawi either by direct use or indirect use of its data, including organizations that fund the implementation of the EHR. I did not include patients as part of this study due to scope and feasibility issues. I present a comprehensive list of stakeholders in Table 1.3. I also show possible data uses of the stakeholder groups with examples from the ART clinic in the same Table. These stakeholders were targeted for interviews in this study, and I provide more details on how I selected them in chapter 2.

Table 1.3 Stakeholder group roles and their functions

Stakeholder role	**Point of Care**	**Privileges/Use of Documented encounter**	**Example of DQ Challenge**
Clerk	Reception	Registration of patients for patient identification. Generating reports. Documenting vital signs	The same patient may have multiple IDs.
Nurse	Nurse room	All of the clerk + patient assessment such as adherence to medication; dispensing of prescription; referral to the clinician.	Out of range or missing, weight and height observations may affect programs that result in referral of patients for therapeutic feeding.
Clinician	Clinician room	All of the nurse + initial patient assessment (HIV stage defining conditions); prescription; diagnostic requests; treatment plan.	Wrong prescription identified by a mismatch with dispensation
MoH Supervision team	Facility/central office	Generate reports; Assess outlier patient cases, e.g., patients that changed from 2nd line treatment to first-line treatment.	Wrong quantity of drugs to procure.
MoH central office	Facility/central office	Review reports; interpret aggregate data for drug redistribution such as a number of patients alive and on treatment.	High numbers of patients flagged as defaulters/lost to care may result in under or over procurement of various essential health commodities.
Researchers	Research institution; research department;	Access EHR data in response to specific research questions that they may have.	Interpretability of the data.
Donor groups	Donor organizations within the country	Use EHR reports for their specific program goals and objectives.	The flexibility of reporting and availability of the data.

As the data in Table 1.3 suggest, different stakeholders have different documentation needs.

1.7 Defining Data Quality

Data quality is usually defined according to different characteristics (61), which lead to different definitions. Some of the frameworks in the literature include the ISO data quality framework (62), the Kahn et al., harmonized framework of secondary data use (63), the consumer network framework of defining data quality (64). Weiskopf et al. (65), also established a framework for defining electronic data quality, and lastly, the WHO (66) also has a framework for defining data quality. All of these frameworks have specific data quality characteristics that they use in describing data quality. In a LIC setting such as that of my study, the WHO framework is important because most health programs in LICs adopt WHO guidelines. I, however, observed that the WHO framework focuses on aggregate data alone (66) while my interest in this study is on patient-level data. Thus, I did not use the WHO data quality framework for my study.

The other frameworks listed above describe data quality with different terminologies as characteristics of data quality. For example, the ISO framework describes data quality using 15 data quality dimensions. These data-quality dimensions are split across two main categories of inherent data quality characteristics and system-dependent data quality characteristics (62). The consumer network data quality framework also groups data quality dimensions into main categories. Two categories are similar to the ISO framework in that they define intrinsic data quality characteristics, just like the inherent data quality characteristics in the ISO framework. The other category is similar to system-dependent data quality characteristics but describes them as contextual data quality characteristics. The consumer network data quality framework continues with two additional categories for data quality, namely representational and accessibility data quality characteristics (64). Weiskopf et al.(65) in their discussion of electronic data quality, described data quality using an earlier review on the most used data quality characteristics for EHRs. They found that completeness was the most used dimension for data quality in EHRs. Like the other previous two frameworks, they also brought into discussion data quality as being intrinsic or extrinsic, the former defining the data quality of selected variables and the latter describing any factors influencing the observed individual data quality. Another aspect when thinking of data quality in terms of the Kahn framework is the idea of verified data quality or validated data quality. Verified data refers to how well different data quality characteristics are achieved with reference to a single dataset (63). In the case of my study this would be focusing on EHR data records in ART clinics for HIV/AIDS only. Validation is another step of data quality assessment where data quality is assessed with an external data source (63). In the case of my study, I would need to get additional data sources that may be used as a benchmark of observed data had I focused on validation.

My observation was that there is an overlap in terminology across these data quality frameworks. I selected the Kahn harmonized data quality framework, which assessed several data quality frameworks, including some of the ones mentioned above because it is the most recent framework of data quality, and it has harmonized terminology from most of the frameworks reviewed above (63).

1.7.1 The Kahn Harmonized Data Quality Terminology Framework

To achieve consistency in the definition of data quality in the context of EHR, Kahn et al. introduced a data quality framework for secondary data use (63). This work was motivated by the multiplicity of frameworks and definitions regarding data quality and lack of consistent terminology, which presents barriers to comparing data quality across different settings. To address these issues, they harmonized the critical data quality dimensions or characteristics, mapping them to the dimensions of completeness, conformance, and plausibility.

Completeness means having a value that fully represents observations that occurred as expected. **Conformance,** on the other hand, refers to observed values following the prescribed format. And lastly, **plausibility** is more concerned with the credibility of observed values in terms of range and external validity. For example, having a male patient with pregnancy observation would not be plausible (63).

The Kahn framework gives examples of secondary data use as quality improvement initiatives and research purposes, usually in large research settings. Therefore, the Kahn framework for secondary data use would be appropriate in thinking about data quality in the context of my study. Although public health reporting motivated the EHRs in Malawi, the country adopted a transactional system where health care providers use the EHRs at the point of care (15,21). This adoption of transactional systems is in contrast to other regions in LICs where reporting systems have been adopted. The first example maps to the transactional use of the EHR in Malawi, where clinical practice guideline recommendations that have been incorporated in the EHR are for quality improvement through enforcing adherence to best practice in managing patients through the HIV care process.

Therefore, in the case of Malawi, there is the primary use of the data, and the secondary data uses come in as a by-product of the primary data collection by the healthcare providers.

The challenge of data quality in Malawi has been echoed in the implementation of EHRs in other countries that have deployed EHRs in ART clinics for HIV too. For example, a study at a high-volume clinic in South Africa pointed out accuracy as a challenge in documented clinical encounters (67). Accuracy is harmonized with plausibility in the Kahn framework when assessed with previous frameworks whose terminology have been mapped to the proposed terminology of the Kahn data quality framework.

Different stakeholders perform different documentation tasks giving them different end goals. This variation in goals implies the need of further understanding of the context of any observed data quality. The characteristics harmonized by the Kahn framework are firstly intrinsic to the data (40). A study that used the Kahn framework in six data sharing networks to conduct a data quality assessment recognized the shortfall of focusing on intrinsic data quality alone and adopted the view that data quality should be fit for purpose (40). In this study different data quality characteristics were attached to different stakeholders in order to understand the context of their data quality assessment (40). The Kahn framework also states the need to adopt this view of data quality "being fit for purpose" (63).

The definition that data quality is "fit for purpose" recognizes that different stakeholders who use EHR data will have different purposes of data use. For example, in the case of audit and feedback to improve adherence to clinical practice guidelines, the expectation is that data should be complete and accurate for these purposes (38). In the case of audit and feedback, if a performance measure on prescribing the right drug is to be calculated, healthcare providers may need to know several aspects of a patient's health before prescribing a drug. One piece of important information to a clinician may be whether a patient is allergic to sulfa, which would prevent the prescription of sulfa drugs. On the other hand, public health officials may only be interested in knowing what prescriptions have been dispensed, which would be important in deciding which drugs to order.

In the case of Malawi, there are public stakeholders who use secondary data for purposes such as procurement of medical supplies and design of public health programs. On the other hand, healthcare providers such as nurses and clinicians have primary purposes of data use from the EHR on direct patient care decisions that are dependent on collected EHR data such as the drug regimen that a patient is on and the quantity that was dispensed. This data would be useful for calculating patient adherence to taking their drugs according to the prescribed dosages. With reported EHR data quality challenges, I therefore used the view of data quality being fit for purpose to explore EHR data quality in ART clinics in Malawi.

1.7.2 Understanding Malawi EHR Data Quality "fit for purpose"

Data quality remains a problem in ART clinics using EHRs in Malawi (50). There have also not been studies that have reported any new approaches to understanding data quality beyond the intrinsic data quality in this setting. I believe this approach does not provide a comprehensive understanding of data quality. Additionally, with the reported poor data quality in Malawi, any interventions to improve data quality that would not take purposes of data use into the context of data quality assessment would fall short of improving data quality. Purposes of electronic data use have been recognized by the WHO in a classification of digital health interventions that have been rolled out in LMICs for different stakeholder groups (67). These role functions have not been confirmed as purposes of data use in the context of data quality assessment. I therefore used this study to identify relevant and significant purposes of data use in ART clinics in Malawi with reference to the electronic data purposes of use in the WHO classification of digital health interventions. Given the absence of a systematic assessment of ART data quality in ART clinics for HIV in Malawi, one that takes into consideration purposes of data use of stakeholders, I asked the question below, with consideration that stakeholders continue to report poor EHR data quality: ***How do stakeholders' varying purposes of EHR data use inform interventions for EHR data quality improvement?***

AIMS

The aims of my research were conducted within the context of the 90-90-90 efforts of eradicating HIV/AIDS in Malawi by 2030:

1. To identify the importance of data quality characteristics in the context of EHR stakeholders' purposes of data use.
    a. To identify purposes of data use for different stakeholders using the EHR data.
    b. To identify characteristics of data quality that stakeholders using the EHR data expect.

    c. To rank the importance of data quality characteristics in accordance with purposes of data use across different stakeholders using the EHR data.

2. To conduct a “fit for purpose” data quality assessment and evaluate the drivers of the observed data quality.
    a. To conduct a data quality assessment of EHR data based on Kahn framework data quality characteristics and purposes of use in ART clinics in Malawi

Chapter 2 Research Methods

2.1 Chapter Overview

In this chapter, I present methods for the research question and aims presented at the end of chapter 1 and restated below:

1. To identify the importance of data quality characteristics in the context of EHR stakeholders' purposes of data use.
    a. To identify purposes of data use for different stakeholders using the EHR and EHR data.
    b. To identify characteristics of data quality that stakeholders using the EHR and EHR data expect.
    c. To rank the importance of data quality characteristics by purposes of data use across different stakeholders using the EHR and EHR data.
2. To conduct a data quality assessment and evaluate the drivers of the observed data quality.
    a. To perform a data quality assessment of EHR data based on Kahn framework data quality characteristics and purposes of use in ART clinics in Malawi

I conducted a mixed methods study to address the aims above. I present the methods for each aim with its sub-aims, starting with the study population, sampling, data collection, and the data analysis approaches. The aims above link to key documents and frameworks, introduced at the end of Chapter 1, including: the Kahn data quality framework, the WHO Classification of Digital Health Interventions, and a practical application for assessing data quality, namely the 90-90-90 targets for HIV/AIDS eradication. I mapped these documents to inform the studies I conducted to address the aims as illustrated in the figures 1 below for my qualitative studies under Aim 1:

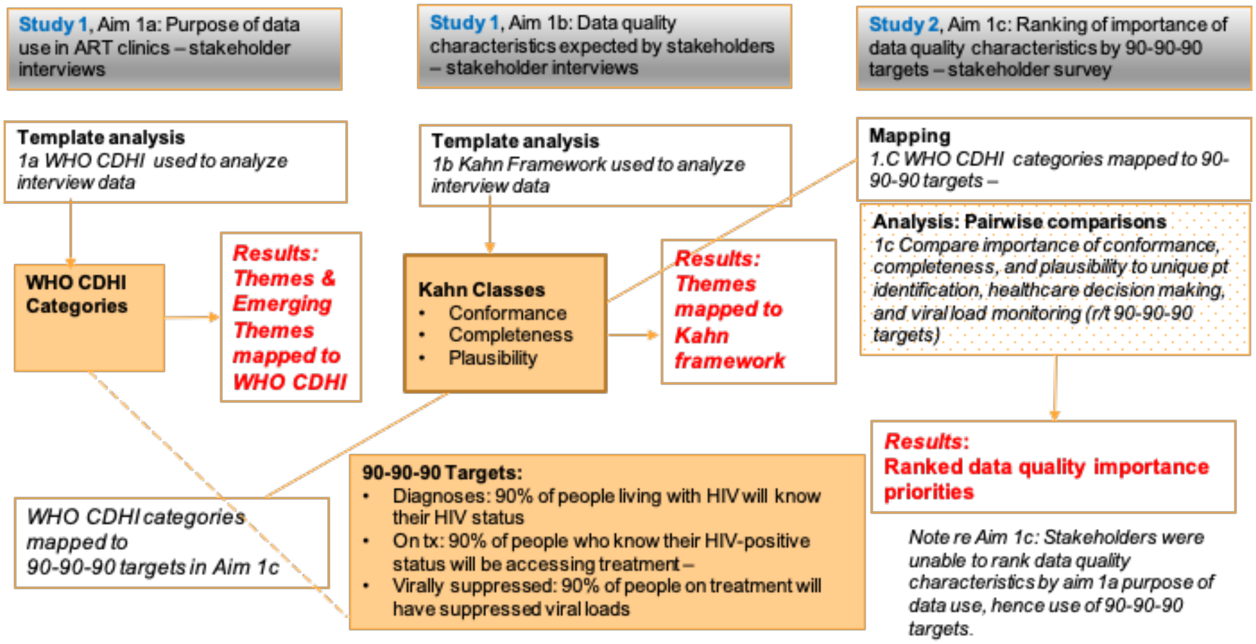


Figure 2.1: Schematic of mapping research methods and study results for Studies 1 and 2

I discuss the detailed research methods represented in Figure 2.1 in sections 2.6.1 and 2.6.2. for study 1, and in section 2.6.3 for study 2. The results are detailed in chapter 3. I illustrate the linkages of the different documents and frameworks for study 3 in Figure 2.2.

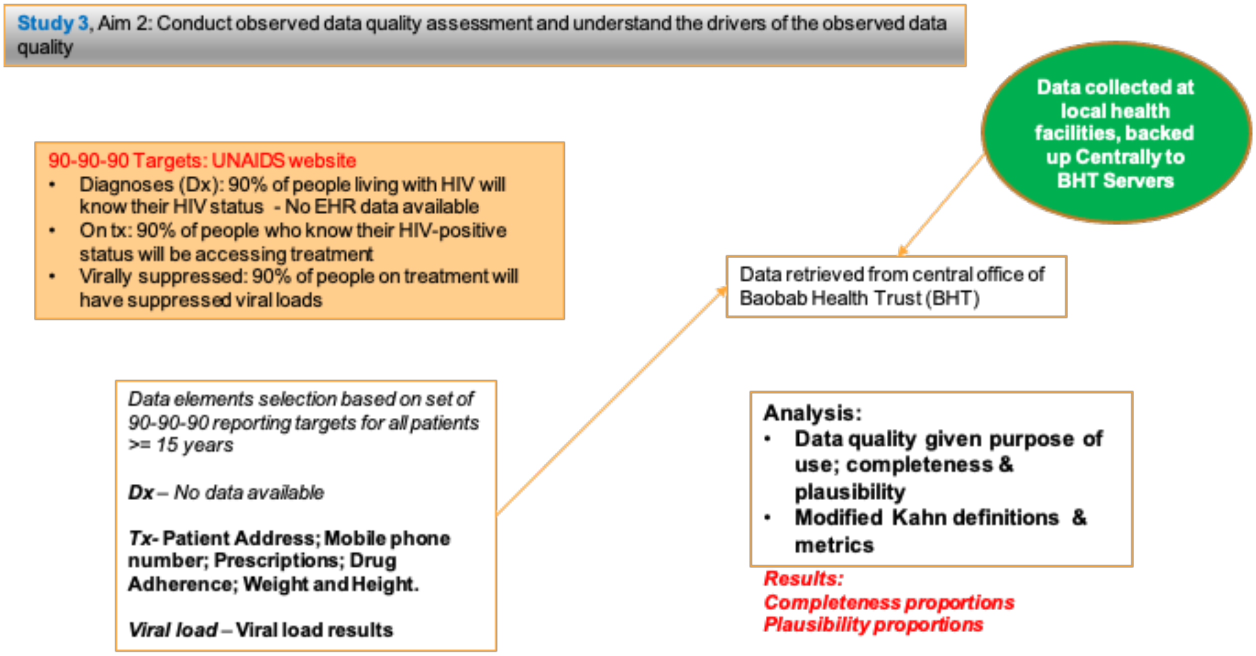


Figure 2.2: Schematic for research methods and results for study 3

I discuss the detailed research methods in Figure 2.2 in section 2.6.4.

2.2 Overview of Methods

I used mixed methods with the belief that measuring data quality using quantitative methods alone would be insufficient because it would miss other factors that may influence data quality. These factors could be different purposes of data use by stakeholders selected in this study. For instance, in a review by Hong et al., they observed that quantitative data quality assessment alone was insufficient as other factors may explain observed data quality after conducting a qualitative assessment. Many studies followed up the quantitative studies with qualitative research to understand or explain the results of the quantitative data quality assessment (61).

I purposively selected health facilities at three different levels of care for all the aims to achieve site representativeness. I selected the list of health facilities from a list of health facilities that use the EHR. After I purposively selected a list of health facilities in each health zone, I shared it with the EHR implementer, Baobab Health Trust, and the MoH to get their approval.

Healthcare levels vary in terms of infrastructure and resources such as human capacity as well as the approved medication list and services rendered (54). The sites were purposively selected to meet geographical and health level representation.

Data collected at different healthcare levels may differ in data quality due to these resource differences. I made this distinction based on data quality studies similar to mine, where they made site selections based on different health facility levels and resources (37). I used the target site population in Table 2.1 below to address both Aims 1 and 2.

Table 2.1: Target Health facilities with EHR for Aims 1 and 2

Health Zone	Health Facility Type	Health Facility	Health Level
South East	District Hospital	Mangochi District Hospital	Secondary
	Central Hospital	Zomba Central Hospital	Tertiary
South West	Health Center	Limbe Health Centre	Primary
	Central Hospital	Queen Elizabeth Central Hospital	Tertiary
Central West	District Hospital	Salima District Hospital	Secondary
Central East	Health Center	Kawale Health Centre	Tertiary

	District Hospital	Dedza District Hospital	Secondary
	Central Hospital	Lighthouse Hospital	Tertiary
Northern	Health Center	Mzuzu Health Center	Primary
	District Hospital	Nkhatabay District Hospital	Secondary
	Central Hospital	Mzuzu Central Hospital	Tertiary

2.3 Research design

This research was an observational study design using mixed methods to address my research aims. I followed a guideline titled "Reporting mixed research in the field of counselling" from Enhancing the Quality and Transparency of Health Research (EQUATOR) web site (68,69). The guideline recommends that researchers should choose a framework which guides their research design (70).

I collected retrospective quantitative data for the period of January 1st, 2017 to December 31st, 2018. I targeted this period based on an assumption that there would not be significant changes in the EHR data since the most recent guideline that was in use was implemented in 2016. My study followed a convergent framework for conducting a mixed-methods study design (69,71). A schematic adapted from Steinmetz-Wood et al. (69) illustrates the method I followed in doing my mixed-methods study design (Figure 2.3).

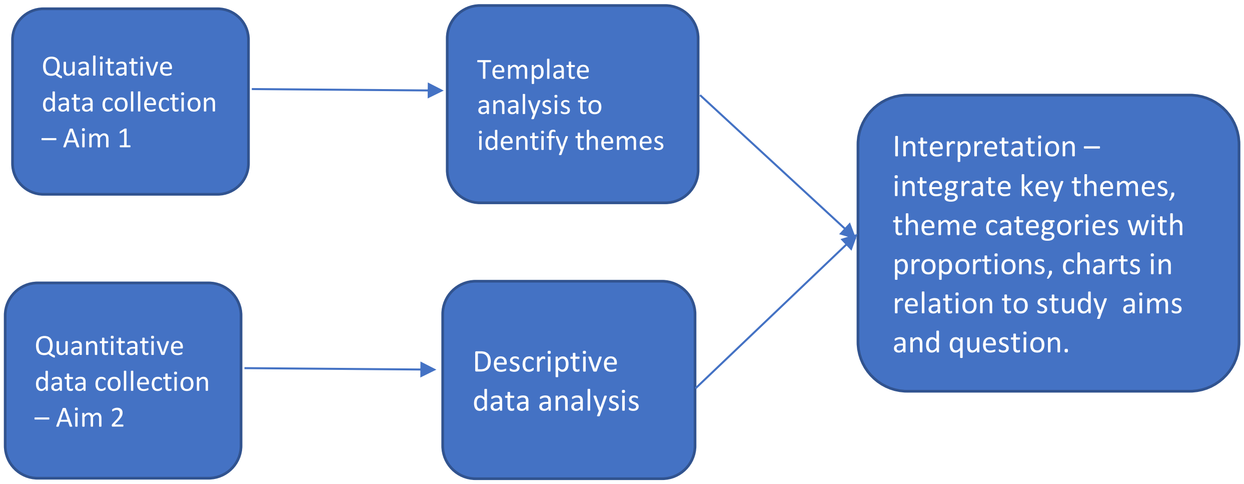


Figure 2.3 Mixed Methods Design using a Convergent Design

2.4 Mixed Method Study Design Integration Approach

Theme by statistics or theme side by side comparisons are two ways of displaying the integration of convergent mixed study results (72). I used side by side theme comparison for Aim 1 and theme by statistics for integrating Aims 1 and 2. Side by side theme comparison presents qualitative themes for each objective side by side in a table, whilst theme by statistics adds quantitative results to the qualitative themes identified. Figure 2.4, adapted from Wood et al. (69), illustrates how I conducted integration between Aims 1 and 2.

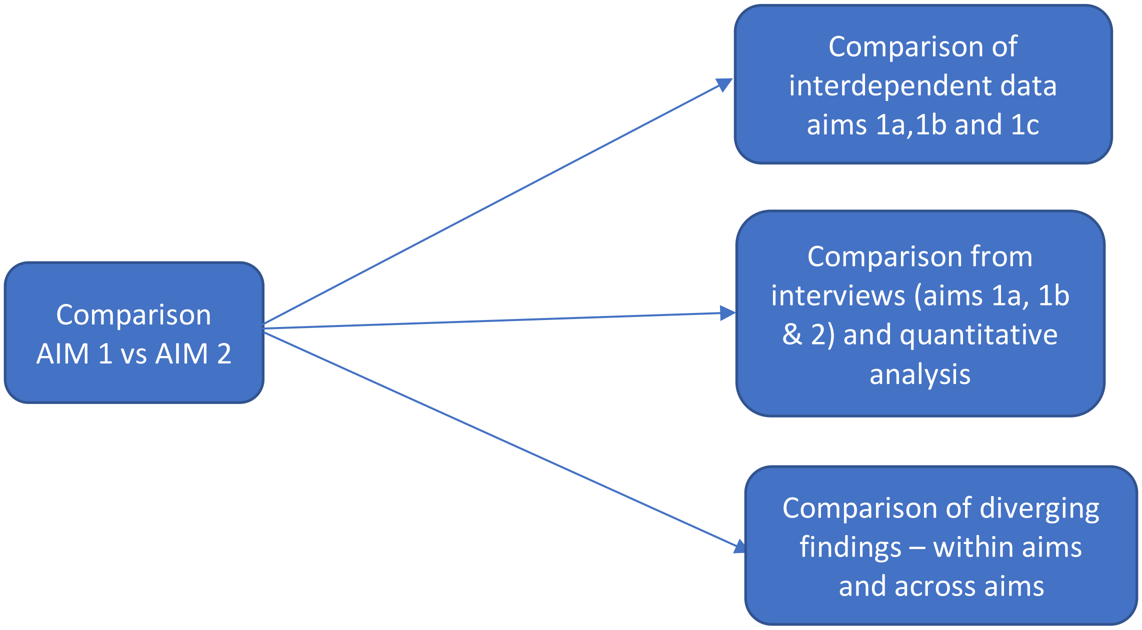


Figure 2.4 Integration approach for different aims in the study adapted from Wood et al.

I presented the integration of both Aims 1 and Aim 2 using tables which were a buildup of the findings under each aim.

2.5 Overview of Study Measures

Measurement is the process of assigning a value corresponding to the presence, absence, or degree of a specific attribute in a specific object (73). I considered the purposes of data use as an object with the WHO digital classification of digital health interventions as attributes for my first study. Additionally, I measured data quality as an object with the data quality characteristics as the attributes for my second and third studies. Measurement requires defining or using existing instruments to ensure that the attribute of the objects that are intended for measurement are precisely captured. In the case of my study, I used instruments defined by the researcher and informed by experts and other existing frameworks. The expertise for validating the researcher defined instruments such as interview guides was provided by my research committee. Additionally, I used the Malawi HIV/AIDS treatment guideline recommendations to define database queries for responding to Aim 2. I present my measurement approaches in Table 2.2 below.

Table 2.2 Summary of Measurement Approaches for Study Aims

Aim	Sample	Concepts & their definition	Variables & their definitions	Measures of variables, or methods for calculating derived variables	Analysis Plan
1a	**34[1] participants:** 10 Nurses 5 Clinicians 3 Public Health officials 3 Health center managers. 2 ART zonal coordinators 9 Data services 2 Donors	Purposes of data use - Stakeholders' EHR data use.		Researcher defined Semi Structured interview guide - Open ended questions: Do you use data in your job? - What/how do you use data in your job?	Thematic analysis and Template analysis using the WHO classification of digital health interventions
1b	**34 participants:** 10 Nurses 5 Clinicians	Anticipated stakeholder data quality characteristics - Data quality dimensions/chara		Researcher defined Semi structured interview guide Open ended question:	Thematic analysis and Template analysis using

---

[1] I used these 34 participants for both Aims 1a and 1b.

	3 Public Health officials 3 Health center managers. 2 ART zonal coordinators 9 Data services 2 Donors	cteristics expected in the data accessed.		- What data quality characteristics do you anticipate when using data for your job? - Of the data quality characteristics, which ones matter most?	the WHO classification of digital health interventions
1c	**11[2] participants:** 3 data services 2 nurses, 1 clinician, 1 health center manager 2 donors 2 public health officials.	Stakeholder Importance priority levels for data quality characteristics - Level of importance attached to different data quality characteristics for selected purposes of data use in the context of the 90-90-90 targets for eradicating HIV/AIDS by 2030.		Researcher designed questionnaire. Importance priority levels of data quality characteristics across selected 90-90-90 purposes of data use. (Patient identification; Patient retention; Patient HIV viral load monitoring) - Likert scale (Very important; Important, Moderately Important, Slightly important, Not important) style of question:	Pairwise comparisons between data quality characteristics for each purpose of data use and counts of the more highly ranked characteristic in each pairwise comparison using Frequency tables
2a	Over 160,000 patient records and 549,000 patient visits from 10 health facilities for 5 variables. Over 37,000 records for viral load	Completeness, Atemporal Plausibility and temporal plausibility. - Completeness: presence or absence of data values regardless of the structure of the value present representing a variable.	Patient physical address: Current place of residence (Three parts: District, Traditional Authority and Village. Body Weight Body Height	Database queries using[3] MySQL and later counts of numerator and denominators for different data quality characteristics based on the Kahn Framework for each data quality characteristic using a computer program in R https://github.com/Jintha/dqa_malawi.	Proportions for each data quality characteristic. Charting for selected variables.

---

[2] These 11 participants for Aim 1c, were a subset of the 34 participants from Aims 1a and 1b.

[3] Queries and measures checked for external and internal validity with Software Development Manager at Baobab Health Trust, the implementer of the EHR. The measures were also checked by the Director of Monitoring and Evaluation of the Lighthouse Trust for validity.

records at 3 tertiary health facilities for viral load results.	- Atemporal plausibility: believability of an observed value in terms of accepted guideline recommended range checks - Temporal plausibility: believability of an observed value in terms of the precedence of recorded events represented.	Drug adherence: Percentage of drugs taken as recorded. Drug prescription: Quantity of drugs dispensed as recorded Viral load results: Viral load results for each patient at three tertiary (referral) level hospitals.	Completeness = Total Number of Records with a recorded value / Total number of records in line with that variable. Plausibility: Measured Temporal Plausibility and Atemporal Plausibility Temporal plausibility = Total number of records falling after a patient is enrolled in the ART clinic / All records within the study period for that particular variable. Atemporal plausibility = Total number of records that meet the range definitions according to the Malawi treatment guidelines for HIV/AIDS or reference tables / All records within the study period for that particular variable.

2.6 Methods

2.6.1 Aim 1 Study Sample

The target sample included stakeholders such as healthcare providers, public health officials in the MoH, and selected donors of EHR investments in the country. I used this target sample for all sub aims under Aim 1 since I felt this target sample would identify individuals with different data needs based on the key stakeholders presented in Table 2.3. Healthcare providers were targeted due to their interaction with the EHR at the point of care while offering various services in the ART clinics for HIV. Their interaction with the EHR as a tool for documenting the services they provide to patients prompted me to target them since I believed have probably experienced effects of poor data quality on their use of the EHR. I targeted other stakeholders because they use by-products of the point of care data collection in their day-to-day activities. These activities may be the design of public health interventions and the procurement of commodities such as drugs. I present the stakeholders targeted in this study in Table 2.3 below:

Table 2.3 Health Area and Stakeholders Targeted for Interviews

Health Area	Stakeholders	Sample
Patient care	Health facility healthcare providers (nurses, clinicians) Data services (Data clerks, Health management information systems officers)	The target population of: 27 based on health facilities in selected health zones at different health levels.
Supervisory Staff	MoH Supervision team or Health Facility managers in Zonal offices	Target population of: 5 based on the total number of health zones in the country.
Donor representatives	Project officers of Donor funding agencies that fund organizations that implement the EHR.	Target population of: 3 based on the main donors of EHR investments in the country.
Public Health Policy	Program/section directors; National Monitoring and Evaluation (M & E) coordinator; Drafters of Clinical Practice Guidelines;	Target population: 2 based on the MoH officials that directly use the outputs of the EHR.

2.6. 2 Methods: AIM 1a and AIM 1b

2.6.2.1 Sampling: AIM 1a and AIM 1b

I used purposive sampling to select the study participants for Aim 1 based on the criteria that stakeholders use the EHR and the EHR data either at ART clinics while seeing patients or in other capacities such as donor programs or HIV/AIDS public health needs. Since this was a convenience sample, the results would only be applicable to similar group of participants.

Site sampling

To take into consideration the site representativeness of responses of healthcare providers, I interviewed participants from different healthcare levels, such as primary, secondary, and tertiary healthcare levels. These healthcare levels attend to different patient catchment populations, increasing across each level, as stated in section 2.2 above (54). Other stakeholders were in central monitoring offices either for the MoH, EHR donor organizations and zonal public health offices.

Study participant stakeholders

Study participants for Aim 1 were also purposively selected based on their direct and indirect interaction with the EHR in their capacities as healthcare providers, data service providers and management services, supervisors, donor representatives and public health policy officials of the EHR implementation in the country.

Health care providers and data services

Each health facility has roles of healthcare providers such as data entry clerks, data officers, nurses, and clinical officers (74), medical assistants, and medical doctors.

Clinical officers or clinicians are trained to perform general medical and surgical procedures such as cesarean sections and are required to complete an internship program before they may start practicing. Nurses may also become clinical officers, but only in the specialty that they train in. For example, if they train in anesthesia, then they are referred as clinical officers only in that area unlike the general training of clinical officers. These cadres exist in Malawi to fill the gap of the low number of medical doctors (74).

Medical assistants get only two years of clinical medicine training and are not required to fulfill internship requirements (74).

I used purposive sampling to select at least one of these roles for Aims 1a and 1b. Specifically, I attempted to have at least 2 participants from each role so that I could validate individual responses. After interviewing the first study participant in these roles, I used snowballing techniques to identify additional stakeholders in the health areas of patient care and data services (75). At the health facilities, I asked for a name of an individual to interview next based on a role that I had not interviewed yet at that facility. So, for example, if I had already interviewed a nurse, I would ask for a clinician that uses the EHR in the ART clinic and is available for an interview.

Supervisory staff

Supervisory staff at the health facilities were healthcare providers who lead different care and administrative activities. These were the point of contact at each health facility. I selected the first participant for an interview with the guidance of the clinic manager with whom I met to give an overview of my study and get permission to conduct interviews.

Public health officials and Donor representatives

For other stakeholders such as public health officials working in government offices and non-profit organizations that also use the EHR, I made a list of potential interviewees and contacted them to set up the interviews. I compiled the list of stakeholders after meeting one key EHR stakeholder at a center of excellence for HIV/AIDS treatment. I selected this individual based on his key role in using EHR HIV/AIDS data at the HIV/AIDS treatment center of excellence. He then made suggestions of names of people to be considered for interviews. I then sent e-mails to this list of stakeholders to book an appointment for an interview. Afterwards, I conducted interviews with the stakeholders that had responded to be interviewed, whilst following up on those that had not responded yet. To identify additional stakeholders in this stakeholder group, I asked the interviewee to recommend other possible participants.

I identified donor representatives of the EHR by first discussing with the implementer of the EHR on who their active donor funding agencies were. I then scheduled meetings with the donor representatives to conduct my interviews.

2.6.2.2 Data Collection Aims 1a and 1b

I used the semi-structured interview guide in Appendix 1 to address Aims 1a and 1b. I developed the semi-structured interview guide with questions from literature on understanding the uses of data and anticipated data quality characteristics (63,67). To ensure further validity of my data collection tool, an expert in my research committee reviewed the questions for Aims 1a and 1b. I interviewed study participants at their workplace, in a room with the door closed or a quiet open space away from other individuals where a private room was not available. I recorded the interviews on a voice recorder after reading out the overview of the study to the study participants as part of seeking informed consent. Interviewees signed informed consent documents as evidence of their consent.

2.6.2.3 Data Analysis: AIMS 1a and 1b

Since this was an observational study design using mixed methods for data collection and analysis, I conducted a thematic analysis to identify common themes on purposes of data use and the data quality characteristics that they anticipated in the data for Aim 1. I followed the steps below for the qualitative data analysis:

Thematic Analysis Steps

1) Coding: I first gained familiarity with the data through transcribing the interviews. After gaining familiarity with the data, I conducted a template analysis based on the World Health Organization (WHO) classification of digital health interventions (CDHI) and the Kahn data quality framework for secondary data use and (63,67). I did this for Aims 1a and 1b respectively. The WHO CDHI is a list of digital health interventions mapped to broad areas of stakeholder groups and their purposes of use. After transcribing, I searched for words or phrases that were closest to the defined codes and mapped them to the codes in NVivo(76) for both Aims 1a and 1b. I have included the codebooks used for this stage as appendices 2B and 2C. I used the qualitative data management analysis software NVivo(76) to manage the data by importing the transcribed interviews, defining a codebook in the software, and then assigning the responses of the interviews to relevant codes in my codebook. I used the WHO CDHI to develop a codebook for defining purposes of data use.
2) Identifying themes: Keeping the research question in mind, I continued to identify emerging themes from the data. The emerging themes were explored by identifying phrases that linked to specific codes.
3) Recoding: I then reviewed the transcripts to assess for the presence new codes. I proceeded as the sole coder by revisiting the themes that I previously identified. To add

objectivity to the coding, I took breaks of up to two weeks before starting to recode the transcripts so that I bracketed myself from the earlier coding that I performed (75).

4) Theme significance: I ranked the themes based on emphasis and frequency of repetition of emerging concepts in the themes (75,77). Additionally, I validated the important themes with the WHO CDHI (67) in the case of purposes of data use. I used the Kahn harmonized data quality framework for secondary data use (63) to code expected data quality characteristics. The WHO CDHI is the most recent publication on digital health interventions mapped to stakeholder purposes of data use. The Kahn framework harmonizes data quality characteristics from many other frameworks and therefore was relevant for my coding approach.
5) Report writing. I then produced a report based on the themes that I identified.

I summarize further my thematic analysis process described above in Table 2.4.

Table 2.4 Thematic Analysis Table

1. Activity	2. Output	3. Follow Up Activities	4. Output
Transcription	Transcripts	Re-listen audios	Final transcripts
Type and translate any audio interviews			
Coding by Objective	Initial codes by data use and stakeholder group classification	Secondary coding	Final codes by data use and stakeholder group classification from the WHO CDHI
To establish the ways that data are used or would like to be used by different stakeholders working to manage HIV/AIDS patients and are using or expect to use an EHR			
To identify characteristics of data quality that would help stakeholders determine whether the data are fit for their purpose/fit for their intended use of the data.	Initial codes by data use and cadre	Secondary coding	Final codes by data use and stakeholder group classification mapped to the Kahn Data Quality Framework
To determine the relative importance of the characteristics of data quality characteristics expected based on the Kahn framework.	Initial codes by data use and cadre	Secondary coding	Final codes by data use and stakeholder group importance from the Kahn Data Quality Framework.
Emerging Themes	Initial themes per code	Theme significance	Final codes: 1) Purposes of data use mapped to the WHO CHDI 2) Expected data quality characteristics mapped to the
How do purposes of EHR data use of stakeholders inform interventions for EHR data quality improvement?			

			Kahn data quality framework. 3)Importance of the Kahn data quality framework characteristics in line with the 90-90-90

2.6.3 Methods: Aim 1c

2.6.3.1 Study population Aim 1c

I targeted a subset of the stakeholders from Aims 1a and b for this Aim. I used convenience sampling to select the participants for AIM 1c based on who was readily available. I identified the participants from the same list of participants as Aims 1a and 1b. The number of participants for this Aim were guided by geographical access and the representation of stakeholder types. To understand the relative importance of data quality characteristics, I used an interviewer-administered questionnaire provided in Appendix 4. I asked stakeholders to rank the importance of data quality characteristics concerning selected data uses mapped to the 90-90-90 HIV eradication strategy. The selected purposes of use were further mapped to examples of digital health data uses as classified in the WHO CDHI. This was done to enable integration between anticipated purposes of data use from stakeholders and practical use of these purposes with the example of the 90-90-90. The 90-90-90 HIV eradication strategy was chosen because they are an essential global strategy that has been adopted by Malawi (58).

2.6.3.2 Sampling Aim 1c

Public health officials such as zonal quality control managers, donors, national monitoring and evaluation experts as well as healthcare providers at ART clinics from the northern zone were selected for this Aim. I chose the participants as a continuation of the interview process for Aims 1a, and 1b. After completing questions for Aims 1a and b, I would proceed to ask questions for Aim 1c. I administered the questionnaire to the study participants and recorded the responses they gave me. I interviewed 11 participants only for this Aim 1c because the interview questionnaire was not ready at the time the participants of Aims 1a and 1b were being interviewed. I thus could only feasibly access the 11 participants that were conveniently located to participate in Aim 1c.

2.6.3.3 Data Collection Aim 1c

I used data quality characteristics harmonized in the Kahn data quality framework for secondary data (63). The data quality characteristics that were ranked are conformance, completeness and plausibility. I mapped these data quality characteristics along with purposes of data use as classified by the WHO (67). Table 2.5 below presents how I mapped the WHO classification of digital health intervention data uses to the 90-90-90 goals for HIV eradication.

Table 2.5 WHO Data Uses Mapped to 90-90-90 Goals

90-90-90 Example Requirement from EHR	**WHO CDHI data use mapping**	**Question for establishing the importance of characteristics.**
Identification of HIV positive patients The first 90 is a goal to identify all people that are HIV positive and get them enrolled in care. This may be possible with unique patient identifiers and other patient demographic information.	Supply chain management (Patient identifier will help in calculating number of patients to receive drugs) Location mapping (Patient identifier will help in identifying clients' place of residence) Client health records (Patient identifier is used to pull individual patient records)	Given unique patient identification, which is an example of client health records data use, compare importance of the two characteristics below (pairwise comparison of completeness, plausibility and conformance)
Retention into care with adherence to Clinical Practice Guideline Recommendations (BMI assessment, adherence, prescription management) Once patients are enrolled in care, the appropriate care should be given to patients by following clinical practice guideline recommendations. This recommendation is dependent on transactional data at the point of care. It may also be dependent on the availability of essential commodities such as the actual medications.	Healthcare provider decision support (Thresholds of BMI or Adherence trigger decision support to healthcare providers) Prescription and medication management (Dosing information calculated via the EHR) Healthcare provider communication (Data entered at one point of care available for use at another)	Given patient retention into care which may have applications towards healthcare provider decision support, prescription and medication management and healthcare provider communication compare the importance of the two characteristics below (pairwise comparison of completeness, plausibility and conformance)
Viral load suppression at required milestones To prevent transmission to other people, HIV positive people need to achieve non-detection of their viral load. This goal requires viral load monitoring at the correct milestones so that 90% of the population reaches the viral load suppression.	Laboratory and diagnostics imaging management (Viral load results saved in the ART EHR)	Given viral load suppression which is an example of laboratory and diagnostic imaging data use, compare the importance of the two characteristics below: (Pairwise comparison of Plausibility, Completeness and Conformance)

I used the purposes of EHR data use above as examples in mapping the importance of data quality characteristics across selected stakeholders. I have included the interviewer-administered questionnaire for this Aim, as Appendix 4.

I interviewed study participants using the data collection tool in Appendix 4 and entered the responses on a paper questionnaire. I later entered the interviewees' responses onto an online form designed using Google forms. I used a paired psychometric method (78) when administering the questionnaire to deal with the challenges of extreme response syndrome when respondents may reply to every characteristic as being very important or not important at all. In this approach, I asked participants to compare the importance of data quality characteristics relative to one another.

2.5.3.4 Data Analysis Aim 1c

I used frequency tables to count the number of times a data quality characteristic was ranked higher in importance compared to another data quality characteristic for a given purpose of use. I calculated the number of possible pairwise comparison combinations of data quality characteristics for a selected purpose of data use with the formula below (78):

$$possible\ number\ of\ pairwise\ combinations = \left(\frac{c(c-1)}{2}\right)$$

The letter "c" denotes the total number of candidates or options that may be chosen from. Data quality characteristics were the candidates in the case of my study. Therefore, the possible combinations for comparison were:

$$(3(3-1)/2) = 3$$

I asked stakeholders to rank importance of the data quality characteristics of plausibility, completeness and conformance across 3 selected purposes of data use from the 90-90-90 that I mapped to the WHO CDHI. The 3 possible pairwise of data quality characteristics comparisons were:

Plausibility vs Completeness
Plausibility vs Conformance
Conformance vs Completeness

I asked stakeholders to rank the highest level of importance between the data quality pairs above for a purpose of data use, for example viral load monitoring data use, in order to determine a data quality characteristic with the highest priority of importance. I further provided definitions for each data quality characteristic with an example for them to understand the data quality characteristic.

For each purpose of use, I assigned a score of 1 for each data quality characteristic of the pair that was ranked higher, and 0 for each characteristic that was ranked lower in each pair. I assigned 0.5 of a point in the event that stakeholders attached the same importance to a pair of data quality characteristic for a selected purpose of data use (79). I then, summed total votes for each characteristic across all participants for each purpose of use. Table 2.6 below shows how I assigned the votes.

Table 2.6 Sample Table of Priority Summations Across Study Participants

Comparison of Characteristics	V-1	V-2	V-3	V-4	V-5	V-6	V-7	V-8	V-9	V-10	V-11	**Sum of Votes**
Purpose of use example: Unique patient identification												
Plausibility												
Completeness												
Plausibility												
Conformance												
Conformance												
Completeness												

I then arrived at the importance for a given data quality characteristic by comparing the sum of data votes across the different data quality pairwise comparisons. The data quality characteristic with the most votes had the highest priority. This process was repeated across the different purposes of data use for the 90-90-90 HIV/AIDS eradication.

2.6.4 Methods: Aim 2a

2.6.4.1 Study Population: Aim 2a

For the quantitative part of this study, adult patient EHRs were the target population. Adults are defined as all patients above 14 years of age by the Malawi MoH (54,80). I again selected the variables on which to perform a data quality assessment concerning the HIV/AIDS eradication goals stated in earlier sections (81). My decision to select variables was informed by identifying variables that are useful to meet the strategic goal of the 90-90-90 targets for HIV/AIDS eradication. I targeted the variables below from the EHR as per Table 2.7 below:

Table 2.7 90-90-90 Goals and EHR Variables selected for the study.

90-90-90 data use	Variable
Patient tracing	Patient Physical Address – the current place of residence Patient phone number – mobile phone number for patients.
Quality of patient care	Prescribed and dispensed drugs Quantity of ARV drugs and Cotrimoxazole Prophylactic Therapy (CPT) drugs dispensed to patients Patients' drug adherence – The presence of adherence calculations and whether they are within range. Body mass index – weight and height observations.
Viral load suppression	Viral load results present as a proxy for viral load orders. Date viral measured in comparison to starting of the Viral load program. Interval of viral load testing.

The data uses above represent examples of data that are collected electronically and are directly used to respond to the HIV 2^{nd} and 3^{rd} program goal of eradicating HIV by 2030 (90-90-90) by achieving these targets by 2020(58). I selected the patient variables that facilitate patient tracing because they ensured that patients remained on treatment in the event that they missed their hospital appointments and would need to be followed up at their homes using their address or phone number. I further selected some variables that may be associated with the quality of care that patients receive at hospitals, such as the drugs prescribed or continuous monitoring of vital signs such as body mass index because good care also determines retention of patients into the program. The last variable of viral load results would enable monitoring of the viral load, hence the I selected this as well.

2.6.4.2 Sampling: Aim 2a

Patient study population

I purposively sampled adult patient (older than 14 years) records of HIV positive patients at 3 different healthcare levels in the country so that there is site representation. Adults are all patients that are above 14 years in the Malawi health system (80). These were patient records of all patients that were alive during the study period. I did not exclude records of patients who died during the study period. In addition to the healthcare level, the health zone (54,82), in which the health facility falls, was also taken into consideration when selecting the health facilities.

Site sampling

I selected all health facilities targeted in this study as presented in Table 2.1. At the time of this study, the EHR had been deployed in about 193 health facilities out of over 700 health facilities in the country (83). Although this represented site coverage of only 28%, these were mainly high-volume ART health facilities, and they cumulatively managed over 50% (~370,000) of all patients alive and on treatment in the country (~780,000) (50). I used purposive site sampling to select health facilities for this study. I selected at one health facility at each health level in each of the 5 health zones in Malawi.

Since the patient records at the health facilities are in electronic format, I selected all records that fell between January 2017 to December 2018. I chose this period because I believed that there had not been any significant changes with the EHR system within my research period with regards to clinical practice guideline recommendations (80). I found that there were some changes that happened to the EHR during the period when I was conducting my electronic data collection. Specifically, I found that the EHR viral load order record function was not updated in the EHR. Additionally, I found that the viral load results were available in only three of my 10 target health facilities. To address the first change, I opted to use viral load results as a proxy for assessing the data quality of viral load order requests. Secondly, I modified my site population to target the three health facilities that had viral load results data. These three health facilities were Queen Elizabeth Central Hospital, Zomba Central Hospital and the Lighthouse Trust at the Kamuzu Central Hospital.

2.6.2.3 Data Collection: Aim 2a

Variables were collected in line with the 90-90-90 targets for data quality assessment in line with the Kahn data quality characteristics of completeness and plausibility. I did not focus on conformance because the process of achieving this measure was beyond the scope of my work as I would need to do more preliminary work to understand the database structure of the EHR. I have summarized these variables in Table 2.7 of this chapter. I dropped the unique patient identifier for the first 90 because I discovered that this variable was in a new EHR module that was not available in most of the targeted health facilities, hence the data not being available. The first 90 has also been observed as a challenge to measure at the national scale (51). The three health facility data that I used for the final 90 assessment were central hospitals that maintained the same organization and were complete.

I used the MySQL database management system in writing the database queries to extract these data and exported the data to a comma-separated value (CSV) format (84). Two developers from Baobab Health Trust, the organization that designed and implemented the EHR in Malawi, checked and edited as necessary, and ran queries on the site databases (85). They provided me with the anonymized datasets by uploading them to a secure Dropbox drive. I then transferred these datasets to a secure drive. Some datasets were corrupted with transposed column values and required regenerating. After I reported this to the developers, they generated the new datasets and re-uploaded them. Again, I immediately transferred the datasets to a secure Box location for further data analysis.

2.6.4.4 Data Analysis: Aim 2a

I used the Kahn harmonized data quality framework for secondary data use to operationalize the definitions for my data quality assessment in relation to the 90-90-90 HIV eradication targets. The Kahn framework helped to come up with broad classes of definitions for data quality. I then made the data quality definitions more specific by selecting variables that map to the 90-90-90 targets for HIV eradication, which have been adopted in Malawi (58). The variables were selected after confirmation that they were being collected in the EHR. I followed up on the variables in Table 2.7 with the EHR implementer to confirm that the data were indeed collected. Initially, I planned to conduct the data quality dimensions assessments based on validation and verification as the Kahn framework suggests. Verification has the key feature to determine expected values and distributions using resources within a local environment, while validation focuses on alignment of data values with respect to external benchmarks. However, I did not have access to external data sources to conduct validation data quality assessment, so I only focused on verification (63). Additionally, I observed the need to make the definitions of my data quality assessment specific. I thus modified some definitions of the proportions from the ones that I made in my proposal. I have provided a table of the modified definitions in Appendix 5. I present the variables selected, specific definitions and status of any changes in definitions that I followed for my data quality assessment in Table 2.8 below. These data were measured using MySQL database queries.

Table 2.8 Definition of data analysis proportion

90-90-90 data use	Variable	Completeness	Atemporal Plausibility	Value Conformance	Completeness (Validation)	Status of Change in the Methods Chapter	Remark
Identification of HIV positive individuals	Unique patient identifier	all assigned patient identifiers/all registered patients with a dispensation	All patient identifiers not compliant with format/all patient assigned identifiers	All patient identifiers issued at registration with matching data dictionary definition/all patient identifiers in the database	All patient identifiers issued at registration verified at central database/ All patient identifiers issued for that site	No longer assessed for all characteristics	Unavailability of some datasets – The EHR component had not been implemented by the time of my study.
Patient tracing	Patient Address – current place of residence	All patients registered with a current place of residence/all patients registered with a dispensation	All non-existent current places of residence /all patients with a current place of residence	All patients with current place of residence matching domain of residences documented at registration/all patients registered with a dispensation	All patients with current place of residence documented at registration verified with master/ All patients with current place of residence documented at registration	Made the definitions specific by stating the denominator as “All study patients during the research period” Made numerator for atemporal plausibility specific by adding “look up table” Added other plausibility type of Temporal plausibility to be consistent with other variables.	Making the definitions exact – Added population and duration to the data quality assessment proportions. Consideration of other sub-types of plausibility
	Patients with contact phone number	All patients registered with a phone number/all	All patients registered with a wrong sequence of a phone number/ All patients registered	All patients registered with a phone number documented at registration with correct database format/ all patients registered with a dispensation		No longer preforming validation data quality assessment due to	

90-90-90 data use	Variable	Completeness	Atemporal Plausibility	Value Conformance	Completeness (Validation)	Status of Change in the Methods Chapter	Remark
		patients registered with a dispensation	with a phone number			unavailability of Master database for validation.	
Quality of patient care	Prescription versus dispensed drugs	Number of prescriptions filled/number of prescriptions made	Number of prescriptions matching dispensations/number of dispensations made	Number of prescriptions filled on visit with correct database definition/Number of prescriptions filled	N/A	No longer performing conformance as part of the data quality assessment due to in part its inherent inclusion in the design of the EHR.	Making the definitions exact – Added population and duration to the data quality assessment proportions for prescriptions filled.
	Patients with changed regimens and reasons for changing	Number of patients with changed regimen and documented reason for changing/ Number of patients with changed regimen	Number of patients with documented reason for changing that does not match the expected guideline / number of patients changed to a specific regimen and documented reason for changing	Number of patients with changed regimen and documented reason at changing saved from domain of reasons/number of patients with changed regimen and documented reason for changing	N/A		Consideration of other sub-types of plausibility for prescriptions filled. Unavailability of some datasets – Did not analyze the variables on changes in regimens because the dataset was not provided by the system implementer.
	Body mass index –	Number of patients without	Number of patients with weight that	Number of patients with weight and height documented at	N/A	No longer performing conformance as part	Making the definitions exact – made changes for

90-90-90 data use	Variable	Completeness	Atemporal Plausibility	Value Conformance	Completeness (Validation)	Status of Change in the Methods Chapter	Remark
	weight and height	height/number of patients registered with a dispensation	is out of bounds/ Number of patients with weight and height Number of patients with height that is out of bounds/number of patients with weight and height	registration with expected domain range/number of patients with weight and height documented		of the data quality assessment due to in part its inherent inclusion in the design of the EHR.	numerator to be "patients with" and "not without". Changed the denominator with exact patient population and time period. Consideration of other sub-types of plausibility for BMI variables of height and weight.
Viral load suppression	Viral load results present	Number of patients with one observation of viral load result 6 months or more after registration/all patients with a registration into the ART program 6 months or more after registration	Number of patients with out-of-range viral load results/Number of patients with one observation of viral load 6 months or more after registration	Number of patients with a viral load result within expected timeframe of milestone with correct domain range/all patients that reached a viral load milestone within 6 months or more after registration	Number of patients with viral load that match with Central Database/Number of patients with viral results for a site 6 months or more after registration	No longer performing conformance as part of the data quality assessment due to in part its inherent inclusion in the design of the EHR.	Making the definitions exact – Added the exact range checks for Atemporal plausibility. Consideration of other sub-types of plausibility for Viral load results in order to describe in detail the different possibilities of a plausible viral load result.

I used descriptive statistics such as proportions, count statistics and charts to conduct data analysis for my data quality assessment as specified in Table 2.8 using the R statistical package(86).

Denominator definitions formed a core part of the data quality assessment that I performed. I have provided the updated table definitions as Appendix 6. I operationalized the data quality dimensions used in the analysis above with the definitions below:

Completeness: This was measured by assessing the frequencies of data attributes in a data set without reference to the data values. The restriction to assessing only the presence or absence of values without assessing whether these values fall in allowable limits differentiates completeness to plausibility. Completeness may be measured as a single instance or over several points in time. The time variations were applied differently based on the data variables for this study.

I had two groups of denominators for measuring completeness in this study. The study population from the 10 health facilities was the denominator for the first group of variables. Patient physical address, patient mobile phone number and patient height fell in this group. Patient viral load also used site population as its denominator although only 3 central hospitals were selected as the population for this denominator due to availability of site data. These variables had the site population as its denominator because each patient is expected to have at least one active observation during their time of attending the ART clinic. On the other hand, I determined the total number of study patient visits as the denominator for variables that should be collected at every patient visit. These included patient weight, patient drug adherence and patient drug prescriptions.

Plausibility: This deals with believability or truthfulness of values of data. Unlike completeness and conformance that deal with presence and structure of data values, plausibility deals with how closely data values represent the real world. Plausibility may be assessed in different ways:

Atemporal plausibility: This assesses whether observed data values meet internal expected values such as range checks or externally validated density distributions, such as weight for age checks for example. For example, weight values for adults in this case were expected only within a given range. Atemporal plausibility was the most common type of plausibility that I employed. I assessed atemporal plausibility for all variables as specified in Table 2.8 above.

Temporal plausibility assessed whether the observed values follow expected precedence, with an example of viral load measurements being expected after an ART start date and not before their start date. The specific definitions for the analyses that I used are in Table 2.8 above and in Appendix 6 and the level of plausibility was confirmed with the Malawi national guidelines for treating HIV/AIDS (80).

2.6 Ethical approval

I received ethical approval to conduct this study from the Institutional Review Boards of the University of Michigan Medical School (IRBMED) and the Malawi National Health Sciences Research Committee (NHSRC) (87). The study identification numbers for my IRBMED and NHSRC applications were HUM00161418 and 2341 respectively. I did not interview patients, hence there was no protected health information (PHI) involved for Aim 1. However, PHI was accessed during the extraction of patient records for Aim 2. These records were extracted by the by the EHR implementing organization (Baobab Health Trust) and de-identified to minimize the risk of compromising privacy and confidentiality of the patient records before being handed over to me. Additionally, I stored the collected information securely on an approved, password-secured drive and ensured that only research team members had access to the information.

Chapter 3 Study Results

3.1 Overview of the results section

I present the results of my research following the methods described in Chapter 2. My project was a convergent mixed methods study using convergence to integrate my quantitative and qualitative results.

3.2 Study Participants and Study population for Aims 1a and 1b

I interviewed 34 participants for this research rather than the planned 36 because one donor and one public health official were not available. I interviewed stakeholders holding the roles described in Table 3.1 below. The roles were not mutually exclusive, as some stakeholders had multiple responsibilities. For example, health managers, such as health center managers, were usually also clinical officers or nurses. Healthcare providers were the highest number of EHR users I interviewed.

Table 3.1 Study Participants

Main EHR Stakeholder	Roles		Total Count
Data Services	Data officer or Monitoring and Evaluation officer	2	9
	Health Management Information Officer (HMIS)	1	
	ART data clerk	6	
Healthcare providers	Clinical officer	4	15
	Community health nurse	2	
	Registered nurse midwife	8	
	Medical assistant	1	
Health Managers and Donors	Health Center managers	2	10
	ART clinic coordinator	2	
	ART clinic manager	1	
	Public health officials	3	
	Donors	2	
Total number of study participants			34 (out of an expected 36, 94%)

3.3 Qualitative Data Analysis Results

3.3.1 Aim 1a: Identifying purposes of data use for different stakeholder

My semi-structured interviews ranged from 25 to 30 minutes. After transcribing the interviews and defining the codebook as described in the methods chapter (Included as Appendix 3), I coded the transcripts into codes derived from the WHO Classification of Digital Health Interventions (CDHI) as purposes of data use. I observed that some purposes of data use were clinical facing, with linkages to direct patient care processes. Other purposes of data use focused on administrative end goals such as supply chain decisions. I summarized the data purposes into two categories. These categories were administrative purposes of data use and clinical-facing purposes data use, as presented in Table 3.2 below.

Table 3.2 Stakeholders' purposes of data use

Clinical Facing purposes of data use		
WHO CDHI category	**Exemplar themes**	**Exemplar Quotes**
Client identification and registration	Unique patient registration and identification;	*"For example, the name of the patient is very important because we can have two "Mercy's" or two "Elizabeths" at the same time"* – Nurse at a health center.
	Unique patient registration and identification;	*"The data helps me when doing patient follow-up, for example, the patient address"* – a nurse at a District Hospital.
Patient health tracking	Trace patients lost to follow up;	*"We also have an advantage after seeing that data and so that, aah, we have a lot of defaulters in the program, so we entice them to come back to the clinic."* – Clinician at a District hospital.
	General health tracking. - Maintaining the quality of care or improving quality improvement projects	*"For those patients who have had side effects; then we can know that these side effects that they now have (present with); they have improved from their previous side effects"* – Clinician at a central hospital.
Prescription, and medication management	Give medications;	*"So, we ask the question for the visit that they (patients) are here, we as well record the side effects in case they have any, and then dispense the*

		appropriate medications as well." – Nurse at a central hospital.
Healthcare decision support	Clinical decision support;	*"To change regimens, we also use the same data, where we can actually determine whether the patients are failing the treatments after they have collected their, aah, viral load. And also, sometimes we do CD4 counts which we can enter into the system." –* Clinician at a District hospital.
	Counseling decision support.	*"Some are young (new HIV positive pregnant patients), for example, 17-year-old or 14-year-old, while some are older, so it helps in determining how to explain to these women. Some are older (HIV positive pregnant patients), and they understand what we are saying. "–* Nurse at a health center
Referral coordination	Intra-health facility referrals;	*"As I said, if there is something I cannot manage, I am given options in the same system that I should send the patient to the doctor. If they are easy things (presenting complaints), then the nurse can act*."- Nurse at a health center.
	Inter-hospital referrals.	*"For example, this facility is just a health center, and we refer patients to other hospitals; for example, to Kamuzu Central Hospital (KCH). So, in the case that the patient dies, we can use the data to explain how we cared for that patient before being referred to them at KCH." –* Nurse at a health center.
Administrative purposes of data use		
	Exemplar Themes	**Exemplar Quotes**
Data collection, management, and use	Reporting;	*"This data is important because, at the end of the month, we can know the number of women on ART treatment for HIV and those that are adherent. In the end, we can know the number of patients that require drugs." –* ART Clerk at a health center

	Data analysis and interpretation	*"The kind of more granular or visit level data that is more useful. So, one example is a discussion we have had recently of TB preventive therapy on what kind of policy options we might want to implement. There we wanted to look at the time intervals between initiation of ART and TB treatment for patients who develop TB because this could inform a strategy, we want to implement a preventative therapy"* – MoH public health official.
Facility management	Allocation of physical infrastructure resources;	*"With the data, we can plan that we have this data we think we should plan for the future using available resources. So, we can also lobby other resources (like rooms, chairs, tables) using the data."* – A health center manager.
	Human resource allocation	*"The data also helps when looking at the capacity of the clinic. The nurses, so how many nurses should we have in terms of the nurses, because we know the patients, so the data is helping in terms of how many nurses we can allocate."* – Clinic director at a cetre of excellence as donor.
Supply chain management	Ordering drugs or medical supplies;	*"In times of maybe ordering drugs; we need to have the number of people so that the drugs you order should be within that consignment"* – an ART clerk at a health center.
	Drugs or medical supplies forecasting	*"Use data for the logistics, supply chain management, procurement planning and of course including reporting of performance to global fund primarily and other funding sources. So, I think it's pretty much the whole cycle..."* - MoH public health official.
Defining and monitoring policy	Project decisions;	*"Use data to inform programming of the baseline for a project. Key decisions are based on data. Identifying health facilities where*

		EMRs can be deployed." – health manager from a MoH donor partner.
	Guideline or policy Guideline definition or patient monitoring;	"*So, in that case, we made a model, which is a Lighthouse brand which we called Advanced Late and Unstable Patients (ALUP), which makes sure that we have a care package for such patients. So, when these patients come in, we make sure that we are packaging them to make sure that they virally controlled with treatment, enhance their nutrition, and at the same time we are also diagnosing other conditions. So for me, the decision of putting up a package of advanced HIV is coming in because I can see the backend of the data, which is coming up to say, what proportion of my patients are coming up with severe disease so that everyone should be well." –* Clinic director at a center of excellence as a donor.

I have provided more detail about the findings related to the purposes of data use presented in Table 3.2 above in Appendix 3.

3.3.2 Aim 1b: To identify characteristics of data quality that stakeholders using EHR data expect.

I explored the concept that "data quality is fit for purpose" to establish the data quality characteristics that study participants required for their purposes of data use. I mapped themes of expected data quality characteristics to the Kahn harmonized data quality framework (63). Most stakeholders mentioned a group of data quality characteristics that they expect in the data that they use rather than pointing to a single characteristic. The group of commonly found characteristics included completeness, accuracy, and timeliness or currency. No stakeholder talked about data quality in terms of the structure of the data. This data quality characteristic is defined as conformance. This observation is not surprising given that conformance is more of a technical data quality characteristic, and the study interviewees were non-technical. I present the main data quality characteristic categories and themes that stakeholders mentioned as expecting in their data use in Table 3.3 below.

Table 3.3 Anticipated data quality characteristics

Anticipated Data Quality Characteristic Main	**Themes**	**Exemplar Quotes**
Completeness	Accessible	*"Another thing, it (the data) should be accessible because if we have the data and people cannot access the data or the information, it will be difficult for people to use the data."* – Nurse at a health center.
	Available	*"Usually, once the data is available, I am a clinical epidemiologist, so I try to make sure that you look at the completeness of the data."* – MoH public health official.
	Available or representative	*"The number of health facilities that have submitted their reports to the district hospital is important. If I have 50% of the health facilities and also only a few reports are available, then it's a problem."* – HMIS officer at a District Hospital.
	Contextual	*"Aah ,make sure that you can define all that you may not find that sort of data aah; also qualitatively, I think or descriptively you may not find the data to tell you"* – MoH public health official
	Up to date or Current or Timely	*"Obviously, the first one is that timeliness is important. When you want to make a decision, you have to make sure that the data is available."* – A donor of the EHR.
Plausibility	Accurate	*"Because viral load is figures, they can all be recorded as normal, but it's not the same right; because a count of less than 40, the other one a count less than 800 something, the other one a count of Zero,…but then all of them are described as suppression. Good suppression, but they are confusing."* – a Clinician at a tertiary hospital. *"The data should represent what is on the ground. It should not be guess-work"* – a Clerk at a district hospital.
	Simple	*"And also, it should be simple, because if the data is complicated, people are not able to understand the data it will be difficult to use the data"*- a nurse at a health center.

3.3.2.1 Completeness Category

I observed the highest repetitions of these data quality characteristics from all the stakeholders. Stakeholders described completeness with different themes such as availability, contextual, and in some cases, with either currency or timeliness.

Availability

Availability was mainly cited in the context of both clinical facing purposes of data use and administrative purpose of use. The responses spoke directly to the purpose of data use for healthcare providers, that viewed "availability" of data as enabling their roles of patient health tracking.

On the other hand, administrative purposes of use tended to describe the "availability" of data by making sure that they have a complete dataset that should enable them to have a complete picture and make decisions such as the quantity of drugs to procure.

Contextual

I found that context was also mentioned as an anticipated data quality characteristic. I mapped context to the Kahn data quality characteristic of completeness. This theme of contextual data quality has been reported as a dimension for describing data quality before by Wang et al. and has also been harmonized as completeness in the Kahn Harmonized data quality framework for secondary data use (63,64).

I found that for clinical facing purposes of data use such as patient assessment, the suggestion seemed to have a flexibility to add more details in free text about a patient's condition.

Timeliness or up to date

I found timeliness mentioned concerning data being available for use to meet a purpose of use of any stakeholder group. Timeliness as an expected data quality characteristic has also been mapped to completeness in the Kahn framework. I categorized timeliness for both administrative purposes of data use and for clinical facing purposes of data use. Health managers for example view timeliness in relation to the purpose of policy or program decisions as shown in the example in Table 3.3 above.

On the other hand, healthcare providers that have more clinical facing purposes of data use look at timeliness in relation to data being available in the EHR at the time that they want to make a decision that would affect the purpose of use of patient health tracking as provided in Table 3.3 above.

3.3.2.2 Plausibility category

I coded the second highest number of themes into this category. Although numerically, it came second to completeness, many stakeholders in describing their expected data quality characteristics emphasized plausibility as being important in that it would help them to correctly achieve their purposes of data use. The themes that were coded into this category were accuracy, and simplicity. Correctness was also mentioned as a theme but was mentioned alongside accuracy.

Accuracy

Stakeholders spoke of accuracy as having data that represents the true meaning of what has been observed. The Kahn framework maps accuracy to plausibility and describes different types of plausibility across three classes as described in the methods chapter. Accuracy was expected across the two main categories of administrative and clinical facing purposes of data use. I provide an illustration of accuracy as an anticipated characteristic from a clinician in Table 3.3 above.

Simplicity

This theme was mentioned concerning both clinical facing and administrative purposes of data use. Stakeholders said their desire to quickly understand any data that is presented for their purpose of use. I have provided an example of one healthcare provider that performs patient health tracking by monitoring the patient adherence to medication, among others, as presented in Table 3.3 above.

3.3.3 Aim 1c: To rank the importance of data quality characteristics by purposes of data use across different stakeholders using EHR and EHR data.

I used two approaches to ascertaining the importance of data quality characteristics.

I first asked for the importance or priority of each characteristic as an open-ended question to participants on data quality characteristics they anticipate in their purposes of data use. I found that most stakeholders could not choose a single data quality characteristic as the most important for their purposes of data use. They rather first responded that all of the data quality characteristics were very important. However, through further probing, I observed that stakeholders with clinical facing purposes of data use gave priority to plausibility themes compared to completeness. On the other hand, stakeholders mentioned completeness as the most important data quality characteristic for administrative purposes of use. Although I made this observation, I could not rank the importance of data quality characteristics based on the open-ended question that I asked during the semi-structured interviews.

Since my study applied actual purposes of use in ART clinics for HIV/AIDS, I assessed the priority of the importance of data quality characteristics concerning HIV eradication goals.

A total of 11 participants were interviewed for this Aim. Although these study participants were the from the same target study population for Aims 1a and 1b (34 participants), I interviewed the first 23 study participants for Aims 1a and 1b in a different geographical location before I had developed the data collection tool for Aim 1c. This is why I ended up interviewing only 11 of the study participants in Aim 1c (Table 3.4):

Table 3.4 Study participants for Aim 1c

Main CDHI Stakeholder Category	**Roles**		**Total Count**
Data Services	Data officer or Monitoring and Evaluation officer	1	3
	Health Management Information Officer (HMIS)	1	
	ART data clerk	1	
Healthcare providers	Clinical officer	1	3
	Registered nurse midwife	2	
Health Managers	Health Center managers	1	5
	EHR Donors	2	
	Public health officials	2	
Total number of study participants			11

These 11 participants formed the total number of votes across the different data quality characteristics pairwise comparisons of the priority of importance. I present the results of the data quality characteristics comparisons in Table 3.5 below based on the table from the methods chapter. Each voter in Table 3.5 below is represented by the letter V.

Table 3.5 Sum of Votes and Weighted Scores for each data quality characteristic

Comparison of Characteristics	V- 1	V- 2	V- 3	V-4	V-5	V-6	V-7	V-8	V-9	V-10	V-11	**Sum of Votes**
Purpose of use: Unique patient identification												
Plausibility	0	1	1	0	0	1	1	0.5	1	0	0.5	**6**
Completeness	1	0	0	1	1	0	0	0.5	0	1	0.5	**5**
Plausibility	0	1	1	0	0	1	1	0.5	1	1	1	**7.5**
Conformance	1	0	0	1	1	0	0	0.5	0	0	0	**3.5**
Conformance	0	1	1	1	1	1	1	0.5	1	0	0	**7.5**
Completeness	1	0	0	0	0	0	0	0.5	0	1	1	**3.5**
Purpose of use: Healthcare decision making												
Plausibility	0	0	1	0.5	0.5	0	0	0.5	0	1	0.5	4
Completeness	1	1	0	0.5	0.5	1	1	0.5	1	0	0.5	7
Plausibility	1	0	1	0	1	0	0	0.5	0	1	1	5.5
Conformance	0	1	0	1	0	1	1	0.5	1	0	0	5.5
Conformance	1	0.5	1	0	0.5	0	0	0.5	0.5	1	1	6
Completeness	0	0.5	0	1	0.5	1	1	0.5	0.5	0	0	5
Purpose of use: Viral load monitoring												
Plausibility	0	1	0	1	0	1	0.5	1	1	0.5	1	**7**
Completeness	1	0	1	0	1	0	0.5	0	0	0.5	0	**4**
Plausibility	1	0	0	0	0	1	1	1	1	1	0.5	**6.5**
Conformance	0	1	1	1	1	0	0	0	0	0	0.5	**4.5**
Conformance	0	0	0	1	0	0	0.5	0	1	0.5	1	**4**
Completeness	1	1	1	0	1	1	0.5	1	0	0.5	0	**7**

I found that plausibility had the highest importance for the purpose of the use of unique patient identification followed by completeness and lastly conformance.

I found completeness to be of the highest importance followed by plausibility and lastly conformance for the data purpose of use of healthcare decision making.

I found that plausibility had the highest importance for the purpose of the use of viral load with a tie in the priority of importance for the data quality characteristics of completeness and conformance.

My results above show variation in ranking the importance of data quality characteristics for the selected purposes of data use for the 90-90-90 HIV/AIDS eradication goals. This observation implies that stakeholders may need to consider the importance of priority assigned to different data quality characteristics when considering a data quality assessment or even interventions to improve any observed gaps in data quality.

3.3.2.3 Integrating Purposes of Data Use, Expected Data Quality Characteristics, and Importance of Data Quality Characteristics

I used a side-by-side comparison of themes to integrate the qualitative results for Aim 1(72). The stakeholder purposes of data use are not exclusive to one data quality characteristic anticipated, hence the repetition. I present the integration of Aim 1 in Table 3.6 below.

Table 3.6 Side by side theme Integration of Aims 1a and 1b

Categories of Purposes of Data Use Reported	Categories of expected data quality characteristics	90-90-90 EHR example purpose of data use	Rank of the importance of expected data quality characteristics
Data management and use; Supply chain management	Completeness Plausibility	Unique patient identification.	1. Plausibility 2. Completeness 3. Conformance
Supply chain management; Defining and monitoring policy	Completeness Plausibility	Healthcare decision support	1. Completeness 2. Plausibility 3. Conformance
Facility management; Supply chain management; Defining and monitoring policy; Data management and use	Completeness Plausibility	Healthcare decision support	1. Completeness 2. Plausibility 3. Conformance
Patient health tracking; Healthcare decision support; Referral coordination;	Plausibility Completeness	Viral load monitoring	1. Plausibility 2. Conformance, Completeness

All stakeholders were able to report data quality characteristics that they anticipate. It was clear that the data quality characteristics were linked to stakeholders' reported purposes of data use. I found that stakeholders who mentioned clinical facing purposes of data use, expected their data to be plausible compared to stakeholders that mentioned administrative purposes of use. These stakeholders emphasized anticipating their data to be complete first to achieve their administrative purposes of use. My integration above showed convergence in stakeholder purposes of data use, expected data quality characteristics, and the priority of importance attached to data quality characteristics for all of the stakeholder purposes of use apart from the purpose of data use of unique patient identification. For this unique patient identification, stakeholders mentioned expecting completeness with greater emphasis but there was divergence in the priority of data quality characteristics. The divergence showed that stakeholders attached greater importance to the data quality characteristic of plausibility.

3.3.4 Aim 2: To conduct a data quality assessment of EHR data based on Kahn framework data quality characteristics and purposes of use in ART clinics in Malawi

I present the patient study population on which I collected patient visit data for data quality assessment of Aim 2, in Table 3.7 below. I had a patient population of 160,647, which had 549,826 patient visits during the study period from January 2017 to December 2018.

Table 3.7 Study population Aim 2 used for patient physical address, patient mobile number and patient height

	Primary Healthcare Facilities **Health Center** **(n=33,451)**	Secondary Healthcare Facilities **District** **(n=37,863)**	Tertiary Healthcare Facilities **Central Hospitals** **(n=89,333)**	**Overall** **(n=160,647)**
Sex				
Female	22,418 (67.0%)	24,583 (64.9%)	52,655 (58.9%)	**99,656 (62.0%)**
Male	11,033 (33.0%)	13,280 (35.1%)	36,678 (41.1%)	**60,991 (38.0%)**
Age groups				
15-24	6,736 (20.1%)	6,099 (16.1%)	9,578 (10.7%)	**22,413 (14.0%)**
25-34	14,454 (43.2%)	14,477 (38.2%)	33,973 (38.0%)	**62,904 (39.2%)**
35-44	8,574 (25.6%)	10,776 (28.5%)	28,686 (32.1%)	**48,036 (17.6%)**
45-54	2,638 (7.9%)	4,500 (11.9%)	12,248 (13.7%)	**19,386 (12.1%)**
55+	1,049 (3.1%)	2,011 (5.3%)	4,848 (5.4%)	**7,908 (4.9%)**
Age				
Mean (SD)	32.7 (10.6)	34.8 (10.6)	36.1 (10.2)	**35.3 (10.3)**
IQR[Q1,Q3]	12[26,38]	14[27,41]	13[29,42]	**13[28,41]**

The study population results in Table 3.7 reflect the number of patients at different health facility types. The health facility population sizes above match the expected catchment population of the health facility type. Specifically, health centers that are the smallest unit have the lowest population, followed by district hospitals and lastly central hospitals with the highest population. The gender representation of 62.0% female patients and 38.0% male patients was in agreement with national proportions on gender that are 64% females and 36% males (50). The study patient population forms the denominator for assessing the completeness of some of the variables for data quality assessment. These variables are the patient physical address, patient mobile number, and patient height. I intended to use the patient population above for viral load as well. However, since I collected viral load data only in three central hospitals, I used the population in these hospitals, which I have presented in Table 3.8 below:

Table 3.8 Population of viral load patients from three tertiary hospitals

	Queen Elizabeth Central Hospital (QECH) (n=16,298)	**Zomba Central Hospital (ZCH)** (n=7,662)	**Kamuzu Central Hospital (Lighthouse - KCH)** (n=13,360)	**Overall** (n=37,320)
Sex				
Female	9,590 (58.8%)	4,769 (62.2%)	7,851 (58.8%)	**22,211 (59.5%)**
Male	6,708 (41.2%)	2,893 (37.8%)	5,508 (41.2%)	**15,109 (40.5%)**
Age groups				
15-24	1,790(11.0%)	805 (10.5%)	1,188 (8.9%)	**3,783 (10.1%)**
25-34	6,043(37.1%)	2,922 (38.1%)	4,820 (36.1%)	**13,785 (36.9%)**
35-44	5,560 (34.1%)	2,442 (31.9%)	4,774 (35.7%)	**12,776 (34.2%)**
45-54	2,137 (13.1%)	1,096 (14.3%)	1,955 (14.6%)	**5,188(13.9%)**
55+	768 (4.7%)	397 (5.2%)	623 (4.7%)	**1,788 (4.8%)**
Age				
Mean (SD)	35.9 (10.6)	36.2 (10.6)	36.6 (10.2)	**36.2 (10.3)**
IQR[Q1,Q3]	13 [29,42]	13 [29,42]	12 [30,42]	**13[29,42]**

The gender and age distributions for this population of the viral load has similar proportions with the tertiary population in Table 3.8 above. Similarly, the proportions of age distributions are similar to the aggregate age proportions of tertiary hospitals in Table 3.8 above.

I proceed by first presenting completeness of the patient visits in line with the variables for measuring the 90-90-90. Afterwards I present plausibility data quality assessment results.

As expected in terms of catchment population, tertiary hospitals had the highest number of visits, followed by secondary health facilities and, lastly, primary hospitals. I present the breakdown of the patients' visits in Table 3.9 below

Table 3.9 Patient visits at different health facilities

Health Facility Name	Health Facility Type	Number of Patient Visits
Queen Elizabeth Central Hospital	Tertiary Level	106,516 (19.4%)
Zomba Central Hospital	Tertiary Level	56,266 (10.2%)
Kamuzu Central Hospital: Lighthouse	Tertiary Level	98,950 (18.0%)
Mzuzu Central Hospital	Tertiary Level	38,640 (7.0%)
Mangochi District Hospital	Secondary Level	58,167 (10.6%)
Dedza District Hospital	Secondary Level	48,784 (8.9%)
Nkhatabay District Hospital	Secondary Level	17,193 (3.1%)
Limbe Health Center	Primary Level	55,670 (10.1%)
Kawale Health Center	Primary Level	36,173 (6.6%)
Mapale Health Center	Primary Level	33,467 (6.1%)
Total patient visits		549,826

I present my analysis of the completeness data quality characteristic for my study variables in Table 3.10 below.

Table 3.10 Completeness proportions for all health facilities

Variable Name	Total number counted	Denominator value	Proportion Completeness Per Variable (%)
Current patient physical address	8,717	160,647	5.4
Patient Mobile phone number	89,569	160,647	55.7 %
Height of patients	124,001	160,647	77.2 %
Weight of patients	549,742	549,826	99.9%
Drug Adherence	247,773	549,826	45.1%
Prescriptions made	419,826	549,826	76.4%
HIV Viral load	30,036	37,320[4]	80.5%

The variables above map to the 90-90-90 efforts for eradicating HIV/AIDS by 2030 in Malawi. There is incompleteness in all variables, which is disturbing for the national efforts.

For the patient retention purpose of use, which is enabled by complete current physical address and complete patient mobile phone number, I observe that only 5.4 % of patients have a

---

[4] This denominator is the total population of patients alive at the beginning of the study period for tertiary health facilities where I collected viral load data.

complete physical address that may be useful for tracing patients. 55.8% have a complete mobile number meaning that this could potentially be used as an alternative to trace patients that missed their ART clinic appointments. 39% of the population in the country had mobile phone coverage as reported in 2018 (88). The coverage at the health facilities represents a higher proportion than the national average probably because these health facilities are located in urban and semi urban areas. Completeness of patient address was analyzed by counting records of patient addresses that had all components of physical address, namely: village, the traditional authority, and their district documented. Taking into consideration that tracing patients was a purpose of use that was mentioned by healthcare providers and those working in data services but had meager completeness proportions, I sought to understand the completeness of this variable more, as shown in Figure 3.1 below.

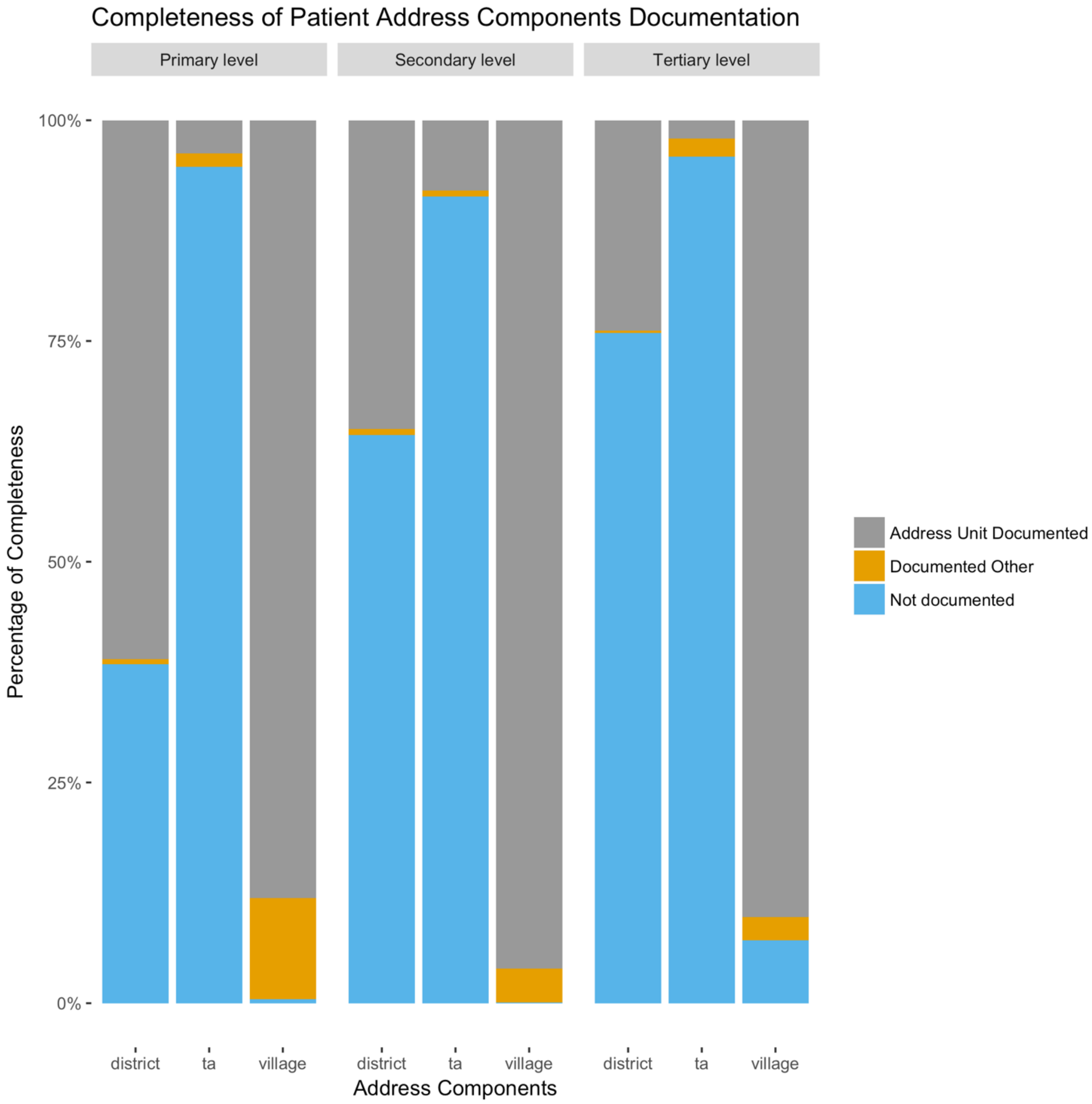


Figure 3.1 Analysis of individual address components

The x-axis contains the different components of the patient physical address. The y-axis is a percentage of address components that have been documented for each health facility type.
I observed that the Traditional Authority was the most undocumented component of physical address across all health facility types. This may be particularly high for health facilities that are in the city and hence the current residence of residents not being assigned to a traditional authority but to a ward. The health centers and tertiary hospitals were both in cities. I however observed a similar high proportion in undocumented Traditional Authority at the district hospitals level which are in rural areas. As much as this is an integral part of patient physical

address, I observed that the village is documented the highest across all health facility types with 88% completion proportion at Health centers and 96% and 90% completion at District and Tertiary health facilities respectively. This observation implies that despite the meager address completeness rates, it should be possible to follow up on the patients that have a village documented and hence meet the patient tracing purpose of data use.

Height and weight are variables that may contribute to patient retention at health facilities as they constitute part of standard care by facilitating the monitoring of body mass index (BMI) for patients enrolled in the clinic (80). I proceeded to assess their completeness since a continuous decline of BMI is an indicator of ART treatment failure (89,90). Treatment failure would result in poor outcomes of patients on treatment such as death or even patients being stopped on treatment to prevent further harm to the patients. This would in turn reduce the number of patients that are alive and retained on treatment. I observed that almost 100% of patient visits had a weight encounter. This finding is not surprising in the case of my study since weight is one of the first interactions of patients during their visit to a clinic. Almost all the study patient visits had a weight observation with coverage of about 99%. However, only 77.2% of patients had a complete height. This proportion implies that it will not be possible to calculate BMI for over 20% of patients.

Drug adherence is another measure that is important in the 90-90-90 target of retention. Each patient visit is supposed to have a drug adherence encounter documented. Only 45.1% had a drug adherence interaction recorded during the patient visits in this study. This result creates a challenge for the treatment strategy to eradicate HIV since the HIV/AIDS program will not know how well patients are taking their medications for over 50% of the patients enrolled in the program.

The goal of the 90-90-90 target is to achieve viral load suppression in 90% of patients that have been retained in treatment. Viral load is a required measurement six months after a patient starts taking medication. I observed that 80.5% of the patient population had their viral load measured in the tertiary hospitals where I collected the data. Given the 90-90-90 targets, I should expect 73% of the study population to have had their viral load measured. Since there are 37,320 patients in the tertiary health facilities, the expected count at 73% would be 27,244. Since I found 30,036 patients had a complete viral load measurement, this surpasses the expectation and hence is in line with national expectations. The country achieved 90% of patients with viral load measurement and suppression (91). I explain this high finding based on the type of health facilities being tertiary health facilities hence being well resourced to test the viral load of patients. As I explained in the methods chapter, the scope of viral load records assessed in this study is only for three tertiary health facilities out of the targeted ten health facilities. Therefore, my results for viral load may only be generalized to within these three health facilities and at the tertiary level.

Having assessed the completeness of the records, I proceeded to explore their plausibility. I present the plausibility results in Table 3.11 below:

Table 3.11 Plausibility of various variable observations

Variable names	**Total number of plausible records**	**Denominator Value**	**Proportion Plausibility Per Variable (%)**
90-90-90 Purpose of Data use: Patient Tracing Variables			
Atemporal Current Patient Physical Address	3,514	8,717	40.3
Temporal Plausibility Patient Physical Address	117,472	160,647	73.1
Atemporal Patient Mobile Phone Number	46,530	89,569	52.0
Temporal Patient Mobile Phone Number	89,569	160,647	55.8%
90-90-90 Purpose of Data use: Quality of patient care			
Atemporal Weights of Patient	548,825	549,742	99.8
Temporal Weights of Patient	385,480	549,742	70.7
Atemporal Heights of Patient	122,051	124,001	98.4
Temporal Heights of Patient	100,421	124,001	81.0
Atemporal Prescription Made	331,613	333,492	99.4
Temporal Prescription Made	323,924	333,492	97.1
Atemporal Drug Adherence	300,177	339,697	88.4
Temporal Drug Adherence	212,036	339,697	62.4
90-90-90 Purpose of Data use: Patient HIV Viral load monitoring			
Atemporal HIV Viral Load	17,286	30,036	57.6
Temporal HIV Viral Load	29,417	30,036	97.9

I calculated proportions for atemporal plausibility and temporal plausibility based on the 90-90-90 variables, as presented in the Table 3.11 above.

My results show low proportions of atemporal plausibility for both physical address and patient phone number. I counted patient physical addresses as being plausible at the atemporal level if the documented addresses matched standard location address components in a particular district for a patient. This would mean validating whether a documented district in the patient address is indeed a district that exists first. In the next step, I assessed whether traditional authority exists in that district and lastly whether the village exists in that village. I counted an address as plausible at the atemporal level if I could validate all these components of patient address. With regards to the low atemporal plausibility, it means that although stakeholders may rely on phone numbers to trace patients that have incomplete physical addresses, they could do this accurately for only 52.0% of the patients with a documented phone number value. Atemporal plausibility

I, however, find that quality of care variables for BMI, such as height and weight, have high atemporal plausibility concerning their complete observations. This observation could be explained via the design of the EHR, which has built-in range checks that ensure correct data is entered in the system (15).

I found that temporal plausibility varied within a range of 70% to 97% across all study variables. Drug adherence has atemporal plausibility of 88.4%. Plausible adherence percentages fall between 0 to 105% based on the Malawi treatment guidelines for HIV/AIDS (6). The Malawi national guidelines for the treatment of HIV/AIDS include an adherence up to 105% considering the scenario that some patients may come sooner than their appointment dates but with less medication than they should have because they either lost some drugs or shared then with partners. All patients that had their adherence percentage outside the possible range of 0 to 105% were counted as not plausible.

I found that 57.6% of viral load patient records had met atemporal plausibility. Atemporal plausibility counted patients that had their first viral load measured within 5 to 36 months of starting ARVs. I used such a flexible definition, considering that most patients had not yet had their viral load measured since the program for the viral load was being rolled out during my study period. Therefore, most patients had been on treatment beyond the 5 to 12 months period in the guideline recommendations. I thus took the maximum milestone for viral load measurement to accommodate the initial tests.

3.3.5 Integrating Data Quality with Expected Purposes of Use

I observed that purposes of use were not exclusive to stakeholder roles and the domains in which they work. For example, stakeholders such as public health officials and healthcare providers both reported purposes of data use with themes of data analysis, interpretation and reporting. I regarded the common unit for integrating my qualitative and quantitative results based on the stakeholder data purposes of use, expected data quality characteristics, ranking of the priority of importance of data quality characteristics, and then the results of my data quality assessment. I used a theme by statistics approach to integrate the qualitative and quantitative results of my study in Table 3.12 below.

Table 3.12 Theme by Statistic Integration of Aims 1 and 2

Purposes of Data Use Reported (Aim 1a)	**Categories of expected data quality characteristics (Aim 1a)**	**Priority of expected data quality characteristics (Aim 1b)**	**Data quality variables in line with 90-90-90 priorities (Aim 1c)**	**Data quality variables proportions (Aim 2)**
Data management and use; Supply chain management	Completeness Plausibility	1. Plausibility 2. Completeness 3. Conformance	Current patient physical address	40.3% (Atemporal Plausibility – (AP)) 73.1% (Temporal Plausibility (TP)) 5.4 % (Completeness – (Comp.))
			Patient Mobile phone number	52.0 % (AP) 55.8 % (TP) 55.8% (Comp.)
Supply chain management;1 ` Defining and monitoring policy	Completeness Plausibility	1. Completeness 2. Plausibility 3. Conformance	Prescriptions made	76.4%(Comp.) 99.4 % (AP) 97.1% (TP)
Patient health tracking; Healthcare decision support;	Plausibility Completeness	1. Plausibility 2. Conformance Completeness	Drug adherence percentages reported	88.4 % (AP) 62.4 % (TP) 45.1% (Comp.)
			BMI monitoring:	98.4%(AP) 81.0%(TP) 77.1%(Comp.)

Referral coordination;	Heights of patients	
	BMI monitoring: Weights of patients	99.8% (AP) 70.7% (TP) 99.9%(Comp.)
	HIV Viral load tracking	57.6% (AP) 97.9% (TP) 80.5%(Comp.)

As I explain my integration of the qualitative and quantitative results in the table above, I go back to my integration strategy where I introduced the concept of divergence or convergence of results. Divergence describes the situation where there is disagreement in the qualitative findings and the quantitative findings. The disagreement may be high expectation of a qualitative aim but finding low results in the quantitative aims or vice versa. Convergence on the other hand describes the situation where there is agreement between qualitative and quantitative results. A challenge that I faced were the acceptable thresholds in terms of an acceptable proportion for the quantitative results.

I ranked plausibility the highest importance for the data purpose of the use of data management and supply chain. Taking the current place of residence and patient mobile phone number shows that they have 40.3% and 52.0% plausibility proportions. These results show divergence between expected data quality characteristics, the importance and the observed data quality assessment proportions.

My comparison of integration for the next stakeholder reported purposes of data use of supply chain and defining policy shows some convergence between the qualitative results and the quantitative results. I make this observation considering that the ranked importance priority of the expected data quality characteristics matches their quality proportions. Although the prescription completeness proportion is 76.4% based on all patient visits, I make the observation that some of the visits may have been just patient assessments without a drug prescription being made and thus documented.

The third group of purposes of use (patient health tracking, healthcare decision support, and referral coordination) has differing overall comparisons for the different examples of variables for the quantitative analysis. I observed convergence in the stakeholder expected characteristic of completeness and its low importance priority and the quantitative results. My data quality assessment found only 45.1% of adherence records being complete.

On the other hand, for the last tracking variable of viral load, stakeholders anticipated the records to be plausible and complete. However, they placed higher importance on plausibility.

Integrating these qualitative findings and the quantitative results, however, show a divergence in that there is low plausibility of viral load results despite high importance placed on plausibility as a data quality characteristic.

Chapter 4 Discussion and Conclusion

I used mixed methods to answer the research question, “How do varying purposes of data use inform interventions for improving electronic data quality in ART Clinics in Malawi.”

My study established that the importance of data quality characteristics was not the same across different purposes of health data use. I made this observation during integration of my qualitative and quantitative results where I found that stakeholders placed different importance on data quality characteristics for different purposes of use. This finding is consistent with a study conducted in Haiti, where stakeholders working in HIV clinics attached different priorities to data variables that were selected for automated audit and feedback to improve the data quality of EHR records at multi-site ART clinics for HIV(35). The finding of different stakeholders having different importance for different characteristics further suggests that data quality is indeed fit for purpose(63), a position that was also adopted by the study in Haiti (63). The integrated findings in chapter 3 have implications on how the data collected in the EHR may be used. As I have explained in the results, I had convergent and divergent results upon integrating quantitative and qualitative results. Convergent results were those that had agreement between the views expressed in the qualitative findings and the quantitative data. On the other hand, divergent results had qualitative views that did not have quantitative data to support the expected views.

In the case of divergent findings, there may be a need to make decisions on the utility of the data that is available. In the example of patient tracking, there may be a need to further understand the extent of the plausibility of the records in order to meet the purpose of use of tracing patients and bringing them back into care.

On the other hand, in the case of convergent results between the purposes of data use, importance priorities for data quality characteristics and the observed proportions for data quality, there may be greater confidence to meet the stakeholder purposes of data use.

My convergent and divergent integration results for Aims 1 and 2 suggest implications in terms of key part of my research around the view that data quality is fit for purpose. In the divergent examples the view of data quality is fit for purpose may not hold since the user expectations do not match the assessment. Such a scenario, therefore, calls for a further examination in terms of the source of any divergence. A suggestion could be for example, based on the stakeholder purpose of data use, to modify how data is documented in hope of improving a given data quality characteristic.

For convergent results, the recommendation may be to continue current practices so that the observed data quality is maintained.

In the context of integrating my qualitative and quantitative results, I believe that my study results suggest a process towards answering my research question. The finding of my study on different purposes of use having different importance priorities for anticipated data quality characteristics and the observed data quality assessment motivates me not only to consider previous solutions that have been implemented as possible interventions to improve data quality. I believe my study suggests that interventions should not be a one size fits all but should take into consideration several factors. These factors include:

- Priorities of importance attached to data quality characteristics for different stakeholder purposes of data use.
- The observed data quality after conducting a quantitative data quality assessment
- Any reported barriers or facilitators that may contribute to the observed data quality.

Identifying and implementing interventions that improve data quality should consider the factors listed above. However, since my study did not have the scope and feasibility to identify interventions that would be appropriate to improve poor data quality, I reviewed literature to identify interventions that have been implemented in similar settings.

4.1 What are some interventions that have been implemented to improve data quality

Introduction

I conducted a literature review on interventions that have been implemented to improve data quality by searching several databases such as Google Scholar, PubMed and other reports recommended to me by my research committee. I proceeded on this route with the understanding that there have been previously reported challenges that result in poor data quality and efforts to improve poor data quality. Studies by Forster et al., Kumar et al., and Noureldin et al. (92–94), mentioned environmental challenges such as infrastructure, knowledge challenges such as the limited capacity to use the EHRs and inadequate training of healthcare providers as some contributors towards poor data quality in resource-constrained settings. These studies suggested investing in interventions to tackle the challenges they established to improve data quality.

Methods

I conducted a literature review to identify journal reports with interventions to improve data quality in low-income countries using the Preferred Reporting Items for Systematic and Meta Analyses (PRISMA) guideline (95). I used the search terminology listed in the PRISMA checklist in Appendix 9 to identify articles for review in two databases namely Google Scholar and PubMed:

I list the inclusion and exclusion criteria for the articles that I extracted for literature review in Table 4.1.

Table 4.1:Inclusion and exclusion criteria for systematic review of interventions to improve data quality in LMICs

Item	Inclusion Criteria	Exclusion criteria
Settings	Low- or Middle-Income Country (LMIC) setting in Sub-Saharan Africa	Studies whose settings are not in LMICs in Sub-Saharan Africa
Type of medium used	Data captured in an electronic health record or other electronic health media.	Studies that do not use electronic health records or other electronic media
Time of publication	Published from 2008 onwards (Taking into consideration that the EHR scale up in Malawi started in 2007)	Studies published earlier than 2008
Intervention	Papers that describe interventions to improve data quality.	No description of any intervention to improve data quality.
Language	Studies conducted in English	Language other than English

Results

I found a total of 122 articles after conducting my search in the databases specified above. I narrowed down articles to 15 to review further in terms of the content regarding my interest on interventions to improve data quality. I have included a PRISMA flow diagram in Appendix 8 that details how I arrived at the articles that I present in Table 4.2. I categorized the different data quality interventions reported in the journal articles into the following main categories:

1. Supervision as a data quality improvement intervention: I classified articles that reported the intervention of a supervisor checking the quality of the data into this intervention.
2. Automated data quality assessments: I classified interventions that employed automated database queries for data quality assessment on existing electronic data using the mechanism of dashboards or reports into this category.
3. Capacity building via training and staff recruitment: I classified all interventions that used training on data quality improvement approaches or used recruitment of special cadres to improve data quality into this intervention.
4. Computerization of the documentation process or use of computer equipment: I classified all articles that reported the design and implementation of technology interventions such as EHRs, mobile health applications to improve data quality in one way or another.
5. Modification of data collection tools: I classified interventions that used change of the structure of the data collection tools as the intervention to improve data quality.
6. Development of standard operating procedures: I classified interventions that used job aides to improve data quality.

I observed that most articles reported that they implemented a combination of interventions to improve data quality. I use the numeric numbers above to represent the group of combinations present in the studies that are in the Table 4.2.

Table 4.2: List of published interventions to improve data quality in LMICs

First Author	Year	Title	Intervention Used	Intervention Categories
Scobie M. Heather et al.(96)	2020	Improving the quality of Immunization and surveillance data: Summary report of the Working Group of the Strategic Advisory Group of Experts on Immunization.	New electronic platforms; Routine monitoring of data quality; Automated validation checks and analyses; Data triangulation.	4,1,2
Karuri et al.(97)	2014	DHIS2: The Tool to Improve Health Data Demand and Use in Kenya.	Using available data will improve data quality; Computerization of records improves data quality.	4,1
Abiy et al. (98)	2018	A Comparison of Electronic Medical Record Data to Paper Records in Antiretroviral Therapy Clinic in Ethiopia: What is affecting the Quality of the Data?	Trainings and continuous support to health workers; Implementation should be single EHR.	3,4
Mphatswe et al. (99)	2012	Improving Public health information: a data quality intervention in Kwa Zulu-Natal, South Africa	Specific training for health-care workers on importance of public health information; Monthly data reviews and feedback; Regular audits.	3,1

Mutale Wilbroad et. al(100).	2013	Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative	Simplified data capture and streamlined reporting designed to lead to more time to focus on quality; Ongoing feedback on missing data and outliers, and ongoing data quality assessments; Quarterly data quality audits and automated data quality report based on logic errors generated when administrative and clinical reports are generated; Facility supervisors review community health agent reports and provide data feedback; Standardized protocol for data capture with real time query of data gaps; subsequent follow up during monitoring visits.	5,1,2,6
Puttkammer Nancy et al(35).	2017	Identifying priorities for data quality improvement within Haiti's iSante EMR system	Automated data quality features via validation; Data quality dashboard over time on automated queries.	2
Githinji Sophie et al (101).	2017	Completeness of Malaria Indicator data reporting via the District Health Information System .	Technical expertise from the MoH; Strong and robust monitoring via biannual surveys; Regular data reviews and technical working groups at both national and county levels.	1
Kiberu Vincent Micheal et al(102).	2014	Strengthening district-based health reporting through the district health management information software system: the Ugandan experience.	Computerization improves data quality reporting; Continued training of HMIS focal persons.	4,3
Wagenaar H. Bradley et al(103)	2015	Effects of a health information system data quality intervention on concordance in Mozambique:time-series analyses from 2009-2012 (24)	Facility based audits, targeted training and supervision, equipment purchase/maintenance, data audit and feedback meetings.	1,3,4
Muthee Veronica et al(37).	2018	The impact of routine data quality assessments on electronic medical record data quality in Kenya	Routine data quality assessments; Audit and feedback to health facilities.	1
Gimbel Sarah et al(104).	2017	Improving data quality across 3 sub-Saharan African countries using the Consolidated Framework for Implementation Research (CFIR):results from the African Health Initiative.	Routine data audits linked to on going on the job training; Performance review meetings.	1,3

Mpofu Mulamuli et. al(105)	2014	Strengthening monitoring and evaluation (M&E) and building sustainable health information systems in resource limited countries: lessons learned from an M&E task-shifting initiative in Botswana	Capacity building by recruiting new roles.	3
Admon A.J. et al (106).	2013	Assessing and improving data quality from community health workers: a successful intervention in Neno Malawi	Task shift role of data aggregation; Localization of data collection forms.	3,6
Gadabu O.J. et al (39).	2011	Is transcription of data on antiretroviral treatment from electronic to paper-based registers reliable in Malawi?	Computerization of paper-based records.	4
Elizabeth Hazel et al(107).	2017	Using Data to Improve Programs: Assessment of Data Quality and Use Intervention Package for Integrated Community Case Management in Malawi	Provision of calculators to improve accuracy of aggregation; Training and supervision.	4,3

I discuss each of the interventions discussed in the Table 4.2 in the context of my study in the sections below. I further discuss how each of these interventions would be relevant for purposes of my study.

4.*1.1 Supervision as a data quality improvement intervention.*

Health facility data supervision approaches have been considered in many settings similar to Malawi (37,100). The general idea is that the EHR data use stakeholders need to be accountable to higher authorities via mechanisms such as routine data quality analyses. In Malawi, supervision has mostly been on HIV/AIDS data for patients enrolled in the ART program. The EHR programs at health facilities maintain a paper-based system as well as an electronic system. This model of implementing EMRs has been reported in other settings such as in Ethiopia and Mozambique (98,103). The data collected via the EHR is compared against the data collected on paper-based records with the paper-based records acting as the gold standard most of the times (39). In settings outside Malawi, others have pointed out the resource intensive nature of supervision as a method for ensuring data quality which results in an expensive process of ensuring data quality (35). Regardless of cost being a barrier, the intervention of routine supervision is very popular and shows some results in terms of improving data quality as shown by the frequency it has been shown in Table 4.2.

4.1.2 Automated data quality assessments of EHR data

Some settings that have an EHR have proposed implementation of automated queries to assess data quality and generate data quality reports (35). These reports would give feedback to different stakeholders on variables that they should focus on and improve data quality on. For example, I came across a paper in an HIV clinic in Haiti where they identified important variables for different stakeholders to improve data quality. In this report, they advocated for implementation of these reports although they did not actually implement them.

However, in another intervention in a widely used aggregate data documentation system in settings similar to Malawi called the District Health Information System version 2 (DHIS 2), data quality assessment reports are part of the system. In this setup, the data quality reports focus on completeness and timeliness of aggregate data reports (108). Completeness and timeliness are usually presented via reporting dashboards in this piece of reporting software. For example, the use of the reporting dashboards in DHIS version 2 in Sierra Leone during data review meetings led to an increase of completeness of selected indicators (97).

I thus would recommend implementation of automated data quality tools in the setting of Malawi to improve the EHR data quality for different variables within the HIV/AIDS program. Given the experience, in other countries, such as the report carried out by Gimbel et. al (104), I would recommend combining the intervention of the automated reports with meeting reviews so that there should be stakeholder engagement. This would be important to ensure by-in by all parties involved as reported by Gimbel et al (104).

4.1.3 Capacity building via training and staff recruitment

Capacity building via targeted trainings is a common intervention for data quality improvement that I encountered in many of the journal articles in Table 4.2. Capacity building has been adopted for different types of data systems including aggregate and EHR based systems similar to the one I based my study in Malawi. For example in two settings with an EHR in ART settings in Kenya and Ethiopia, training was reported to result in improved data quality (37,98). A similar observation was made for aggregate data systems, where the introduction of dedicated cadres to enter collated data into the DHIS version 2 increased completeness of reported indicators (1,2). Training has further been emphasized as an appropriate intervention

to improve understanding of the data collected and techniques for analyzing the data to improve data use with the hope that this would in turn improve data quality. In a study that assessed interventions to improve data quality in three sub-Saharan African countries namely Mozambique, Zambia and Rwanda, an implementation science framework called the Consolidated Framework for Implementation Research was used to have group discussions that would lead the way into identifying appropriate interventions (104). On-going on-the-job training was emphasized as an important intervention to ensure data quality.

4.*1.4 Computerization of the documentation process or the introduction of computer equipment*

Some challenges in data quality are a result of errors that arise due to the limitation of not being able to build in computation or validation rules in paper-based system. I reviewed papers that have used different levels of automation to improve data quality at different levels. In a study to improve accuracy of data collected by community health workers for the Integrated Management of Childhood illnesses program, the Ministry of Health and its partners introduced calculators to improve the accuracy of aggregating data before documenting this data into facility reporting form (107). Despite, this basic intervention, there was some improvement in the accuracy of the aggregated data. In other more complex interventions, EHRs have been implemented with the observation of reduction in missing patient records which usually result due to damage of physical patient records (39). Computerization is however, also sought for as an intervention to improve data quality because of the possibility to build validation rules in the EHRs as seen in the implementation of the EHR in Rwanda (1). In the case of my study, the observation of high degrees of data quality for some variables such as weight, may be an indication that there are robust built-in rules in the EHR that ensure the plausibility of these variables.

4.1.5 Modification of data collection tools

Other interventions have required that current tools that are used for data collection should be modified so that they easy to understand by healthcare providers and also that they have clearer workflows (100). For example, an intervention to improve the data quality of reported data translated data collection forms from English into the local language Chichewa in Malawi to ease understanding by community health workers (106). It is reported that this change, coupled with training resulted in higher completeness of forms submitted for reporting from the communities. Although, localization could be an example of change in structure intervention to improve the EHR data quality, I would aim at identifying other areas where change in the design would lead to better data quality. I would not localize at the level of the clinic since the Clinical Practice Guidelines at the level where I conducted my study are in English. Structural interventions that could be relevant for my study would be those around streamlining the documentation process so that plausible and complete data are entered in the EHR.

4.1.6 Development of standard operating procedures

Job aides have also been reported as interventions that result in improved data quality. This has been reported as presenting standardized protocols so that healthcare providers know exactly how to go about documentation (100). This may be similar to training; however, the difference may be with giving the healthcare workers job aides that they may use to fulfill data quality improvement interventions. One approach was to give community healthcare workers templates for making charts to achieve data use as well as to identify easily gaps in data quality and address the gaps there and then (106). As I established purposes of data use in my study, reporting was one of the purposes of data use. Although the EHR has some built in reporting tools, introduction of job aides that would go beyond only how to use the current features of the EHR, but also to improve data quality at the health facilities would help improve the overall data quality.

4. 2 How do the varying purposes of stakeholders' purposes of data use inform interventions for improving data quality?

I suggest that stakeholders' purposes of data use, as have been confirmed with WHO classification of digital health interventions in my study may follow the process illustrated in figure 1 below:

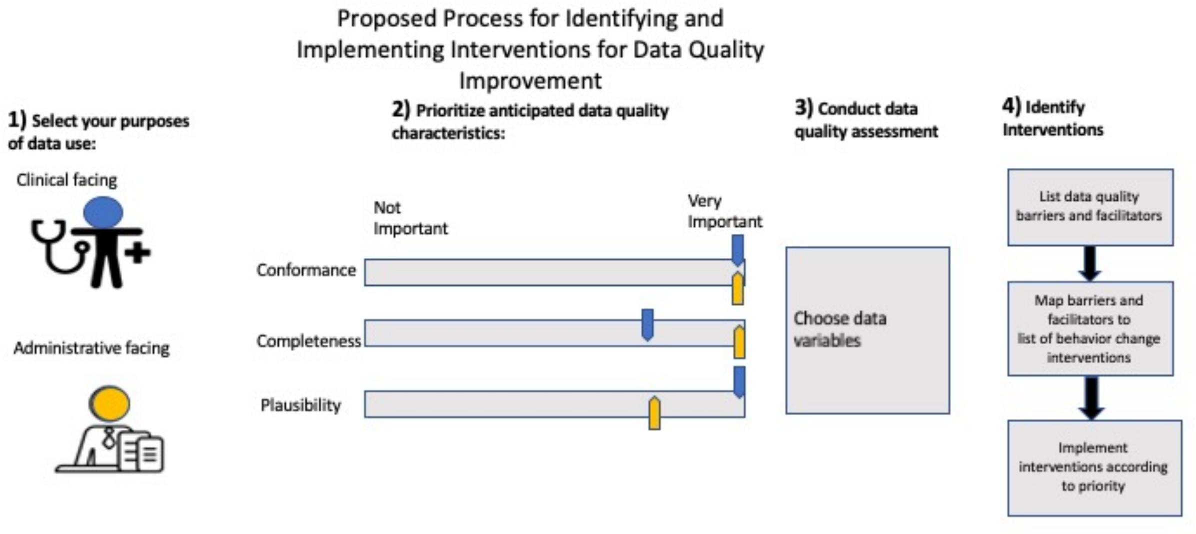


Figure 4.1: Suggested Process for identifying and Implementing Interventions for improving data quality in ART clinics

Each intervention that may be implemented to improve data quality should be done considering the priority that a stakeholder has attached to the purpose of use being considered. The implementation should also take into consideration the observed data quality following a data quality assessment. The observed data quality assessment results alone are regarded as intrinsic data quality(63).

To illustrate the suggested process in Figure 4.1, I give the example of the stakeholder clinical-facing purpose of the use of tracking patient adherence which placed high importance for the plausibility data quality characteristic. I found that it had high plausibility but low completeness when I conducted the data quality assessment. In this example, there would be less emphasis to identify interventions to improve data quality for this variable immediately since the stakeholders' purpose of data use, anticipated data quality characteristics and the importance ranking converge with plausibility data quality.

On the other hand, the stakeholder client-facing purpose of use of viral load monitoring had high importance priority ranking for plausibility expectation and poor data quality in terms of atemporal plausibility. This divergence in the results emphasizes that efforts should be made to identify barriers towards the observed data quality and identify interventions to improve the observed data quality.

4.3 Strengths and Weaknesses

4.3.1 Strengths of My Study

This study used a guideline for reporting mixed methods from an article on the Enhancing the Quality and Transparency of Health Research (EQUATOR) network called “Reporting mixed research in the field of counseling and beyond” hence has followed best practice in conducting and reporting mixed methods research(109). I have attached the guideline as Appendix 10. This study also had a large patient population selected from different health levels; thus, it was representative of the health facility types in Malawi. Therefore, the results of the study could apply to many other sites with similar settings in the country.

4.3.2 Weaknesses of the study

Nonresponse from a critical donor participant in the study was a weakness in the qualitative part as that participant plays a significant role in the rollout of the EHR in Malawi. To address this weakness, future work may consider engaging the donor with the results of this study to get their feedback. This weakness affected my first two studies which were fell under Aim 1. I take this as a major weakness given the background that EHR investments in Malawi are funded by this main donor.

Trustworthiness for the qualitative arm of my research was also a major weakness due to the fact that I was the only person that conducted the coding. I tried to improve the trustworthiness of my coding for Aim 1 by among other things taking some time off between coding sessions. Multiple coders are recommended so that a researcher may report Intercoder Reliability (ICR) alongside reported codes.

Another limitation I faced was the reduction in the number of participants for Aim 1c due to feasibility, related to the delay in completing the questionnaire used for this aim. The reduced number reduces how representative the results are, because I did not get views from other stakeholders who may have attached different priorities for different data quality characteristics.

Another major limitation was the failure to get viral load datasets for health centers and district hospitals because they had not been documented in the EHR. To address this weakness, I selected records from Tertiary level hospitals since these had complete records. In future work, efforts would be made to replicate the study for the other health facilities.

4.4 Possible Future Work

4.4.1 Conduct follow up implementation study using the data quality assessment results to identify barriers and facilitators to data quality improvement interventions

The data quality assessment results may be a tool that can help in bringing discussions among different stakeholders to identify barriers or facilitators of observed data quality. This would be achieved by presenting the data quality assessment results to different stakeholder groups. I would use implementation science frameworks to understand any barriers or facilitators that may have contributed to the observed data quality, specifically the Theoretical Domains Framework (TDF) and the Consolidated Framework for Implementation Research (CFIR) to identify individual and group barriers or facilitators respectively. I would use the TDF with the observation that there may be some factors that affect individual stakeholders as barriers or facilitators that result in poor data quality. These factors would best be understood by the TDF since this framework has dimensions and constructs that are applicable at the individual level. In the case of my study setting, there may be for example different cadres that have different individual views on their data documentation roles versus their professional roles. That being said, I also recognize that the individual stakeholders usually work in a group under different organizations. This means that there would be organizational or group facilitators or barriers towards data quality such as workplace regulations or culture that would need to be established. The CFIR has dimensions and constructs that are applicable to groups or organizations.

Having done this, I would then use the Behavior Change Technique Taxonomy to identify interventions for improving data quality(57). I would consider these interventions together with appropriate implementation strategies to add more structure and scalability of the interventions to additional health facilities. I would use the data quality assessment proportions as performance measures to monitor any changes in data quality over an agreed-upon time with the different stakeholders. I propose to conduct this as an implementation study so that changes may be assessed over a given period and applied to other health facilities.

4.4.2 Automate Data Quality Assessment in health facilities with the EHR

Future work for this would be to automate the data quality assessment that I have conducted in this study and implement it in all health facilities that have Baobab Health Trust EHR. This may be done by packaging the processes of data extraction and analysis into one software program that may then be installed in all health facilities with the EHR. The software would then be set to run at a scheduled time and disseminate the results to selected stakeholders. This would result in building blocks I believe would not only give a state of data quality but would be an intervention to gradually and consistently improve data quality as routine supervision since manual audits are labor-intensive and expensive (8,35).

4.4.3 Introduce automated tools for addressing identified poor data quality issues.

Understanding the nature of missing data in case of completeness data quality characteristic may assist in addressing the poor data quality. Missing data may be classified as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR).

I observed that some variables are missing not at random (110), such as patient physical addresses that consistently have village and district but do not have any traditional authority value. In this case the missing address component may most likely be due to the location where the patient lives (urban vs. rural). This is an incomplete address that may be fixed by imputing the missing traditional authority from lookup tables. Another example could be a patient visit with missing height or weight value. Again, in this example the missing data may be imputed from previous weight or height values for this individual or from weight and height distributions of patients in the cohort at that clinic if there are no previous values for an individual.

4.4.4 Develop a website that shows DQA metrics for each health facility with an ART EHR

Robust data quality increases data use (35). I have additionally demonstrated that different stakeholders using EHR ART data in Malawi have different purposes of data use. Therefore, consideration for future work is to build health information infrastructure for assessing data quality that recognizes different stakeholders. Given the process illustrated in figure 1, the job might entail designing a web-based tool for the MoH where stakeholders may select their purposes of data use and run a data quality assessment. To ensure that the tool meets various user needs, further steps would be employed to gather requirements around the infrastructure. These requirements would then inform use cases that would map to the data workflows of different stakeholders. This will give them confidence in terms of how they want to use their data. This kind of infrastructure could be customized for other similar settings and used by similar stakeholders.

4.5 Conclusion

My study has made two main contributions to understanding data quality in ART clinics for HIV patients in low-income countries that have EHR infrastructure. The contributions are:

1) Process for conducting data quality assessment and follow up process to identify interventions

   My study has contributed methods for conducting data quality assessments and identifying interventions to improve any observed data quality. Taking the qualitative and quantitative findings, my research presented an approach to conducting data quality assessment and a process for answering the research question of identifying interventions for improving data quality in resource-constrained settings.

2) Provide data quality assessment software code that may be adapted to present performance measures for displaying 90-90-90

   The second contribution is a quantitative analysis program that may be re-used to understand data quality for other health facilities in Malawi. The source code is freely available on GitHub, although it requires customization.

Lastly, the previously reported challenges of EHR data quality in ART clinics in Malawi present a threat to continued rollout of the EHR as infrastructure for the ART program in the country. This threat could lead to loss of trust(8) in EHR data. EHR infrastructure may be a crucial part of creating a learning health system for ART clinics, and my study may provide necessary processes to support data quality that is fit for purpose in Malawi.

References

1. Briet S, Martinet L. What is Documentation?: English Translation of the Classic French Text. Scarecrow Press; 2006. 88 p.

2. Peter Waegemann, Claudia Tessier. Healthcare Documentation: A Report on Information Capture and Report Generation [Internet]. 2007 Jul [cited 2018 Jan 31]. Available from: http://www.providersedge.com/ehdocs/ehr_articles/Healthcare_Doc-Report_on_Information_Capture_and_Report_Generation.pdf

3. Jungraithmayr W, Weder W. Chest surgical disorders in ancient Egypt: evidence of advanced knowledge. Ann Surg. 2012 Mar;255(3):605–8.

4. Kamp MA, Tahsim-Oglou Y, Steiger H-J, Hänggi D. Traumatic brain injuries in the ancient Egypt: insights from the Edwin Smith Papyrus. J Neurol Surg Part Cent Eur Neurosurg. 2012 Aug;73(4):230–7.

5. The Joint Commission,Joint Commission International, World Health Organization. Patient Identification [Internet]. World Health Organization; 2007. Available from: https://www.who.int/patientsafety/solutions/patientsafety/PS-Solution2.pdf

6. WPRO | Medical Records Manual: A Guide for Developing Countries [Internet]. WPRO. [cited 2018 May 9]. Available from: http://www.wpro.who.int/publications/PUB_9290610050/en/

7. Core Indicators for National AIDS Programmes - Guidance and Specifications for Additional Recommended Indicators. :45.

8. Makombe SD, Hochgesang M, Jahn A, Tweya H, Hedt B, Chuka S, et al. Assessing the quality of data aggregated by antiretroviral treatment clinics in Malawi. Bull World Health Organ. 2008 Apr;86(4):310–4.

9. Medical Records and Documentation Standards [Internet]. [cited 2017 Nov 8]. Available from: https://provider.ghc.org/open/render.jhtml?item=/open/workingWithGroupHealth/records-standards.xml

10. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. The Learning Healthcare System: Workshop Summary [Internet]. Olsen L, Aisner D, McGinnis JM, editors. Washington (DC): National Academies Press (US); 2007. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: http://www.ncbi.nlm.nih.gov/books/NBK53494/

11. English M, Irimu G, Agweyu A, Gathara D, Oliwa J, Ayieko P, et al. Building Learning Health Systems to Accelerate Research and Improve Outcomes of Clinical Care in Low- and Middle-Income Countries. PLOS Med. 2016 Apr 12;13(4):e1001991.

12. The Learning Healthcare Project The Learning Health Care Project [Internet]. [cited 2020 Feb 18]. Available from: http://www.learninghealthcareproject.org/section/background/learning-healthcare-system

13. Friedman C, Rubin J, Brown J, Buntin M, Corn M, Etheredge L, et al. Toward a science of learning systems: a research agenda for the high-functioning Learning Health System. J Am Med Inform Assoc. 2015 Jan 22;22(1):43.

14. Edwards PN. Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems. Mod Technol. 2003 Jan 1;185–225.

15. Douglas GP, Gadabu OJ, Joukes S, Mumba S, McKay MV, Ben-Smith A, et al. Using touchscreen electronic medical record systems to support and monitor national scale-up of antiretroviral therapy in Malawi. PLoS Med. 2010;7(8).

16. Braaf S, Manias E, Riley R. The role of documents and documentation in communication failure across the perioperative pathway. A literature review. Int J Nurs Stud. 2011 Aug 1;48(8):1024–38.

17. Pirkle CM, Dumont A, Zunzunegui M-V. Medical recordkeeping, essential but overlooked aspect of quality of care in resource-limited settings. Int J Qual Health Care. 2012 Dec 1;24(6):564–7.

18. Tsegaye AT, Wubshet M, Awoke T, Addis Alene K. Predictors of treatment failure on second-line antiretroviral therapy among adults in northwest Ethiopia: a multicentre retrospective follow-up study. BMJ Open [Internet]. 2016 Dec 8;6(12). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5168604/

19. Cusack CM, Hripcsak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, et al. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting. J Am Med Inform Assoc. 2013 Jan 1;20(1):134–40.

20. Gettel CJ, Merchant RC, Li Y, Long S, Tam A, Marks SJ, et al. The Impact of Incomplete Nursing Home Transfer Documentation on Emergency Department Care. J Am Med Dir Assoc [Internet]. 2018 Oct 29 [cited 2019 Jan 14]; Available from: http://www.sciencedirect.com/science/article/pii/S1525861018304997

21. Jahn A, Harries AD, Schouten EJ, Libamba E, Ford N, Maher D, et al. Scaling-up antiretroviral therapy in Malawi. Bull World Health Organ. 2016 Oct 1;94(10):772–6.

22. Martin C Were CS, William M. Tierney JJM, Paul G. Biondich XL, Sylvester Kimaiyo BWM. Evaluation of computer-generated reminders to improve CD4 laboratory monitoring in sub-Saharan Africa: a prospective comparative study. J Am Med Inform Assoc. 2011;1–6.

23. Medicine I of. Computer-Based Patient Record: An Essential Technology for Health Care [Internet]. 1969 [cited 2017 Nov 9]. Available from: https://www.nap.edu/catalog/18459/computer-based-patient-record-an-essential-technology-for-health-care

24. Edward H. Shortliffe JJC. Biomedical Informatics, Computer Applications in Health Care and Biomedicine. Health Informatics. Third Edition. New York: Springer; 2006.

25. Adler-Milstein J, Ronchi E, Cohen GR, Winn LAP, Jha AK. Benchmarking health IT among OECD countries: better data for better policy. J Am Med Inform Assoc JAMIA. 2014 Jan;21(1):111–6.

26. Martin C. W, Eric M. M. Ethics of Implementing Electronic Health Records in Developing Countries: Points to Consider [Internet]. 2011 [cited 2012 Nov 21]. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3243215/

27. Oluoch T, de Keizer NF. Evaluation of Health IT in Low-Income Countries. Stud Health Technol Inform. 2016;222:324–35.

28. Hamish SF Fraser JB. Implementing medical information systems in developing countries, what works and what doesn't. In AMIA; 2010. p. 232–6.

29. Blaya JA, Fraser HSF, Holt B. E-health technologies show promise in developing countries. Health Aff Proj Hope. 2010 Feb;29(2):244–51.

30. Braitstein P, Einterz RM, Sidle JE, Kimaiyo S, Tierney W. "Talkin' About a Revolution": How Electronic Health Records Can Facilitate the Scale-Up of HIV Care and Treatment and Catalyze Primary Care in Resource-Constrained Settings. JAIDS J Acquir Immune Defic Syndr. 2009 Nov;52:S54.

31. Kang'a S, Puttkammer N, Wanyee S, Kimanga D, Madrano J, Muthee V, et al. A national standards-based assessment on functionality of electronic medical records systems used in Kenyan public-Sector health facilities. Int J Med Inf. 2017 Jan;97:68–75.

32. Hamilton C. The WHO-ITU national eHealth strategy toolkit as an effective approach to national strategy development and implementation. Stud Health Technol Inform. 2013;192:913–6.

33. deRiel E, Puttkammer N, Hyppolite N, Diallo J, Wagner S, Honoré JG, et al. Success factors for implementing and sustaining a mature electronic medical record in a low-resource setting: a case study of iSanté in Haiti. Health Policy Plan. 2018 Mar 1;33(2):237–46.

34. Puttkammer N, Baseman JG, Devine EB, Valles JS, Hyppolite N, Garilus F, et al. An assessment of data quality in a multi-site electronic medical record system in Haiti. Int J Med Inf. 2016 Feb 1;86:104–16.

35. Puttkammer N, Pettersen K, Hyppolite N, France G, Valles JS, Honoré JG, et al. Identifying priorities for data quality improvement within Haiti′s iSanté EMR system: Comparing two methods. Health Policy Technol. 2017 Mar;6(1):93–104.

36. Muthee V, Bochner AF, Kang'a S, Owiso G, Akhwale W, Wanyee S, et al. Site readiness assessment preceding the implementation of a HIV care and treatment electronic medical record system in Kenya. Int J Med Inf. 2018 Jan 1;109:23–9.

37. 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 Apr 18;13(4):e0195362.

38. Landis-Lewis Z, Manjomo R, Gadabu OJ, Kam M, Simwaka BN, Zickmund SL, et al. Barriers to using eHealth data for clinical performance feedback in Malawi: A case study. Int J Med Inf. 2015 Oct;84(10):868–75.

39. Gadabu OJ, Munthali CV, Zachariah R, Gudmund-Hinderaker S, Jahn A, Twea H, et al. Is transcription of data on antiretroviral treatment from electronic to paper-based registers reliable in Malawi? Public Health Action. 2011 Sep 21;1(1):10–2.

40. A Comparison of Data Quality Assessment Checks in Six Data Sharing Networks. - PubMed - NCBI [Internet]. [cited 2018 Oct 24]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/29881733

41. 2018 Malawi Population and Housing Census [Internet]. [cited 2019 Mar 9]. Available from: http://www.nsomalawi.mw/index.php?option=com_content&view=article&id=226:2018-malawi-population-and-housing-census&catid=8:reports&Itemid=6

42. Size of Great Lakes in Africa 2016 | Statistic [Internet]. Statista. [cited 2019 Mar 1]. Available from: https://www.statista.com/statistics/574073/size-of-lakes-in-the-great-lakes-of-africa/

43. Zant W. If smallholder farmers have access to the world market: the case of tobacco marketing in Malawi. Eur Rev Agric Econ [Internet]. [cited 2020 Feb 8]; Available from: https://academic.oup.com/erae/advance-article/doi/10.1093/erae/jbz039/5628209

44. Overview [Internet]. World Bank. [cited 2019 Mar 9]. Available from: http://www.worldbank.org/en/topic/poverty/overview

45. Phiri G. Poverty levels extreme in Malawi—Unctad [Internet]. The Nation Online. 2017 [cited 2018 May 12]. Available from: http://mwnation.com/poverty-levels-extreme-in-malawi-unctad/

46. Malawi Literacy [Internet]. 2016 [cited 2020 Feb 8]. Available from: http://uis.unesco.org/en/country/mw

47. Mwale B. HIV/AIDS in Malawi. Malawi Med J J Med Assoc Malawi. 2002 Sep;14(2):2–3.

48. Makombe SD, Jahn A, Tweya H, Chuka S, Yu JK-L, Hochgesang M, et al. A national survey of the impact of rapid scale-up of antiretroviral therapy on health-care workers in Malawi: effects on human resources and survival. Bull World Health Organ. 2007 Nov;85(11):851–7.

49. Libamba E, Makombe SD, Harries AD, Schouten EJ, Yu JK-L, Pasulani O, et al. Malawi's contribution to "3 by 5": achievements and challenges. Bull World Health Organ. 2007 Feb;85(2):156–60.

50. Government of Malawi Ministry of Health. Integrated HIV Program Report July-September 2018 [Internet]. Department of HIV AIDS; 2019 Jan. Available from: https://www.hiv.health.gov.mw/

51. HIV and AIDS in Malawi [Internet]. AVERT. 2015 [cited 2018 May 12]. Available from: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/malawi

52. Geubbels E, Bowie C. Epidemiology of HIV/AIDS in adults in Malawi. Malawi Med J J Med Assoc Malawi. 2006 Sep;18(3):111–33.

53. Test and Treat offers Malawi hope to reduce new HIV infections [Internet]. Malawi Nyasa Times - Malawi breaking news in Malawi. 2015 [cited 2018 May 12]. Available from: https://www.nyasatimes.com/test-and-treat-offers-malawi-hope-to-reduce-new-hiv-infections/

54. Malawi Ministry of Health. Malawi Health Sector Strategic Plan II [Internet]. [cited 2018 Dec 14]. Available from: http://www.health.gov.mw/index.php/policies-strategies?download=14:malawi-health-sector-strategic-plan-2011-2016

55. BaobabHealthTrust [Internet]. GitHub. [cited 2019 Jan 23]. Available from: https://github.com/BaobabHealthTrust

56. Msosa Y, Densmore M, Maria Keet C. Towards an architectural design of a guideline-driven EMR system. In 2015. p. 1–4.

57. Landis Lewis Z, Mello-Thoms C, Gadabu OJ, Gillespie EM, Douglas GP, Crowley RS. The feasibility of automating audit and feedback for ART guideline adherence in Malawi. J Am Med Inform Assoc JAMIA. 2011;18(6):868–74.

58. Chirwa Z, Kayambo F, Oseni L, Plotkin M, Hiner C, Chitsulo C, et al. Extending beyond Policy: Reaching UNAIDS' Three "90"s in Malawi. Front Public Health. 2018;6:69.

59. Tweya H, Feldacker C, Gadabu OJ, Ng'ambi W, Mumba SL, Phiri D, et al. Developing a point-of-care electronic medical record system for TB/HIV co-infected patients: experiences from Lighthouse Trust, Lilongwe, Malawi. BMC Res Notes [Internet]. 2016 Mar 5;9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4779573/

60. Definition of STAKEHOLDER [Internet]. [cited 2019 Jan 15]. Available from: https://www.merriam-webster.com/dictionary/stakeholder

61. Chen H, Hailey D, Wang N, Yu P. A Review of Data Quality Assessment Methods for Public Health Information Systems. Int J Environ Res Public Health. 2014 May;11(5):5170–207.

62. ISO 25012 [Internet]. [cited 2020 Feb 9]. Available from: https://iso25000.com/index.php/en/iso-25000-standards/iso-25012

63. Kahn MG, Callahan TJ, Barnard J, Bauck AE, Brown J, Davidson BN, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS Wash DC. 2016;4(1):1244.

64. Y. Wang R, M. Strong D. Beyond Accuracy: What Data Quality Means to Data Consumers. J Manag Inf Syst. 1996 Mar 1;12:5–33.

65. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. eGEMs [Internet]. [cited 2018 Oct 24];5(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5983018/

66. WHO | Data Quality Review (DQR) Toolkit [Internet]. WHO. [cited 2019 Jan 4]. Available from: http://www.who.int/healthinfo/tools_data_analysis/dqr_modules/en/

67. WHO | Classification of digital health interventions v1.0 [Internet]. WHO. [cited 2018 Oct 2]. Available from: http://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/

68. Guidelines for conducting and reporting mixed research in the field of counseling and beyond | The EQUATOR Network [Internet]. [cited 2020 Jan 31]. Available from: https://www.equator-network.org/reporting-guidelines/guidelines-for-conducting-and-reporting-mixed-research-in-the-field-of-counseling-and-beyond/

69. The planning and reporting of mixed methods studies on the built environment and health - ScienceDirect [Internet]. [cited 2020 Jan 19]. Available from: https://www.sciencedirect.com/science/article/pii/S0091743519302282

70. Guidelines for Conducting and Reporting Mixed Research in the Field of Counseling and Beyond - Leech - 2010 - Journal of Counseling & Development - Wiley Online Library [Internet]. [cited 2020 Jan 19]. Available from: https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.1556-6678.2010.tb00151.x

71. Writing Mixed Research Reports - Nancy L. Leech, 2012 [Internet]. [cited 2020 Jan 14]. Available from: https://journals-sagepub-com.proxy.lib.umich.edu/doi/10.1177/0002764211433800

72. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Ann Fam Med. 2015 Nov;13(6):554–61.

73. Evaluation Methods in Biomedical Informatics | Charles P. Friedman | Springer [Internet]. [cited 2017 May 15]. Available from: http://www.springer.com/us/book/9780387258898

74. Muula AS. Case for Clinical Officers and Medical Assistants in Malawi. Croat Med J. 2009 Feb;50(1):77–8.

75. Holloway I. Basic Concepts for Qualitative Research. Wiley; 1997. 214 p.

76. Buy NVivo Now | NVivo [Internet]. [cited 2019 Oct 1]. Available from: https://www.qsrinternational.com/nvivo/nvivo-products/nvivo-12-pro

77. Ryan G, Bernard H. Techniques to Identify Themes. Field Methods - FIELD METHOD. 2003 Feb 1;15:85–109.

78. Paired Comparison Technique. In: Encyclopedia of Survey Research Methods [Internet]. 2455 Teller Road, Thousand Oaks California 91320 United States of America: Sage Publications, Inc.; 2008 [cited 2018 Dec 3]. Available from: http://methods.sagepub.com/reference/encyclopedia-of-survey-research-methods/n362.xml

79. Pairwise comparisons [Internet]. [cited 2020 Apr 29]. Available from: https://www.math.upenn.edu/~deturck/m170/wk10/lecture/vote5.html

80. Malawi Ministry of Health. Malawi Integrated Clinical HIV Guidelines 3rd Edition (2016) [Internet]. 2016 [cited 2017 Feb 7]. Available from: http://www.hiv.health.gov.mw/index.php/our-documents

81. Office of U.S. Global AIDS Coordinator, and the Bureau of Public Affairs, U.S. State Department. Data for Impact: Accelerating Toward Achieving 90-90-90 & Epidemic Control [Internet]. [cited 2017 Jan 19]. Available from: https://www.pepfar.gov/priorities/data/index.htm

82. (12) (PDF) Baseline Trachoma Mapping in Malawi with the Global Trachoma Mapping Project (GTMP) [Internet]. ResearchGate. [cited 2018 Dec 14]. Available from: https://www.researchgate.net/publication/279967133_Baseline_Trachoma_Mapping_in_Malawi_with_the_Global_Trachoma_Mapping_Project_GTMP/figures?lo=1

83. Q and A With Executive Director of BHT Martha Kwataine - [Internet]. 2018 [cited 2018 Dec 14]. Available from: http://baobabhealth.org/?p=1914

84. MySQL :: MySQL 5.7 Reference Manual [Internet]. [cited 2017 Dec 4]. Available from: https://dev.mysql.com/doc/refman/5.7/en/

85. Welcome to Baobab Health Trust [Internet]. [cited 2020 Feb 1]. Available from: http://www.baobabhealth.org/

86. R Core Development Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2011.

87. National Health Science Research Committee (NHSRC) | [Internet]. [cited 2018 Nov 28]. Available from: https://www.ncst.mw/?page_id=366

88. Mobile cellular subscriptions - Malawi | Data [Internet]. [cited 2020 Aug 26]. Available from: https://data.worldbank.org/indicator/IT.CEL.SETS?locations=MW

89. Ahmed M, Merga H, Jarso H. Predictors of virological treatment failure among adult HIV patients on first-line antiretroviral therapy in Woldia and Dessie hospitals, Northeast Ethiopia: a case-control study. BMC Infect Dis. 2019 Apr 3;19(1):305.

90. Koethe JR, Lukusa A, Giganti MJ, Chi BH, Nyirenda CK, Limbada MI, et al. Association between weight gain and clinical outcomes among malnourished adults initiating antiretroviral therapy in Lusaka, Zambia. J Acquir Immune Defic Syndr 1999. 2010 Apr 1;53(4):507–13.

91. Malawi scores highly on global 90:90:90 HIV targets [Internet]. Malawi Nyasa Times - News from Malawi about Malawi. 2020 [cited 2020 Jan 30]. Available from: https://www.nyasatimes.com/malawi-scores-highly-on-global-909090-hiv-targets/

92. Forster CBM, Martin WG Brinkhof CG, Andrew Boulle MS, Eric Balestre MM, Olivia Keiser AJ, Matthias Egger. Electronic medical record systems, data quality and loss to follow-up: survey of antiretroviral therapy programmes in resource-limitted settings. Bull World Health Organ. 2008 Dec;86(12):939–47.

93. Noureldin M, Mosallam R, Hassan SZ. Quality of documentation of electronic medical information systems at primary health care units in Alexandria, Egypt. East Mediterr Health J Rev Sante Mediterr Orient Al-Majallah Al-Sihhiyah Li-Sharq Al-Mutawassit. 2014 Mar 13;20(2):105–11.

94. Breimaier HE, Heckemann B, Halfens RJG, Lohrmann C. The Consolidated Framework for Implementation Research (CFIR): a useful theoretical framework for guiding and evaluating a guideline implementation process in a hospital-based nursing practice. BMC Nurs. 2015;14:43.

95. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009 Jul 21;6(7):e1000097.

96. Scobie HM, Edelstein M, Nicol E, Morice A, Rahimi N, MacDonald NE, et al. Improving the quality and use of immunization and surveillance data: Summary report of the Working Group of the Strategic Advisory Group of Experts on Immunization. Vaccine. 2020 Oct 27;38(46):7183–97.

97. Karuri J, Waiganjo P, Orwa D, Manya A. DHIS2: The Tool to Improve Health Data Demand and Use in Kenya. J Health Inform Dev Ctries [Internet]. 2014 Mar 18 [cited 2020 Dec 19];8(1). Available from: https://jhidc.org/index.php/jhidc/article/view/113

98. 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.

99. Mphatswe W, Mate K, Bennett B, Ngidi H, Reddy J, Barker P, et al. Improving public health information: a data quality intervention in KwaZulu-Natal, South Africa. Bull World Health Organ. 2012 Mar 1;90(3):176–82.

100. Mutale W, Chintu N, Amoroso C, Awoonor-Williams K, Phillips J, Baynes C, et al. Improving health information systems for decision making across five sub-Saharan African countries: Implementation strategies from the African Health Initiative. BMC Health Serv Res. 2013 May 31;13(Suppl 2):S9.

101. Githinji S, Oyando R, Malinga J, Ejersa W, Soti D, Rono J, et al. Completeness of malaria indicator data reporting via the District Health Information Software 2 in Kenya, 2011-2015. Malar J. 2017 Aug 17;16(1):344.

102. Kiberu VM, Matovu JK, Makumbi F, Kyozira C, Mukooyo E, Wanyenze RK. Strengthening district-based health reporting through the district health management information software system: the Ugandan experience. BMC Med Inform Decis Mak. 2014 May 13;14(1):40.

103. Wagenaar BH, Gimbel S, Hoek R, Pfeiffer J, Michel C, Manuel JL, et al. Effects of a health information system data quality intervention on concordance in Mozambique: time-series analyses from 2009–2012. Popul Health Metr [Internet]. 2015 Mar 26;13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4377037/

104. Gimbel S, Mwanza M, Nisingizwe MP, Michel C, Hirschhorn L. Improving data quality across 3 sub-Saharan African countries using the Consolidated Framework for Implementation Research (CFIR): results from the African Health Initiative. BMC Health Serv Res [Internet]. 2017 Dec 21;17(Suppl 3). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5763292/

105. Mpofu M, Semo B-W, Grignon J, Lebelonyane R, Ludick S, Matshediso E, et al. Strengthening monitoring and evaluation (M&E) and building sustainable health information systems in resource limited countries: lessons learned from an M&E task-shifting initiative in Botswana. BMC Public Health. 2014 Oct 3;14:1032.

106. Admon AJ, Bazile J, Makungwa H, Chingoli MA, Hirschhorn LR, Peckarsky M, et al. Assessing and improving data quality from community health workers: a successful intervention in Neno, Malawi. Public Health Action. 2013 Mar 21;3(1):56–9.

107. Hazel E, Chimbalanga E, Chimuna T, Nsona H, Mtimuni A, Kaludzu E, et al. Using Data to Improve Programs: Assessment of a Data Quality and Use Intervention Package for Integrated Community Case Management in Malawi. Glob Health Sci Pract. 2017 Sep 27;5(3):355–66.

108. 13.4 Using DHIS2 to improve data quality [Internet]. [cited 2020 Nov 10]. Available from: https://docs.dhis2.org/master/en/implementer/html/using-dhis2-to-improve-data-quality.html

109. Lawton R, Heyhoe J, Louch G, Ingleson E, Glidewell L, Willis TA, et al. Using the Theoretical Domains Framework (TDF) to understand adherence to multiple evidence-based indicators in primary care: a qualitative study. Implement Sci. 2016;11:113.

110. R in Action [Internet]. Manning Publications. [cited 2018 Nov 28]. Available from: https://www.manning.com/books/r-in-action

111. Leech NL, Onwuegbuzie AJ. Guidelines for Conducting and Reporting Mixed Research in the Field of Counseling and Beyond. J Couns Dev. 2010;88(1):61–9.

Appendix 1: Semi Structured Data Collection Tool

Data Quality and Fitness for Use in ART clinics for HIV/AIDS using Electronic Medical Record (EMR) systems in Malawi

Study Overview

The purpose of this study is to understand the uses of data for different roles directly or indirectly linked with the EMR in Malawi. The semi-structured interviews will attempt to understand "fitness to purpose" of data based on these uses. Participation in the study is completely voluntary and participants may stop at any time in the study. The study has minimal risk to the participants with confidentiality being the main risk. To minimize this risk, the study will not collect identifiable information and will keep the recordings of the study in a separate file. The study will also collect at least 6 participants for each stakeholder to increase confidentiality of the responses provided.

Question 1
What is your role at this organization?

Question 2
Do you use data in your role? How do you use the data?

Question 3
Are there other ways that you would like to use data in your role?

Question 4
When using the data in your work, what characteristics do you anticipate in your data? Do you find these characteristics in the data you access? What do you do when these characteristics are not there?

Question 5
When you think of data quality, what characteristics or features of the data do you think are important for your uses? (Potential follow-up questions: For example, is accuracy important? Is current data important and if so, how current?)

Question 6
Given that you have an electronic medical record system, what has been your experience with these expectations of data quality?

Question 7

Now I would like to show you a list of uses suggested by other people with a role similar to yours. Do you think these uses apply to your role in organization?

Table 4.3 Adapted from the WHO digital classification of health interventions

Stakeholder	Purpose of data use
Healthcare providers	Laboratory and diagnostics imaging management
	Prescription and medication management
	Healthcare provider training
	Healthcare activity planning and scheduling
	Referral coordination
	Client identification and registration
	Client health records
	Healthcare provider decision support
	Telemedicine
	Healthcare provider communication
	Healthcare provider feedback
Health system managers	Facility management
	Health financing
	Human resource management
	Supply chain management
	Public health event notification
	Civil registration and vital statistic
	Equipment and asset management
Data services	Data collection, management and use
	Data coding
	Location mapping
	Data exchange and interoperability

Question 8
I would also like to show you a list of data characteristics suggested by other people, not necessarily working in health care or in roles like yours. Are these characteristics important to you when you use data? Which are most important?

- Completeness
- Plausibility
- Conformance

Appendix 2: Template Analysis Table for Coding Purposes of Data Use informed by the WHO Classification of Digital of Health Interventions

Table 4.4 Definition of codes for purposes of data use

Purpose of data use category from the WHO CDHI	**Description**	**Themes**
Data collection management and use	Approaches to data collection, management, analysis, storage. This can include standalone interventions focusing exclusively on data collection and management, as well as data services to support other interventions, such as data visualization within supply chain management	Identify phrases that may map to this category
Client identification and registration	Client identity verification and enrollment into health services.	Identify phrases that may map to this category
Facility management	Digital approaches that enable administrative functions related to the management of facilities.	Identify phrases that may map to this category
Patient health tracking	The use of health records, review and track health status.	Identify phrases that may map to this category

Prescription and medication management	Approaches to facilitate the management of prescriptions, including tracking prescription orders and monitoring physical consumption of medication.	Identify phrases that may map to this category
Supply chain management	Approaches for monitoring and reporting stock levels, consumption and distribution of medical commodities. This can include the use of communication systems (e.g. SMS) and data dashboards to manage and report on supply levels of medical commodities.	Identify phrases that may map to this category
Defining and monitoring policy	Approaches to use data to suggest policies and monitor their implementation	Identify phrases that may map to this category
Healthcare training	The management and provision of education and training content in electronic form for health professionals [2]. In contrast to decision support, healthcare provider training does not need	Identify phrases that may map to this category
Human resource management	Digital approaches to manage the health workforce, including the use of databases to record training levels, certifications, and identification of health workers.	Identify phrases that may map to this category

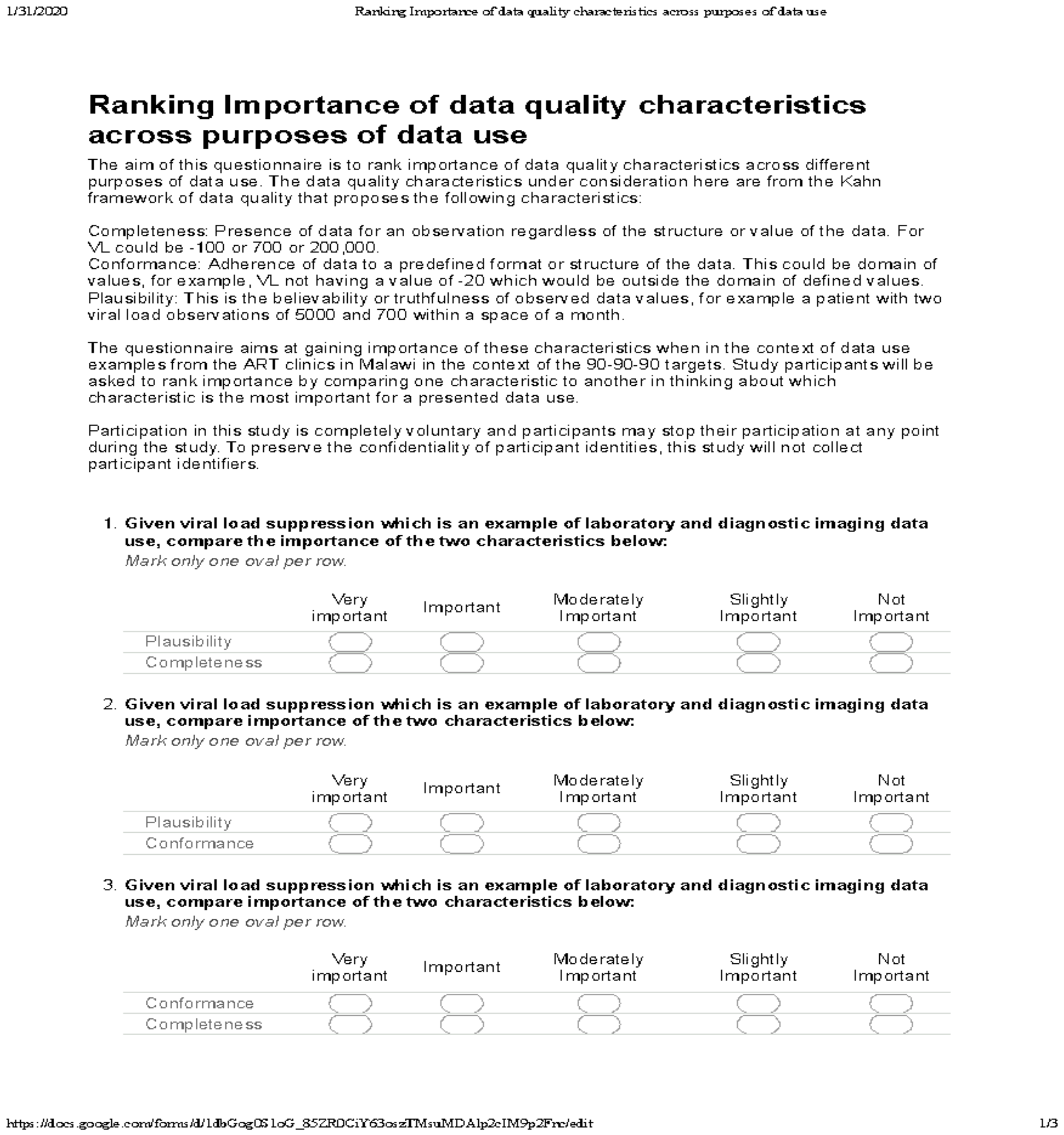
Client identification and registration	Client identity verification and enrollment into health services	Identify phrases that may map to this category
Healthcare provider decision support	Digitized job aids that combine an individual's health information with the health-care provider's knowledge and clinical protocols in order to assist health-care providers in making diagnosis and treatment decisions.	Identify themes that may map to this category
Referral coordination	Approaches to support communication and coordination mechanisms to facilitate referrals, both within the health sector and to other health-related sectors	Identify themes that may map to this category

Appendix 3: Template Analysis Table for Identifying Expected Data Quality Characteristics Informed by the Kahn Data Quality Framework for Secondary Data Use.

Table 4.5 Definition of codes for expected data quality characteristics

Kahn Framework Category	Description	Theme
Conformance	This focused on the agreement of the data against internal or external formatting, relational, or computational definitions. This characteristic did not attempt to measure the completeness or plausibility of the values observed	Identify phrases that may map to this category such as concordance, correctness, consistency, conformance.
Completeness	This was measured by assessing the frequencies of data attributes in a data set without reference to the data values.	Identify phrases that may map to this category such as representative, density, timeliness or currency, completeness.
Plausibility	This deals with believability or truthfulness of the values of data. Unlike completeness and conformance that deal with the presence and structure of data values, plausibility deals with how closely data values represent the world.	Identify phrases that may map to this category such as no duplicates, relative correctness, reliability, plausibility.

Appendix 4: Questionnaire for Ranking Importance of Data Quality Characteristics in Relation to the 90-90-90 Targets for Eradicating HIV/AIDS by 2030



1/31/2020 Ranking Importance of data quality characteristics across purposes of data use

Ranking Importance of data quality characteristics across purposes of data use

The aim of this questionnaire is to rank importance of data quality characteristics across different purposes of data use. The data quality characteristics under consideration here are from the Kahn framework of data quality that proposes the following characteristics:

Completeness: Presence of data for an observation regardless of the structure or value of the data. For VL could be -100 or 700 or 200,000.
Conformance: Adherence of data to a predefined format or structure of the data. This could be domain of values, for example, VL not having a value of -20 which would be outside the domain of defined values.
Plausibility: This is the believability or truthfulness of observed data values, for example a patient with two viral load observations of 5000 and 700 within a space of a month.

The questionnaire aims at gaining importance of these characteristics when in the context of data use examples from the ART clinics in Malawi in the context of the 90-90-90 targets. Study participants will be asked to rank importance by comparing one characteristic to another in thinking about which characteristic is the most important for a presented data use.

Participation in this study is completely voluntary and participants may stop their participation at any point during the study. To preserve the confidentiality of participant identities, this study will not collect participant identifiers.

1. **Given viral load suppression which is an example of laboratory and diagnostic imaging data use, compare the importance of the two characteristics below:**
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Completeness	○	○	○	○	○

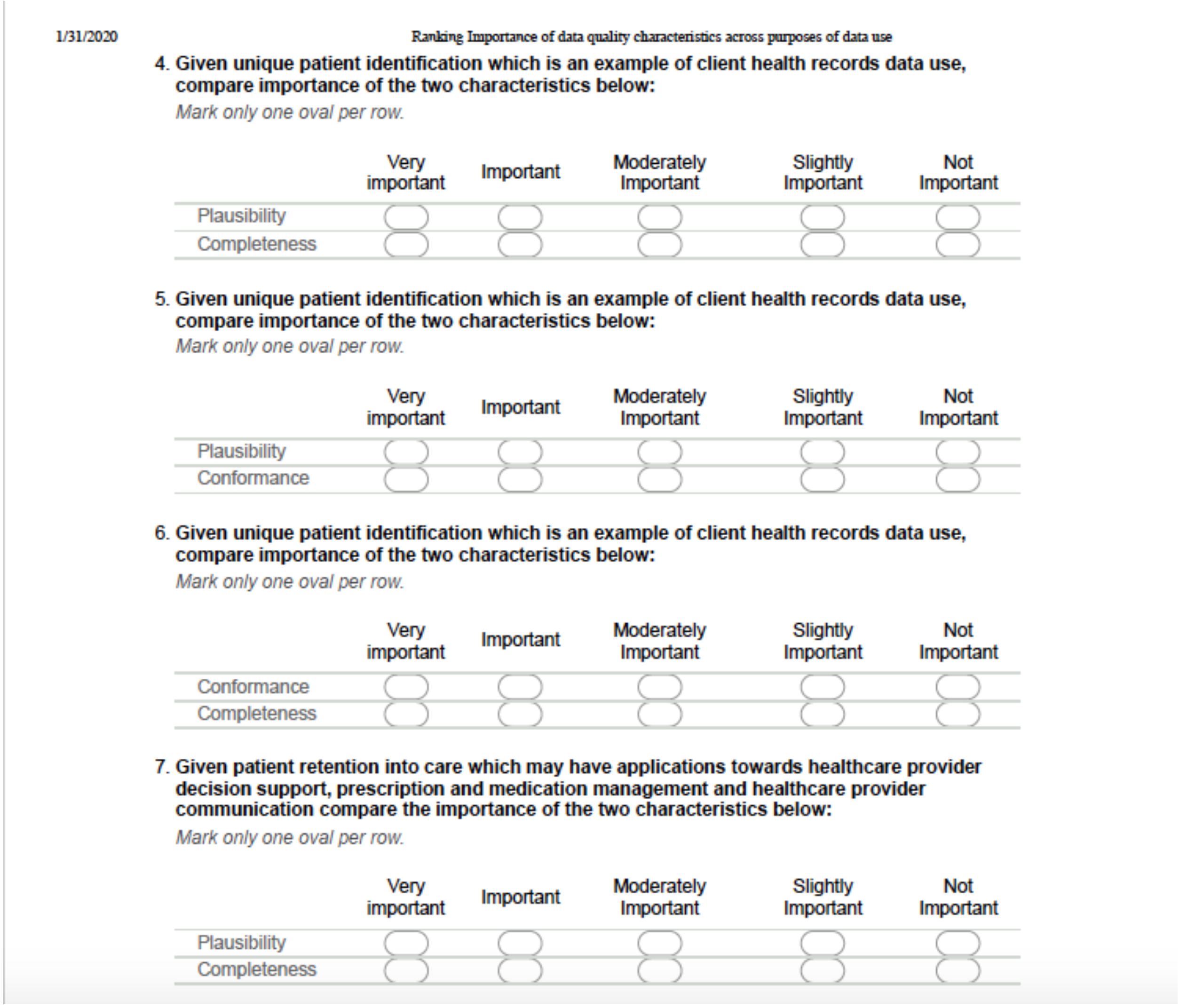
2. **Given viral load suppression which is an example of laboratory and diagnostic imaging data use, compare importance of the two characteristics below:**
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Conformance	○	○	○	○	○

3. **Given viral load suppression which is an example of laboratory and diagnostic imaging data use, compare importance of the two characteristics below:**
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Conformance	○	○	○	○	○
Completeness	○	○	○	○	○

https://docs.google.com/forms/d/1dbGog0S1oG_85ZRDCiY63oszTMsuMDAlp2cIM9p2Fnc/edit 1/3



1/31/2020 Ranking Importance of data quality characteristics across purposes of data use

4. **Given unique patient identification which is an example of client health records data use, compare importance of the two characteristics below:**

Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Completeness	○	○	○	○	○

5. **Given unique patient identification which is an example of client health records data use, compare importance of the two characteristics below:**

Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Conformance	○	○	○	○	○

6. **Given unique patient identification which is an example of client health records data use, compare importance of the two characteristics below:**

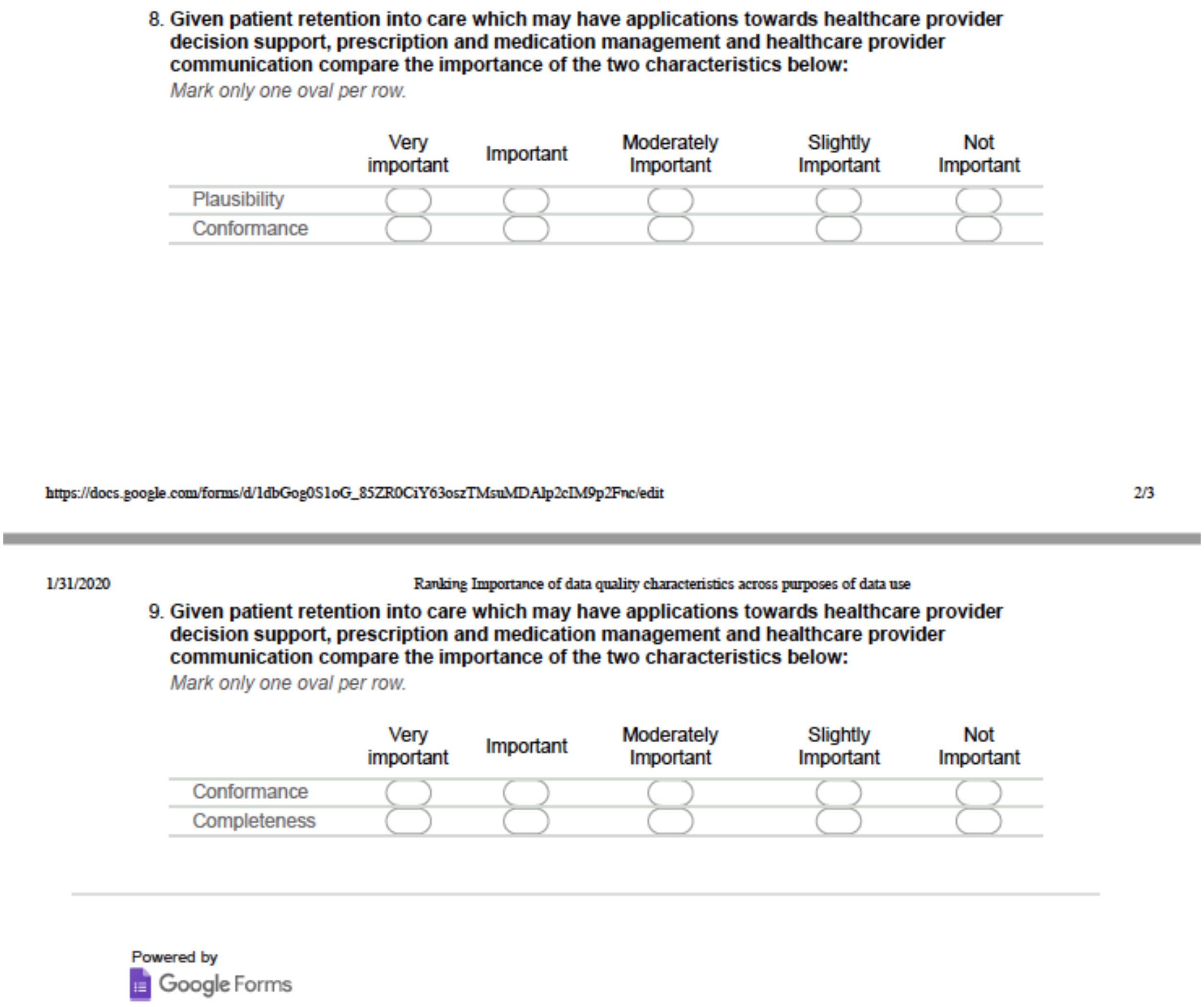
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Conformance	○	○	○	○	○
Completeness	○	○	○	○	○

7. **Given patient retention into care which may have applications towards healthcare provider decision support, prescription and medication management and healthcare provider communication compare the importance of the two characteristics below:**

Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Completeness	○	○	○	○	○



8. **Given patient retention into care which may have applications towards healthcare provider decision support, prescription and medication management and healthcare provider communication compare the importance of the two characteristics below:**
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Plausibility	○	○	○	○	○
Conformance	○	○	○	○	○

9. **Given patient retention into care which may have applications towards healthcare provider decision support, prescription and medication management and healthcare provider communication compare the importance of the two characteristics below:**
Mark only one oval per row.

	Very important	Important	Moderately Important	Slightly Important	Not Important
Conformance	○	○	○	○	○
Completeness	○	○	○	○	○

Appendix 5: Data Extraction table structures

1. Patient identification (for patient tracking) for the Second 90

Table 4.6 Fields extracted for First 90

Pseudo_Patient_ID	Pat_Current_Residence	Pat_Phone_Number	Pat_age

2. Patient care data: Drugs Dispensed for the Second 90

Table 4.7 Fields Extracted for Second 90, Drugs Dispensed

Pseudo_Patient_ID	Patient_Visit_Date	Prescribed_Drug	Dispensed_Drug	Quantity Dispensed

Patient care data: BMI calculation for the Second 90

Table 4.8 Fields Extracted for Second 90 BMI

Pseudo_Patient_ID	Patient_Visit_Date	Patient_Height	Patient_Weight	Pat_age

Patient care data: Adherence data for the Second 90

Table 4.9 Fields Extracted for Second 90, Drug Adherence

Pseudo_Patient_ID	Patient_Visit_Date	Drugs Dispensed on Previous Visit	ARV Drugs Remaining	Drug Dosage	Patient Adherence in Percent

3. Patient care data for the third 90: Viral Load Suppression

Table 4.10 Fields Extracted for Third 90, Viral Load

Pseudo_Patient_ID	Date_Started_ART	Patient_Visit_Date	Patient_VL_Result

Appendix 6: Detailed Updated definitions of Variable Proportions for Aim 2a.

Table 4.11 Updated proportions for AIM 2 (Appendix 6)

90-90-90 data use	Variable	Completeness	Atemporal Plausibility	Temporal plausibility
Patient tracing	Patient Address – the current place of residence	All patients registered with a current place of residence/all study patients registered within the study period.	All complete addresses in look-up master address table /all patients with a complete physical address current place of residence	All patients with physical address registered on or after enrolling in ART program / all patients with a complete physical address current place of residence
	Patients with contact phone number	All patients registered with a documented mobile phone number/all patients registered within the study period.	All patients registered with the expected sequence of mobile phone number (correct carrier prefix) / All patients registered with a complete phone number	All patients with mobile phone number registered on or after enrolling in ART program / all patients with a complete physical address current place of residence
Quality of patient care	Prescription versus dispensed drugs	Number of prescriptions visits/number of all patient visits during study	Number of prescriptions with quantity of drugs dispensed for 30 to 90 days duration/number of dispensations made during study period.	Number of patients that had their ARVs dispensed on their visit date/number of dispensations made during study period
	Patients drug adherence monitoring	Number of patients with patient adherence visits/ Number of all patient visits during study	Number of patients with adherence in percentage to drugs between 95 to 105/ Number of patients	Number of patients that had their adherence measured on their

			with adherence observations	visit date/ number of patients with adherence observations
	Body mass index – weight and height	Number of patients with height visits / number of study patients	Number of patient observations with height that is within bounds (height for age and absolute values)/ Number of patients with height visits	Number of patient height observations made on the same day of the patient visit / Number of patients with height visits
		Number of patients with weight visits / Number of all patient visits during study	Number of patients with weight that is not out of bounds according to weight for age and absolute minimum and maximum values/ number of patients with complete weight observations	Number of patient weight observations made on the same day of the patient visit / Number of patients with weight visits
Viral load suppression	Viral load results present	Number of patients with one observation of viral load result / Number of study patients	Number of patients whose viral load was measured between 6 to 12 Months after their start date / Number of complete viral load measurements	Number of patients with viral load values not measured before their ART start date / Number of complete viral load measurements

Appendix 7: Additional Details for Qualitative Analysis of Aim 1a.

3.3.1.1 Client Identification and Registration

I observed that most interviewees started their description of their purposes of data use with patient registration and patient identification. I coded the theme of patient registration and identification under the WHO category of client identification and registration.

Unique identification was emphasized as an essential purpose of data use to ensure that care is being given to the correct patients. The example in the table above shows the possibility of duplicate patients that may not receive the right care in the event of wrong identification.

Another purpose for conducting patient registration and identification is to facilitate follow up of patients if a patient has missed their appointment date at the ART clinic. I found the patient follow up in responses for other purposes of data use, such as patient health tracking that I have presented below. The purposes of data use of client identification and registration are essential for both the first 90 in the identification of HIV positive patients and the 2nd 90 in ensuring that patients are retained in care.

3.3.1.2) WHO CDHI category, Patient health tracking

Two main themes were coded into this category and were reported mainly in the stakeholder groups of health care providers and data services. This purpose of data uses maps to the second 90 in the 90-90-90 initiative of eradicating HIV/AIDS by 2030. The second 90 intends to retain 90% of all patients enrolled in an ART clinic for HIV(11). Therefore, this data purpose of use links with the targeted quantitative data quality assessment of variables that enable using data to meet this purpose. I established the themes below under this category:

Trace patients lost to follow up

Stakeholders mentioned that they use data for this purpose to identify patients that had missed their appointments and were at risk of defaulting from treatment. Healthcare providers cared about getting patients back into therapy so that they continue to receive medication.

The choice of the word “entices” emphasized the more exceptional picture that healthcare providers may have on maintaining the patients into care. On the other hand, data services stakeholders use data to trace patients mainly to update patient outcomes with the real picture of who is present at the clinic. The real picture helps them in collecting the correct number of drugs for a clinic day from the pharmacy.

Maintaining quality of care or improving quality improvement projects

I coded the theme of quality improvement into the main category of patient health tracking as well. I observed that this theme was mentioned frequently and emphasized as a purpose of data use. This theme maps to the 2nd 90 as well as the 3rd 90 in HIV eradication the goals. I found that the stakeholder groups that emphasized this theme were healthcare providers and health managers. Healthcare providers, mentioned their desire to manage patient side effects properly which improves the quality of life of patients(12) and therefore improves retention in health facilities

3.3.1.3 WHO CDHI category, prescription and medication management

I found this category with mainly stakeholders such as clinicians and nurses. In a task shifting environment, nursing staff also reported being responsible for prescription (13). This purpose of data use is important for the 2nd 90 of retention since the strategy for eradicating HIV/AIDS by 2030 is via treatment (14). Therefore, using electronic data for the purpose of prescription and medication management gives a basis for the data quality assessment that this study did as well. I found the theme below under this category:

Drugs to prescribe based on presenting conditions

Healthcare providers indicated that they use data to understand whether the patient has any conditions that require changing them the medication that they are on or maintaining them on the medication.

3.3.1.4 WHO CDHI category, Healthcare decision support

This formed as a category for interviewee responses that fit to the themes of clinical decisions and counseling decisions. Both these themes are important as purposes of data use in ART clinics since they speak to the HIV targets in the country for retention as well as viral load monitoring (11).

Clinical decisions

Healthcare providers such as clinical officers gave this purpose of data use as an example of how they make decisions regarding management of patients who seem to be failing on treatment. Continued rising of viral load in the presence of ARV treatment may indicate ARV treatment failure. Viral load is measured in copies/ml of blood to understand the extent of HIV disease in patients. Viral load is the leading guide for clinicians to make crucial decisions on management of the patient in case of rising viral load. The recommended viral load for patients to remain healthy and reduce morbidity due to HIV opportunistic infections, is below 1000 copies/ml(6) and this is considered viral load suppression. Patients that have less than 40 copies/ml are considered to have undetectable HIV hence no risk of sexually transmitting the virus to their partners (6)(15). The goal of eradication is, therefore, to ensure that 90% of patients on ART for HIV achieve viral load suppression and having this as a purpose of data use confirms the need for these data. I also performed a data quality assessment for viral load quantitatively, and the results are in the next section.

Counseling decisions

This theme appeared as a purpose of data use with the example of making sure that patients are given the right education in terms of procedures they receive. Healthcare providers indicated that they reference data such as patient age to know how much emphasis they should put on their counseling efforts.

Counseling is vital as a purpose of data use, particularly in the context of this study since patients receive counseling triggered by different data points such as the proportion of drugs that are calculated to have been taken concerning their previous clinic visit. Adherence is one of the variables that this study analyzes regarding the 90-90-90 target of patient retention. Counselling has also been shown to be one of the factors that influence patients, particularly in ART settings, to continue seeking care at health facilities. The importance of counseling was confirmed in a study that assessed factors of retention for pregnant women started on lifelong antiretroviral therapy via option B+(16).

Referral Coordination

I coded referral coordination with two themes of intra-hospital referrals where healthcare providers indicated that they use data as a means of referring patients that require additional care to more experienced doctors. This type of referral would usually happen within the same health facility. Additionally, healthcare providers touched on the data purpose of the use of transferring patients to more resourced health facilities such as tertiary health facilities. This purpose of data use is consistent with the expectation of smaller health facilities, referring patients to more prominent and well-resourced health centers. An interesting point raised as part of the response is the view that the documented care at the source health facility may be used for legal purposes if there is an adverse outcome for the referred patient such as a death.

3.3.1.5) WHO CDHI category, Data collection, management, and use

I found this as a common category for all stakeholders. Although this category was common across all stakeholders in this study, the themes varied. I found that this finding echoed previous research on the goal of implementing EHRs, where the primary purpose of the data collected was mainly for generating routine reports (17). Some of the themes that emerged under this category were:

Reporting

Many healthcare providers mentioned reporting as an essential purpose of data use about the EHR systems that they use. However, a closer assessment of "reporting" as a theme under the category data collection, management, and use, reveals it as a means to an end in terms of the actual professional roles of the stakeholders. For example, clinical facing stakeholders such as clinicians and nurses valued reporting to aid them in their daily work such as knowing the correct number of patients in their clinics so that they may prepare how to treat them,

On the other hand, stakeholders in data services cared more about data collection, management, and use to submit routine reports to the MoH. This view was also true for other stakeholder roles involved in administrative activities that mentioned the theme of reporting as fulfilling a routine requirement to an external body.

Data analysis and interpretation

I found the theme above mainly in the stakeholder roles of health managers, such as the MoH officials or EMR donors. I also found that advanced data service roles such as data officers and monitoring and evaluation officers also mentioned this as their purpose of data use. Most of the time, this was referred to as data analysis to understand certain national interest conditions and inform planned interventions on the part of national programs.

Within the data services stakeholders, there were data officers or monitoring and evaluation officers that mentioned data analysis in the context of cleaning data and responding to questions that their supervisors would want to answer based on the data being collected.

3.3.1.6 WHO CDHI category, Facility management

This was also a primary purpose of use for health managers at different levels of the healthcare system in Malawi. Health managers, those that oversee health facility activities cited examples of data use that affected their immediate needs such as staffing requirements or need for infrastructure. The WHO CDHI has human resources as an area with examples of digital health interventions. Since human resources in this study were mentioned as an example of a data purpose of use in the context of facility management, I coded it under health facility management. The end goal of adequate resources at health facilities may speak directly to the issue of retention of patients since inadequate staff members and resources have been shown to affect optimal care to patients and thus potentially reduce retention of patients (18)(16)(19).

The themes that appeared under this category are, therefore:

Allocation of physical infrastructure

Health managers at facilities cited that they use data to demonstrate the need for additional infrastructure since most of them receive equipment to use via the standard classification of the health facility. So, if a health facility is a health center, they will have designated number of rooms and specific equipment that is expected to be available in such places. A nurse who is the in-charge of one of the health centers attested to this.

Human resource management

This was another common theme that went beyond adequate numbers of members of staff for different departments at a health facility but also using data to understand the various training needs of the staff.

Human resource capacity is an integral part of ART clinics in Malawi since it ensures that patients are receiving care according to national clinical practice guidelines for the management of HIV/AIDS (20). Human resource capacity, in turn, ensures that patients are getting good quality care, which provides progress towards the goals of HIV eradication.

3.3.1.7 WHO CDHI category, supply chain management

All stakeholders mentioned a supply chain theme as one of their purposes of data use. Supply chain management was mentioned as an end goal for collecting data using the EHR. Two main themes appeared here. One for healthcare providers and staff working at health facilities such as data services. The other more for public health officials focusing on the national and sub-national scales.

Ordering drugs or medical supplies

This theme was a purpose of data use that was mostly reported by stakeholders working at sub-national levels, such as health facilities. I found these stakeholders to be mainly healthcare providers such as nurses and those in data services such as data clerks. Ordering of drugs, in this case, meant requesting drugs from their facility drug store to their clinic to dispense to patients during clinic time. For data services people who collect and transport the drugs from the drug storeroom, they want to receive the right quantity.

On the other hand, for health managers who would be clinic managers or public health officials, their purpose of data would target ensuring population-level availability of drugs at different scales of national and facility levels. The availability of drugs is a crucial feature of the national ART program in Malawi as there is a robust supervision model to ascertain patient outcomes every three months (21).

Drug forecasting

Drug forecasting was another theme that appeared mainly in health managers working at the national level. Faced with challenges in timely reporting, one essential technique that I got from the interviewees is to use available data to estimate quantities of ARV drug regimens that are required for the national program.

To an extent, the different purposes of data use reported seem to support the supply chain of various commodities as the end goal, particularly for HIV being a chronic illness that requires taking ARV drug regimens daily.

3.3.1.8 WHO CDHI category, Defining and monitoring policy

Defining and monitoring policy was almost exclusively for health managers with other staff accepting their roles as only supportive in the way they use the data. For example, data services indicated that they produce data for decisions that are made at senior management levels. I feel this purpose is relevant to this study because the stakeholder group of healthcare managers makes decisions about the EHR in the country. I found the themes below:

Project decisions

The purpose of data use of project decisions pointed to the application of electronic data to help in defining specific interventions driven by the MoH partners or by the ministry itself. An example from one respondent was on how they use data to make decisions on scaling the EHR in the country.

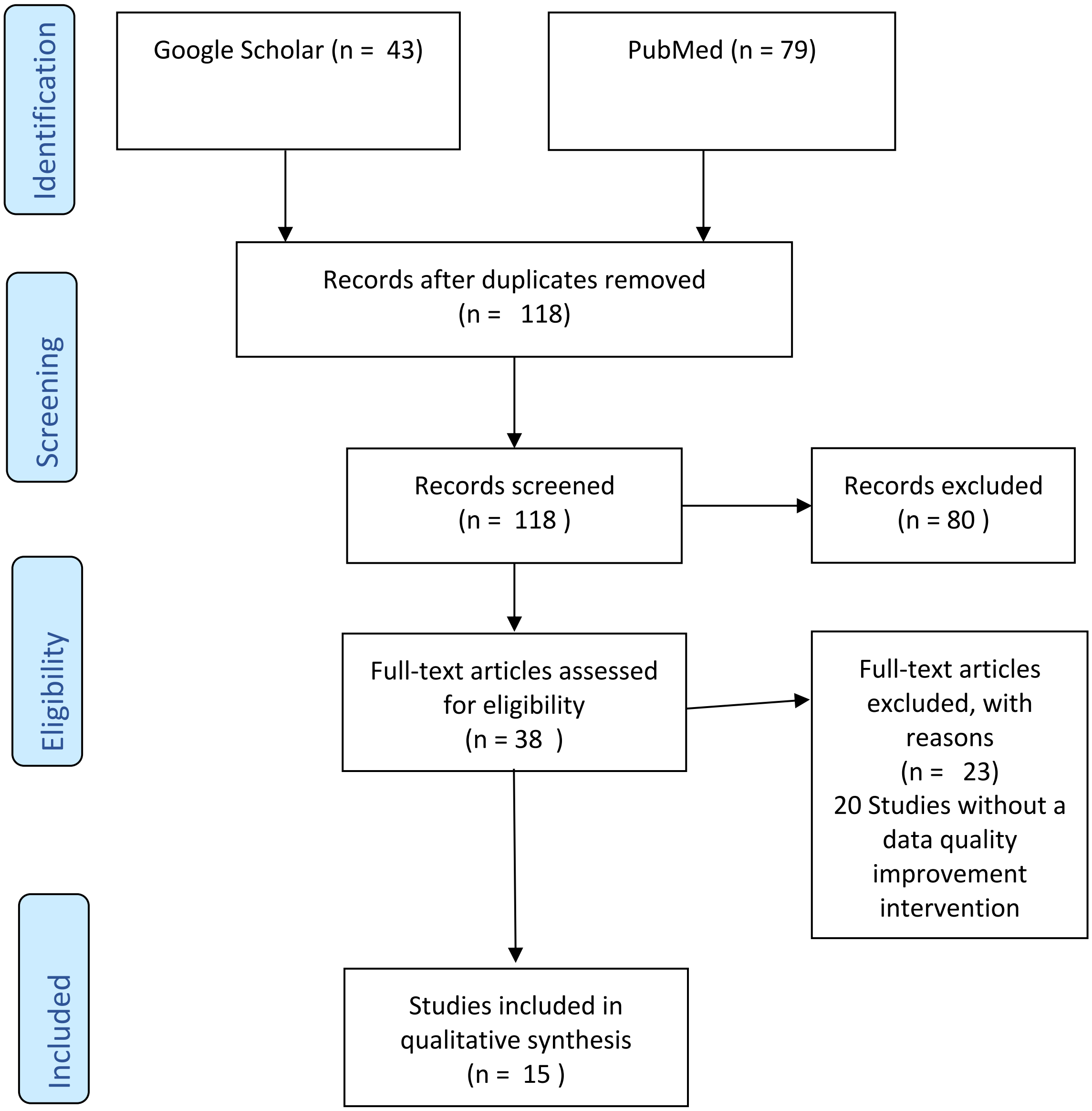
Project decisions such as reported here are not strange as there have been other decisions that started as projects on a smaller scale based on data but were eventually scaled up in the country. An example is an initiative to enroll all women that test HIV positive immediately on ARVs, which began with a few health facilities as a project and then scaled nationally (22).

Guideline definitions or patient monitoring

Guideline definitions or patient monitoring was another common theme for data use with health managers such as public health officials within the MoH and clinic partners at specialized clinics that serve as centers of excellence for treating HIV/AIDS.

Other stakeholders mentioned this theme in line with using data to understand currently implemented guidelines and then making suggestions on how best to improve them. I observe here that this data use is also crosscutting with the category of data collection, management, and use mainly the data analysis and interpretation theme, which would then lead to new ways of implementing existing programs.

Appendix 8: PRISMA 2009 Flow Diagram on Data Quality Improvement Interventions



Appendix 9: PRISMA Checklist(95)

Table 4.12 Checklist for PRISMA literature review

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	77
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	77,78
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	77
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	77
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	78
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	78

Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated. PubMed: ((((((Data quality improvement interventions) AND (electronic health data)) OR (health information systems)) OR (digital health)) AND (Low Income Countries)) OR (developing countries)) OR (LMICs) Google Scholar: "Data quality interventions" for electronic health data or health information systems Low Income Countries	
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	78
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Independently
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	78
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	N/A
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	N/A

Appendix 10: Guidelines for Conducting and Reporting Mixed Research (111)

Table 4.13 Checklist for Conducting Mixed Methods Research

1. Research Formulation	Page Number
1.4.1. Specify the purpose of the study.	19
1.4.2. Specify the purpose for combining qualitative and quantitative approaches (e.g., identify representative sample members, conduct member check, validate individual scores on outcome measures, develop items for an instrument, identify barriers and/or facilitators within intervention condition, evaluate the fidelity of implementing the intervention and how it worked, enhance findings that are not significant, compare results from the quantitative data with the qualitative findings).	3, 30
1.5.1. Avoid asking research questions that lend themselves to yes/no responses.	19
1.5.2. Present mixed research questions (i.e., questions that embed both a quantitative research question and a qualitative research question within the same question) when possible.	19
2. Research Planning	
2.1.1. Specify the initial and final sample sizes for all quantitative and qualitative phases of the study.	37,44,49
2.1.2. Present all sample size considerations made for the quantitative phase(s) (i.e., a priori power) and qualitative phases (e.g., information-rich cases).	49
2.1.3. Present the sampling scheme for both the quantitative and qualitative phases of the study.	38,44,49
2.1.4. Describe the mixed sampling scheme (i.e., concurrent–identical, concurrent–parallel, concurrent–nested, concurrent–multilevel, sequential–identical, sequential–parallel, sequential–nested, and sequential–multilevel).	31
2.1.5. Clarify the type of generalization to be made (i.e., statistical generalization, analytic generalization, and case-to-case transfer) and link it to the selected sampling design, sampling scheme, and sample size(s).	
2.2.1. Outline the mixed research design.	31
2.2.2. Specify the quantitative research design (i.e., historical, descriptive, correlational, causal–comparative/quasi-experimental, and experimental).	38,44,48
2.2.3. Specify the qualitative research design (e.g., biography, ethnographic, auto-ethnography, oral history, phenomenological, case study, grounded theory).	40

3.	**Research Implementation**	
3.1.1.	Outline the mixed data collection strategy.	31
3.1.2.	Present information about all quantitative and qualitative instruments and the process of administration.	39,44,48,50
3.2.1.	Outline the mixed data collection strategy (i.e., data reduction, data display, data transformation, data correlation, data consolidation, data comparison, and data integration).	32
3.2.2.	Provide relevant descriptive and inferential statistics for each statistical analysis.	50,52
3.2.3.	Discuss the extent to which the assumptions (e.g., normality, independence, equality of variances) that underlie the analyses were met, as well as any observations that might have distorted the findings (e.g., missing data, outliers).	NA
3.2.4.	Specify the statistical software used.	56
3.2.5.	Specify where the responsibility or authority for the creation of categories resided (i.e., participants, programs, investigative, literature, or interpretive), what the grounds were on which one could justify the existence of a given set of categories (i.e., external, rational, referential, empirical, technical, or participative), what was the source of the name used to identify a given category (i.e., participants, programs, investigative, literature, or interpretive), and at what point during the research process the categories were specified (i.e., a priori, a posteriori, or iterative).	40, 42
3.2.6.	Specify the name of the technique used to analyze the qualitative data (e.g., content analysis method of constant comparison, discourse analysis, componential analysis, keywords in context, analytic induction, word count, domain analysis, taxonomic analysis).	40
3.2.7.	Specify the qualitative software used.	40
3.3.1.	Discuss the threats to internal validity, external validity, and measurement validity and outline the steps taken to address each of these threats to internal validity, external validity, and measurement validity.	39,35
3.3.2.	Discuss the threats to trustworthiness, credibility, dependability, authenticity, verification, plausibility, applicability, confirmability, and/or transferability of data and outline all verification procedures used.	40
3.3.3.	Discuss mixed research legitimation types (i.e., sample integration legitimation, insider–outsider legitimation, weakness minimization legitimation, sequential legitimation, conversion legitimation, paradigmatic mixing legitimation, commensurability legitimation, multiple validities legitimation, and political legitimation).	33,74,82
3.4.1.	Interpret relevant types of significance of the quantitative findings (i.e., statistical significance, practical significance,	75-80

	clinical significance, and economic significance).	
3.4.2.	Conduct post hoc power analysis for all statistically nonsignificant findings.	NA
3.4.3.	Interpret the significance (i.e., meaning) of qualitative findings.	NA
3.4.4.	Discuss criteria for evaluating findings in mixed research studies (e.g., within-design consistency, conceptual consistency, interpretive agreement, interpretive distinctiveness, design suitability, design fidelity, analytic adequacy, interpretive consistency, theoretical consistency, integrative efficacy).	82,98
3.5.1.	Describe all steps of the mixed research process.	
3.5.2.	Describe the context in which the mixed research study took place.	32
3.5.3.	Ensure that the mixed research report is accurate and complete; does not distort differences within and among individuals and groups; is free from plagiarism or misrepresentation of the ideas and conceptualizations of other scholars; and contains findings that are adequately accessible for reanalysis, further analysis, verification, or replication.	82
3.5.4.	Present all ethical considerations that were addressed in the study (e.g., informed consent, confidentiality, incentives, funding sources, potential conflicts of interest, biases).	57
3.5.5.	Specify study approval in accordance with an institutional review board either in the report or in the cover letter submitted to the editor.	57
3.5.6.	Present recommendations for future research that culminate in a validation, replication, or extension of the underlying study.	104